# A Survey on Deep Learning Methods for Power Load and Renewable Energy Forecasting in Smart Microgrids

Sheraz Aslam[a,b,*], Herodotos Herodotou[a], Syed Muhammad Mohsin[b], Nadeem Javaid[b], Nouman Ashraf[c], Shahzad Aslam[d]

[a] *Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, Limassol, Cyprus*
[b] *Department of Computer Science, COMSATS University Islamabad, Pakistan*
[c] *Telecommunications Software and Systems Group (TSSG), Waterford Institute of Technology, Ireland*
[d] *Department of Statistics and Mathematics, Institute of Southern Punjab, Multan, Pakistan*

**Abstract**

Microgrids have recently emerged as a building block for smart grids combining distributed renewable energy sources (RESs), energy storage devices, and load management methodologies. The intermittent nature of RESs brings several challenges to the smart microgrids, such as reliability, power quality, and balance between supply and demand. Thus, forecasting power generation from RESs, such as wind turbines and solar panels, is becoming essential for the efficient and perpetual operations of the power grid and it also helps in attaining optimal utilization of RESs. Energy demand forecasting is also an integral part of smart microgrids that helps in planning the power generation and energy trading with commercial grid. Machine learning (ML) and deep learning (DL) based models are promising solutions for predicting consumers' demands and energy generations from RESs. In this context, this manuscript provides a comprehensive survey of the existing DL-based approaches, which are developed for power forecasting of wind turbines and solar panels as well as electric power load forecasting. It also discusses the datasets used to train and test the different DL-based prediction models, enabling future researchers to identify appropriate datasets to use in their work. Even though there are a few related surveys regarding energy management in smart grid applications, they are focused on a specific production application such as either solar or wind. Moreover, none of the surveys review the forecasting schemes for production and load side simultaneously. Finally, previous surveys do not consider the datasets used for forecasting despite their significance in DL-based forecasting approaches. Hence, our survey work is intrinsically different due to its

*Corresponding author
*Email addresses:* sheraz.aslam@cut.ac.cy (Sheraz Aslam), herodotos.herodotou@cut.ac.cy (Herodotos Herodotou), syedmmohsin9@yahoo.com (Syed Muhammad Mohsin), nadeemjavaid@comsats.edu.pk (Nadeem Javaid), nouman.ashraf@seecs.edu.pk (Nouman Ashraf), Shahzad.ddp@gmail.com (Shahzad Aslam)

*Preprint submitted to Renewable and Sustainable Energy Reviews* *March 19, 2021*

data-centered view, along with presenting DL-based applications for load and energy generation forecasting in both residential and commercial sectors. The comparison of different DL approaches discussed in this manuscript reveals that the efficiency of such forecasting methods is highly dependent on the amount of the historical data and thus a large number of data storage devices and high processing power devices are required to deal with big data. Finally, this study raises several open research problems and opportunities in the area of renewable energy forecasting for smart microgrids.

*Keywords:* `Energy forecasting; Renewable energy; Deep learning; Artificial neural networks; Machine learning`

## 1. Introduction

The power sector is moving towards renewable energy sources (RESs) because of their low price and massive contributions in reduction of carbon emissions. RESs consist of a number of resources, which include bioenergy, wind energy, hydropower, photovoltaic (PV) energy, etc. Usually, these RESs are operated in islanded and grid-connected modes [1]. Solar and wind energies are generated by installing PV panels and wind turbines (WTs), respectively, and these are handy in most places around the globe. Besides, RESs play an important role in minimizing carbon emissions among various electricity sources [2, 3, 4, 5, 6, 7], as shown in Figure 1. Moreover, Figure 2 indicates the yearly proportion of renewable power contribution to the whole electricity generation of some leading countries of the world. Brazil generates a huge amount of power from renewable sources (see Figure 2) in order to meet the consumers' power demand.

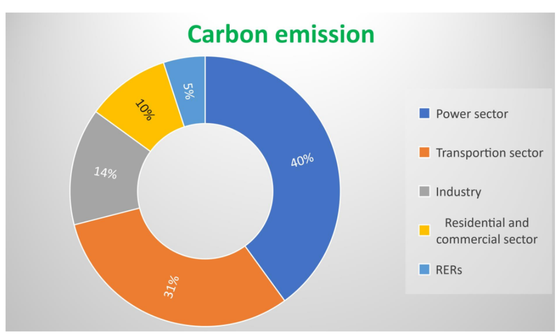


Figure 1: Sector-wise carbon emissions around the world [8]

Solar panels convert direct sunlight to electrical energy, while WTs generate electric power from wind. The key characteristics of these energy sources are limited controllability, limited predictability, and power output variability as the power produced from RESs completely relies upon environmental factors like solar irradiance, temperature, humidity, and wind speed [9]. For example,

Table 1: List of abbreviations

| Acronym | Description | Acronym | Description |
|---|---|---|---|
| ACCE | Adaptive circular conditional expectation | LM | Load monitoring |
| AE | Auto Encoder | LSTM | Long short term memory |
| AEMO | Australian energy market operator | LSTM-EFG | LSTM-enhanced forget-gate |
| AI | Artificial intelligence | MFE | Multistage forecast engine |
| ALM | Adaptive learning model | MI | Mutual information |
| ANFIS | Adaptive neuro-fuzzy inference system | MLR | Multiple linear regression |
| ANN | Artificial neural network | MLP | Multilayer perceptron |
| AR | Auto-regressive model | NARX | Nonlinear Auto regressive network with exogenous variables |
| ARX | Auto-regressive with exogenous input | NMAE | Normalized mean absolute error |
| BEC | Building energy consumption | NN | Neural network |
| BRT | Boosted regression tree | NREL | National renewable energy laboratory |
| CNN | Convolutional neural network | NWP | Numerical weather prediction |
| DBN | Deep belief network | NWS | National weather service |
| DCWT | Dual-tree complex wavelet transform | NWTC | National wind technology center |
| DE | Differential evolution | PDRNN | Pooling-based deep RNN |
| DL | Deep learning | PV | Photovoltaic |
| DLSTM | Deep LSTM | p-WPRF | Probabilistic wind energy ramp forecasting |
| DNN | Deep neural network | RBMs | Restricted boltzmann machines |
| DQR | Direct quantile regression | Relu | Rectified linear unit |
| EMSs | Energy management systems | RESs | Renewable energy sources |
| ENN | Elman neural network | RICNN | Recurrent Inspection CNN |
| EO | External optimization | RNN | Recurrent neural network |
| ESSs | Energy storage systems | SSA | Singular spectrum analysis |
| EVs | Electric vehicles | SVRM | Support vector regression machine |
| EWT | Empirical wavelet transformation | UAVs | Unmanned aerial vehicles |
| GA | Genetic algorithm | V2G | Vehicles to grid |
| GABPNN | GA back-propagation NN | WIND | Wind integration national dataset |
| GBR | Gradient boosting regression | WPD | Wavelet packet decomposition |
| GRU | Gated recurrent unit | WPF | Wavelet packet filter |
| HELM | Hysteretic extreme learning machine | WT | Wavelet transform |
| ICT | Information and communication technologies | WTs | Wind turbines |
| IMFs | Intrinsic mode functions | WTD | Wavelet threshold denoising |
| IoT | Internet of things | | |
| KEPCO | Korea electric power corporation | | |

PV panels produce higher energy in case of high solar radiation (clear sky) and they generate minimum energy (may be 0) during cloudy weather or at night times. On the contrary, WTs generate minimum energy (may be zero) in case of lower and higher wind speed than cut-in and cut-out speed, respectively [1]. Thus, large fluctuations in power generation from PV plants and WTs introduce several challenges, including voltage irregulations as well as reserve power flow problems and power distribution issues [10]. To make matters worse, energy consumers also exhibit intermittent behavior in power consumption because of various factors, like environmental changes, user preferences, etc.

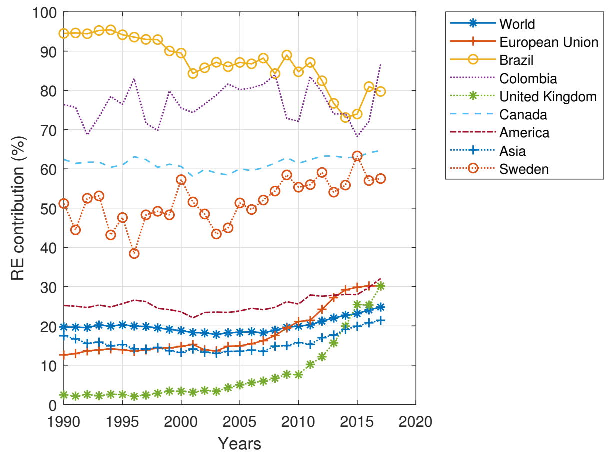


Figure 2: RESs contribution in whole power generation of a few countries of the world [11]

As mentioned above, the integration of RESs complicate the power grid operations and microgrids introduce difficulties in maintaining balance between energy generation and consumption (see Figure 3 for microgrid architecture). Therefore, accurate forecasting of energy generation from RESs (i.e., PV panels and WTs) along with electric load forecasting is an exigent need of the current smart grid era. Accurate load/demand forecasting allows the utility companies to control demand-driven supply effectively and produce surplus power from other resources (traditional power generation portfolios) when RESs are unable to meet consumers' demand.

Reliable prediction of wind and solar power generation form WTs and solar panels, respectively, is a challenging task, as it relies entirely on weather patterns (e.g., humidity, temperature, irradiance, etc.) [9, 1, 12]. Forecasting can be performed using several methods, including physical models [13], machine learning (ML) [14, 15], and (more recently) deep learning (DL) [16, 17]. In the last decade, ML and DL approaches have been applied in several domains of computational intelligence and forecasting, where they demonstrated promising efficacy. For example, they are employed for energy optimization and forecasting in smart microgrids [18, 17], energy prediction in wheat production [19], health services improvements [20, 21, 22], performance improvement in wireless networks [23], flood management [24], and hydrogen production forecasting [25]. However, all the forecasting methods have their own pros and cons. For instance, physical methods are effective in predicting the dynamics of the atmosphere, but they need significant computational resources since a huge amount of data is required to calibrate the dynamics of the atmosphere. Further issues

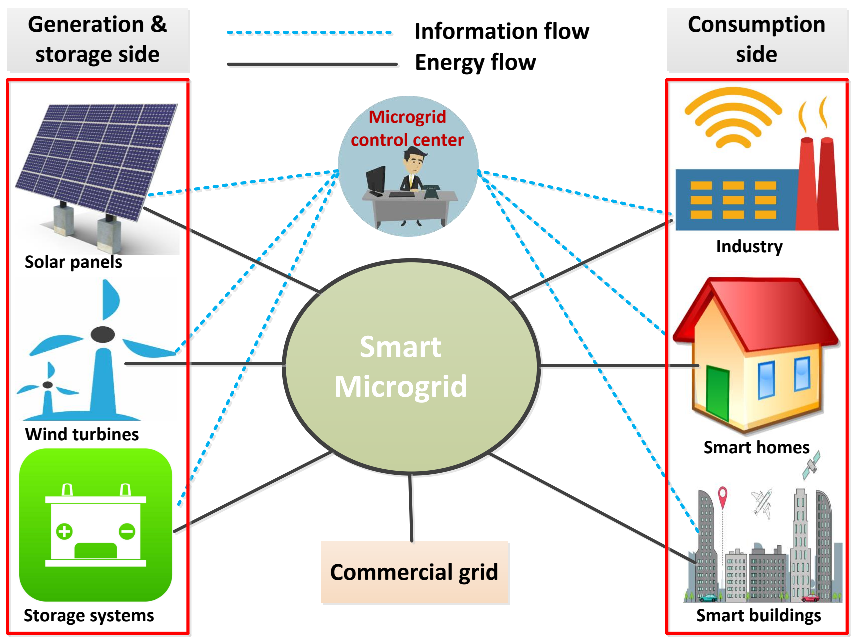


Figure 3: A typical microgrid architecture

arise when physical approaches find unexpected estimation errors, while they are also not suitable for short-term forecasting horizons. Similarly, most of the current renewable energy prediction statistical models are designed as linear models that limits their ability to solve more complex forecasting issues with longer forecasting time horizons.

Contrary to physical models, ML-based forecasting approaches usually offer more accurate results than statistical and physical models due to their advanced data mining and feature extraction capabilities. However, as a general rule, ML-based forecasting approaches use some "shallow" models as their central learning concepts. Typical shallow patterns are trees, regressors, or neural networks with zero or one hidden layer. It is well known that the training of such shallow models requires a great deal of experience and skill. Moreover, the theoretical study of shallow structures is often challenging. Thus, in practical applications, shallow models have significant drawbacks. However, it has been recently established that DL-based energy generation and power load forecasting approaches outperform the aforementioned methods as, unlike ML-based approaches, DL-based approaches do not suffer from hand-engineered feature selection, sample complexity, and weak generalization efficiency. [26].

Even though the forecasting of load demand and energy from RESs is a new research area, it has already gained significant attention from the research community. Lately, a lot of research studies have proposed DL-based approaches for such forecasting, while several survey/review works have been conducted [27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37], where they attempted to survey DL methods for energy or load forecasting from various perspectives and scopes. For instance, some recent survey papers present an overview of microgrid and RESs, such as solar power, wind energy, geothermal energy, hydro energy, etc. [27, 32]. The work presented in [18] reviews ten major ML models that were frequently employed in energy systems. A brief review of the load monitoring (LM) strategies is discussed in [28]. Surveys of building energy consumption prediction and overview of ML methods are given in [29, 30, 37]. A review study presented in [31] discusses DL-based methods for solar irradiance prediction, while the papers [33, 34, 35] disclose recent studies on both solar and wind energy forecasting using DL/ML methods. A more detailed discussion about state-of-the-art survey works is presented in Section 2 and Table 2. List of abbreviations is given in Table 1 of this manuscript.

However, the existing studies only review either some particular topics or consider a specific issue. There is no survey/review study that considers a broad involvement of DL methods in smart microgrids in simultaneous ways, e.g., load forecasting and energy generation prediction from photovoltaic and wind turbines. In addition, none of the existing surveys review datasets that were employed for load and energy forecasting. The above motivate us to deliver this study with *the comprehensive review of the state-of-the-art DL-based approaches developed to forecast the power generation from WTs and solar panels, along with the forecasting of load demand of consumers.* Various datasets reported in the literature for the prediction of wind speed/energy and solar irradiance/energy are also presented in this study. Our comprehensive review and analysis makes this manuscript useful for beginners as well as experts working in this domain. This study further helps the reader in tracking datasets used by researchers and developing real-world forecasting applications. Finally, this study can serve as a technical reference for comparison and selection of effective and efficient forecasting strategies.

The rest of the manuscript is organized as follows. Section 2 discusses past surveys in the area of energy management systems (EMSs) and highlights our contributions. Section 3 outlines the methodology of this survey. Section 4 offers a summary of the main DL techniques, while Section 5 describes the use of DL in EMSs and various forecasting models. This section also reviews the datasets that are used to train and test the reviewed DL-based forecasting models. Section 6 investigates the potential issues of the existing DL-based approaches. Finally, the last section concludes the survey.

## 2. Related Work, Motivation, and Contributions

There are a lot of research works published regarding energy management in smart grid/microgrids that present problems and solutions along with future opportunities in the area of smart energy management [27, 28, 38, 39, 40]. Nowadays, researchers are working to explore ML, DL, and artificial intelligence technologies to tackle smart grid challenges. Such techniques provide powerful tools for the planning, modeling, monitoring, fault diagnostics, and fault-tolerant operation of advanced smart grids and renewable energy systems. In order to organize and summarize the current status of DL-based approaches for energy and load forecasting, several review/survey articles have been presented by the research community. In this section, an overview of these articles is disclosed. At the end, this section also highlights how our manuscript differs from past surveys.

In [27], authors present an overview of RESs, such as solar power, wind energy, geothermal energy, hydro energy, etc. Furthermore, the significant role of artificial intelligence (AI) to improve the performance of renewable energy is uncovered in various aspects, including decision, control, optimization, and simulations. At the end, they conclude that the performance of the smart grid and microgrid can be enhanced by employing AI-based techniques.

The study at [28] presents a brief review of the load monitoring (LM) strategies in energy management systems (EMSs). This work categorizes the energy management in two broad types: (i) intrusive LM that refer to distributed sensing, and (ii) non-intrusive LM that belong to single-point sensing. They also analyze intrusive and non-intrusive based LM schemes for energy management in the smart grid. In addition, this study presents an analysis of current literature as well as future prospects in LM for energy management. Some of the future problems regarding LM raised in their work include accurate disaggregation/recognition, non-intrusive LM application in EMS, non-traditional signatures usage to improve the accuracy of non-intrusive LM, and smart meter usage in EMS.

Amasyali *et al.* covered data-driven prediction studies for building energy consumption (BEC) in [29], where they review the prediction steps in detail (i.e., data gathering, data preprocessing, model training, and testing of the trained model). Furthermore, they present machine learning (ML) based algorithms along with their performance in terms of building energy predictions. Performance evaluation criteria of different studies are also disclosed in this work. Finally, gaps are uncovered in the existing research and future directions are provided to the research community in the field of data-driven BEC prediction.

Another review work on data-driven based strategies is presented at [30]. Unlike [29], the research presented at [30] considers data-driven approaches for prediction as well as for the classification of BEC. Their review work demonstrates that a large amount of building energy applications are addressed by data-driven strategies. These applications include load forecasting/prediction, benchmarking for building stocks, guideline making, and power pattern profiling. At the end, this work paves an opportunity for the researchers to explore the potential in small-scale energy minimization via considering consumers' demands.

Voyant *et al.* presented a review in [31], which unfolds the ML-based methodologies to predict the solar irradiance. It is important to note that solar irradiance must be predicted in order to forecast energy generation from the solar panel. This survey presents ML-based prediction models in terms of classification, data preparation, learning (supervised and unsupervised), and accuracy evaluation. Additionally, a comparative analysis is presented to determine the accuracy of various prediction models.

Research work at [32] presents a critical review of smart microgrid energy management methods, problems, and their solutions. As electricity generation in microgrids is intermittent in nature, [32] summarizes the methods/strategies to tackle the volatile and intermittent behavior of the microgrid. A variety of EMSs are discussed in detail, which are developed through different approaches, e.g., classical methods, linear programming, heuristics schemes, evolutionary approaches, swarm optimization, fuzzy logic, neural network, etc. Moreover, communication technologies used in the microgrid are disclosed and comparative analysis among them is performed. Real-time applications of microgrids and future challenges conclude this study.

Authors of [33] have summarized the studies on solar and wind energy forecasting using DL-based prediction techniques. This study states that robustness, reliability, generalization ability, accuracy, sustainability, and precision are the prominent issues when using DL-based algorithms for energy prediction of renewable energy sources. The performance of DL-based algorithms is much better than other computationally intensive prediction techniques when dealing with big datasets; however, the performance is low in case of small datasets. The authors have broadly categorized the DL-based forecasting algorithms into single and hybrid forecasting methods and concluded that hybrid DL techniques provide better forecasting results compared to single DL techniques.

The research contributions presented at [34, 35] survey wind energy and solar power prediction approaches, respectively. In addition, the authors of [34] also discuss applications of ANN in WT system design and fault detection. Fallah *et al.* presented a review work in [36], which explores and summarizes the efforts of researchers in developing load forecasting algorithms. Another study [37] reviews load forecasting methods, while the authors classify the forecasting algorithms in several types based on short-term, very short-term, medium-term, and long-term load forecasting.

**Contributions.** Table 2 summarizes the closely related surveys/reviews on smart microgrids and reveals our survey's novelty. The aforementioned surveys and review works either focus on a specific production application [28, 29, 30, 31, 33, 34, 35, 36, 37, 41] or failed to present a broad image of energy and load forecasting simultaneously. Furthermore, none of the presented works focused on the datasets used for forecasting. Our survey work is therefore intrinsically different due to its data-centered view, along with DL-based application for load and energy generation forecasting in both residential and commercial sectors. This study presents a detailed review of state-of-the-art DL-based approaches, proposed for power forecasting of wind turbines and solar panels as well as energy load forecasting. Moreover, this survey also presents the datasets used to train and test the different DL-based prediction models, enabling future researchers to identify appropriate datasets to use in their works. Eventually, based on our comprehensive survey, this study outlines several challenges that still remain to be addressed and research opportunities for future.

Table 2: Comparative analysis of our work and existing review/survey studies.
Note: PY: published year; BEC: building energy consumption; LF: load forecasting; WSF: wind speed forecasting; SIF: solar irradiance/energy forecasting; DP: datasets presentation

| Ref. | PY | Duration | BEC/LF | WSF | SIF | DP | Review/survey focus |
|---|---|---|---|---|---|---|---|
| [37] | 2014 | 1973-2013 | ✓ | × | × | × | Solutions to power demand forecasting problem; classifies the applied load forecasting methods in various types, e.g., very short-term, short-term, medium-term, and long-term load prediction |
| [27] | 2017 | 1981-2017 | × | ✓ | ✓ | × | Energy generation from renewable energy sources (RESs) and hybrid renewable systems; the role of artificial intelligence in improving the efficiency of RESs |
| [28] | 2017 | 1992-2016 | ✓ | × | × | × | Intrusive and non-intrusive load monitoring techniques to mitigate the power consumption and energy cost of consumers; load forecasting methods are adapted to forecast the energy consumption to balance demand and supply |
| [31] | 2017 | 1996-2016 | × | × | ✓ | × | Solar energy forecasting using ML techniques, namely, supervised and unsupervised learning; data preprocessing and data classification techniques |
| [35] | 2017 | 1991-2016 | × | × | ✓ | × | Current status of solar energy in India; real-time implication of solar plants in various states of India, energy generation from these plants, and their impact on India's economy; solar energy forecasting methods |
| [29] | 2018 | 2002-2017 | ✓ | × | × | × | Building energy consumption prediction focused on the scope of load predictions, the data properties, and the data pre-processing techniques that are exploited in the literature |
| [30] | 2018 | 1986-2017 | ✓ | × | × | × | Building energy analysis and building energy consumption forecasting through data-driven approaches; data classification methods for building energy consumption management |
| [34] | 2018 | 2000-2018 | × | ✓ | × | × | Artificial neural network (ANN) based studies are exploited to forecast wind energy; applications of ANN in WT system design and fault detection |
| [36] | 2018 | 1979-2018 | ✓ | × | × | × | Machine learning techniques for load demand prediction to make sure the reliable operations of the whole power system |
| [33] | 2019 | 2008-2018 | × | ✓ | ✓ | × | Solar and wind energy prediction using DL-based techniques; this study concludes that hybrid methods are more efficient than single DL methods |
| [41] | 2020 | 2002-2019 | × | × | ✓ | × | Limited to long-term solar radiations forecasting using DL-based models |
| Our work | - | Upto 2020 | ✓ | ✓ | ✓ | ✓ | DL-based forecasting methods for both load and energy generation from solar panels and WTs; first-of-its-type datasets presentation while considering load and energy prediction; current challenges and future research directions |

## 3. Survey Methodology

The primary objective of the research methodology is to identify, classify, and review the DL approaches that are employed for load demand or energy forecasting (for solar and wind energy). The main focus during paper selection was on works that were conducted from the period 2015 to 2020. In our comprehensive review, the methodology consists of five primary steps.

1. **Keyword-based search:** As a first step, we have performed a keyword-based search of research studies using Google Scholar. Since Google Scholar ranks articles based on various factors, i.e., authors, publishers, number of citations, and published year, it is selected for searching high-quality articles. Examples of our keywords include data-driven load forecasting, building energy consumption forecasting, load forecasting, wind energy forecasting, wind speed forecasting, solar energy forecasting, solar irradiance forecasting, as well as machine and deep learning for energy management in smart grids.
2. **Screening of papers:** Next, we performed screening of the retrieved research papers that were found through the previous step. The criteria of screening were that the study focuses on power load or energy prediction and employs single DL, single ML, or hybrid DL/ML approaches.
3. **Identifying extra articles:** In this step, we found some extra articles based on the papers that were identified in step 2. Specifically, articles that were cited in the selected papers and articles citing the selected papers were also screened through our criteria described in step 2.
4. **Considering for review:** All the articles selected in steps 2 and 3 are reviewed to disclose their objectives of forecasting, employed/proposed DL/ML methods, forecasting type (long-term, short-term), data source and type, modeling performance, and compared approaches.
5. **Analyzing review results** In the last step, review results are analyzed in order to find superior approaches for load or energy forecasting. Research gaps and future opportunities were also found in this phase.

### *3.1. Evaluation Criteria*

Since the prediction accuracy is a critical factor in selecting any forecasting model, the performance of DL algorithms in this survey paper is compared on the basis of the potential of the proposed approaches to establish the most accurate predictions. Mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) are selected as the three basic evaluation metrics, since they are the most popular metrics used in the reviewed papers.

## 4. Preliminaries on Deep Learning Models

This section discusses the DL-based approaches that are most widely employed in the current literature for energy management and power prediction.

### 4.1. Artificial Neural Network

An artificial neural network (ANN) is constructed based on the working principle of the human nervous system [42]. The ANN is entirely based on a set of neurons, which are the fundamental parts of a neural network (NN) in which communications happen. In Figure 4, a basic ANN architecture is depicted. An input is received and output is generated by neurons based on their internal activation functions [43, 44]. The weights and parameters determining the activation functions are modified by a mechanism known as learning. For ANNs, the key parameters that control learning are the learning rate parameter, the number of hidden layers, and the maximum number of iterations. The input, hidden, and output layers may contain a different number of neurons. Different activation functions, like Sigmoid, Rectified Linear Unit, and Softmax, are used for computation within the ANNs. The advantages of ANN include: information is stored on the entire network so loss of any piece of information does not affect the performance of ANN, fault tolerance, and it has a parallel processing capability [45]. On the contrary, the disadvantages of ANN include: hardware dependency as it requires processors with parallel processing power, lack of interpretability of the network, and the duration of network is unknown [45, 46].

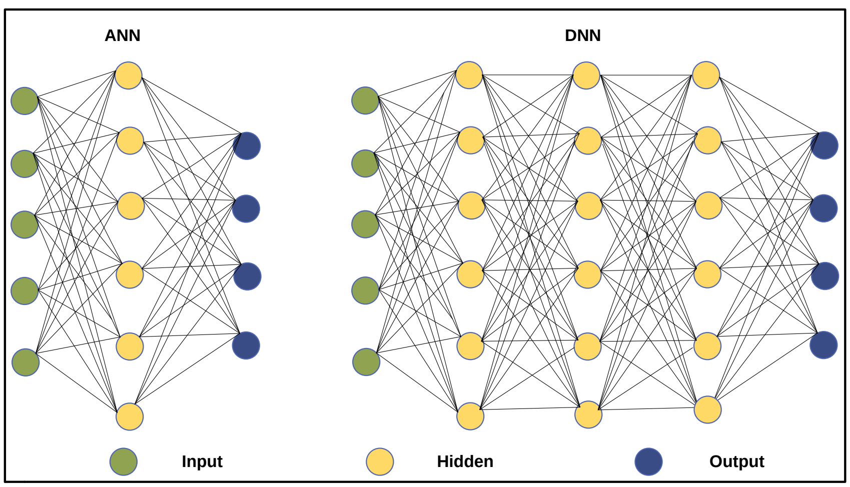


Figure 4: A typical architecture of ANN and DNN

### 4.2. Deep Neural Network

Deep neural network (DNN), also shown in Figure 4, is composed of various hidden layers in addition to the input and output layers [47, 48]. An ANN with two or more hidden layers is called DNN. To generate the output, the DNN investigates the input data using mathematical manipulation. The NN is trained by exploiting the training set resulting typically in the probability calculation for each output. DNNs have similar advantages and disadvantages with ANNs, but since DNNs comprises more layers than ANNs, they often require more training data to attain better results compared to ANN.

### 4.3. Convolutional Neural Network

Convolutional neural network (CNN) is most commonly adopted in energy management, pattern recognition, and visual image processing. It is a revised form of a multilayer perceptron (MLP). The MLP is a fully-connected (FC) layer network, where each neuron is FC with all other neurons of another layer. The completely connected property leads to the problem of over-fitting. Hence, the CNN utilizes different methods for regularization of the results in order to avoid over-fitting issue. CNNs provide an acceptable accuracy especially when dealing with image data; however, large datasets are required for efficient results, which cause high computational cost and the need for high graphical processing units [33].

CNN is also known as a shift variant based on the transition variant [49]. The CNN operates as an NN, and it includes an input, an output, and several hidden-layers [50]. However, unlike ANN, CNN uses a collection of several layers as hidden layers, i.e., convolutional/pooling layers, FC layers, flatten layers, dropout layers, and normalization layers. An activation function hides the input and the output of the hidden layer. In CNN, the linear unit rectifier (Relu) is the most commonly adopted activation function and it includes a back-propagation method to generate more reliable weights.

CNN's convolutional layer is employed to detect patterns and features from the input file. At this layer, filters are applied to the input file and activation maps are generated. The following equation is used to generate the dimension of the activation map [51].

$$\frac{N+2P-F}{S+1}. \tag{1}$$

In the ablove equation, $N$ represents the dimension of input file, $P$ is the padding, $S$ is the stride, and $F$ represents the dimensions of the filter. After the convolutional layer, the spooling layer downscales the data such that processing is simpler, although the actual data remain the same. Through dimensionality reduction, this layer reduces the scale of the input data and minimizes the computational complexity required to process the data. It also extracts the dominant features that help in efficient training of the model. There exist two types of pooling layers: 1) average-pooling layer and 2) max-pooling layer. The average-pooling layer calculates the average values of the data using the kernel and the max-pooling layer uses the maximum values covered by the kernel in the data. Max pooling is commonly used in a CNN. The following equation is used to compute the output file [51].

$$\frac{N-F}{S+1}. \tag{2}$$

The data is passed to the FC layer. In this layer, every neuron of each layer is connected with each neuron of other layer, like MLP. In the FC layer, most of the parameters are occupied, which lead to the over-fitting problem. This problem is resolved by the dropout layer. Using a threshold value starting at 0.5, some of the inputs are removed. The value is often reduced to 0.01 because the increase in dropout leads to losing effective information. The actual weights are then added after training the data at the initial stage. After dropout layer, the data is passed to the flatten layer. It converts the data to a column vector form. The feed-forward NN and the back-propagation methods are then applied at every training step. After the flatten layer, the model is trained enough to distinguish between the dominant features and the low-level features. Finally, the softmax activation function is applied for classification purposes [52].

$$\sigma(Z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}. \tag{3}$$

In this equation, $Z$ represents the input vector of $K$ real numbers. $z$ is an element of input vector $Z$, such that $Z = \{z_1, z_2, z_3, ..., z_K\}$ and $i = \{1, 2, 3, ..., K\}$.

### *4.4. AutoEncoder*

AutoEncoder (AE) is one of the feed-forward NNs, which is employed to copy input neurons to output neurons by passing through single or multiple hidden layers [53]. The AE architecture consists on two key functions, namely, the encoder function $h = f(x)$ and the decoder function $\hat{x} = g(h)$. The mathematical presentation of AE is expressed as:

$$\hat{x} = g(Wx + b) \tag{4}$$

where $x$ and $W$ represent the input and weights, respectively. An activation function is represented by $g$ that can be a rectified or sigmoid function. The term $b$ introduces bias in Equation 4. Figure 5 presents a typical architecture of AE, which shows input, output, and hidden layers. One advantage of AE is that it employs filters to fit a dataset in a better way, which can improve the performance of AE. Consequently, it takes additional training time, which is a main disadvantage of AE [33]

### *4.5. Deep Belief Network*

A deep belief network (DBN) [55] contains multiple restricted boltzmann machines (RBMs) that are considered primary elements of the DBN [56]. The RBM is an updated form of a boltzmann machine [57] by adding node connections. The RBM contains two key layers, namely visible and hidden layers. Moreover, DBN uses both supervised and unsupervised learning. In particular, unsupervised learning is used in the pre-training phase, whereas supervised learning is exploited in the fine-tuning phase. Selection of appropriate initial parameters, weights, and bias is performed by unsupervised learning using independent variables. In this way, the pre-training stage rebuilds training samples by tuning variables to enhance likelihood estimation. Supervised learning further tunes the weights and bias on the basis of initial parameters that are given

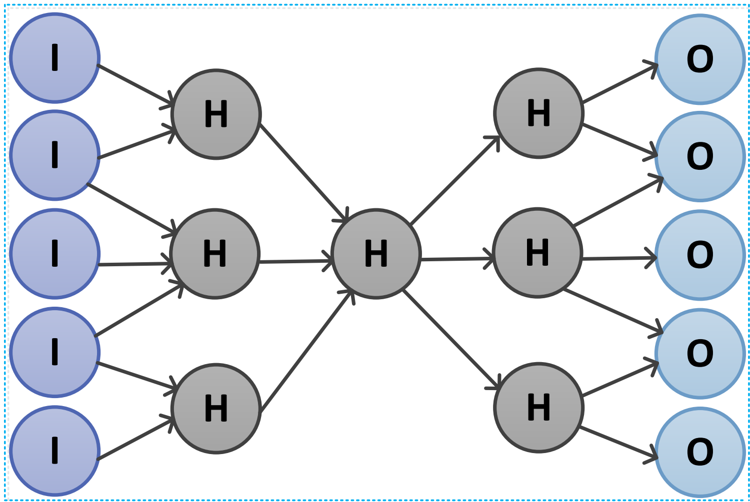


Figure 5: A typical architecture of AE [54]

by the pre-training stage. Overall, DBN networks train differently compared to DNNs and ANNs as they use energy-based training functions to propagate data throughout the unsupervised training mode. Based on a past critical analysis [33], DBN is highly capable to deal with similar image data; however, it has high computational complexity. Figure 6 presents a DBN model with $L$ number of layers, where the input and output layers are presented on the left and right sides, respectively.

### *4.6. Recurrent Neural Network*

For the processing of sequential data, a special form of NN, proposed by the research community, is known as recurrent neural network (RNN). The CNNs typically provide training independently to each sample; however, this form of independent training is not enough, particularly for sound, text, image, and time-related data. RNN solves this problem and it takes input sequentially. It includes feedback connections in the hidden layer units, as opposed to other feed-forward NNs. RNN will, therefore, undergo temporal processing and learn sequentially. In addition, the RNN exploits a hidden layer as a memory in order to store sequential information, unlike other NNs. In addition, the RNN employes the same parameters $(U, V, W)$ for each layer, as opposed to conventional DNNs that use different parameters for each layer (see Figure 7). This figure unfolds RNN into a full network. Moreover, in RNN calculations, $x_t$ shows input at time $t$, while $s_t$ and $o_t$ represent the hidden and output state at time $t$, respectively. The key advantages of RNN are that it remembers complete information based on time, it can deal with sequential data efficiently, and it

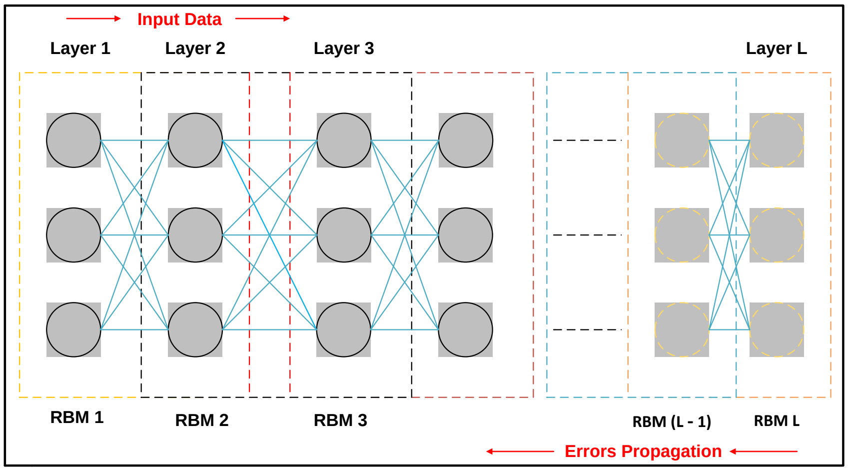


Figure 6: A typical architecture of DBN [57, 58]

provides acceptable accuracy while predicting based on time-series data. However, long-range learning is difficult with RNNs because of exploding or gradient vanishing problems [59, 60]

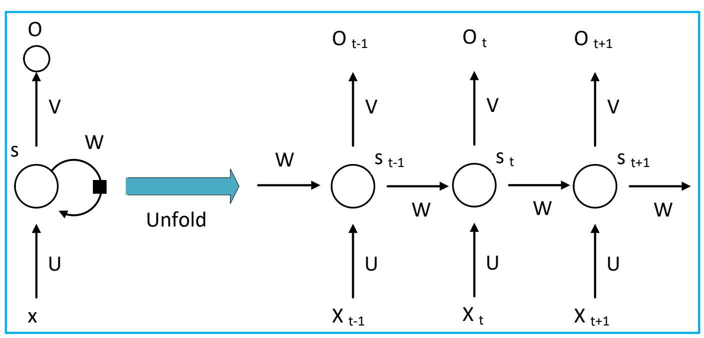


Figure 7: A complex RNN architecture [61]

### *4.7. Long Short Term Memory*

RNNs were developed to process sequential data and are able to establish a temporal correlation of current circumstances with previous information. For instance, RNNs make decision at time step $t$ on the bases of $t-1$ and $t$. This type of RNN characteristics makes it able to efficiently solve the load forecasting and energy generation prediction of solar/wind energy sources. Moreover, RNNs are trained by back-propagation through time [62]. But, long-range learning is difficult with RNNs because of exploding or gradient vanishing problems [59, 60].

To solve the aforementioned problems in RNNs, Hochreiter *et al.* introduced long short term memory (LSTM) by including a memory cell [63], which was further enhanced by adding an extra forget gate [64]. LSTM is considered one of the most efficient NN architectures for time-series forecasting and modeling. Conventional NNs learn the correspondence among input and output from a static perspective. However, information is lost when time-series data is independently trained as input and output of NNs. The RNN makes a link between each pair of "input-output", as presented in Figure 8, where $x$ denotes input data, $y$ shows output data, and $h$ presents the hidden states. The terms $W_{hx}$, $W_{yh}$, and $W_{hh}$ denote the matrices of weights, which show the relationship between $h$ and $x$, $y$ and $h$, and $h$ and $h$, respectively. Furthermore, unlike simple RNN, the LSTM has two hidden states $h_t$ and $c_t$ to capture the long-term dependencies. Hidden states $h_t$ and $c_t$ are designed to keep the short-term and long-term information, respectively. The hidden state $c$ has an additional mechanism that helps LSTM to strategically forget unnecessary information. LSTM has introduced three control gates to keep the information for the long-term, as presented in Figure 9. The LSTM is capable to solve vanishing gradient problems and make shorter the pre-processing of data [33]. The main drawbacks of LSTMs are: they need huge amount of resources to deal with big data, training process is very difficult, and they need high memory-bandwidth because of the linear layers present in each cell, which makes them inefficient [64] .

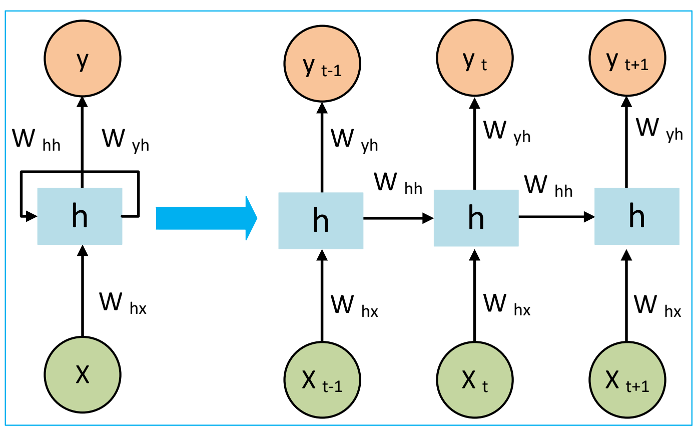


Figure 8: A typical structure of LSTM [65]

As shown in Figure 9, LSTM has three gates: forget gate (denoted by $f_t$), input gate (denoted by $i_t$), and output gate (denoted by $o_t$). The forget gate ($f_t$) determines which information is kept from the last state and utilizes a sigmoid activation function. The second gate is the input gate ($i_t$) that determines which information should be considered as input for the current state. The last gate is known as output gate ($o_t$) and calculates which information is treated as output while using the tanh activation function.

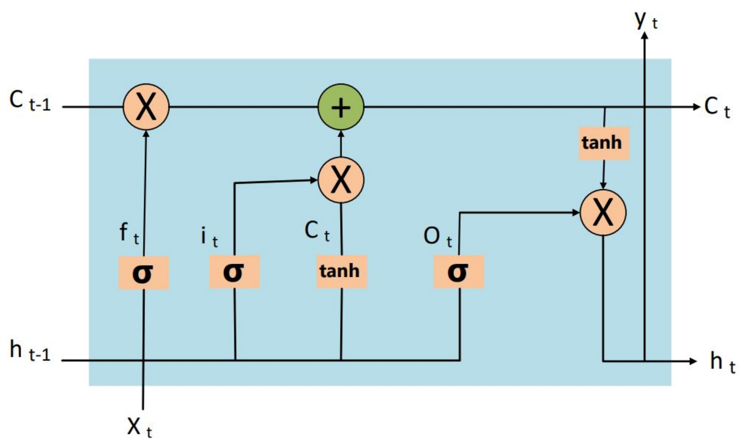


Figure 9: Inner structure of LSTM [65]

## 5. Deep Learning in Energy Management Systems

Energy management is essential to efficiently integrate RESs and energy storage systems (ESSs) in power systems [66]. Energy management is the process of observing, planning, and controlling the operations of energy production and consumption units. With proper energy management, energy consumers can reduce their electricity bills and utility companies can reduce peak creations [1]. Furthermore, an optimal utilization of RESs can be achieved by implementing an efficient energy management strategy, for instance, by shifting all the load and ESS charging to solar energy in day time instead purchasing from utility [1]. On the contrary, energy management is also necessary for enhancing the life of ESSs [67, 68]. Charging and discharging of storage systems up to specific limit can also enhance the life of batteries. For example, according to [69], for achieving higher efficiency of ESS, minimum and maximum storage levels of ESS are 10% and 90%, respectively.

An accurate energy prediction is necessary to attain effective energy management because of the intermittent power production from RESs. Researchers have developed various forecasting methods for load forecasting and renewable energy sources on the bases of their properties, such as wind speed, solar iradiance, temperature, etc. The forecasting of wind energy, solar energy, and load using DL follows three main steps, as presented in Figure 10. First, the data pre-processing step is performed to clean and normalize the input data, as well as to split it into training, validation, and testing datasets. Next, model training is performed for creating an appropriate and validated prediction model. Finally, the forecasting is performed using the trained model and often visualized. In the next section, we uncover the works that use DL-based techniques to forecast wind energy, solar energy, and load demand.

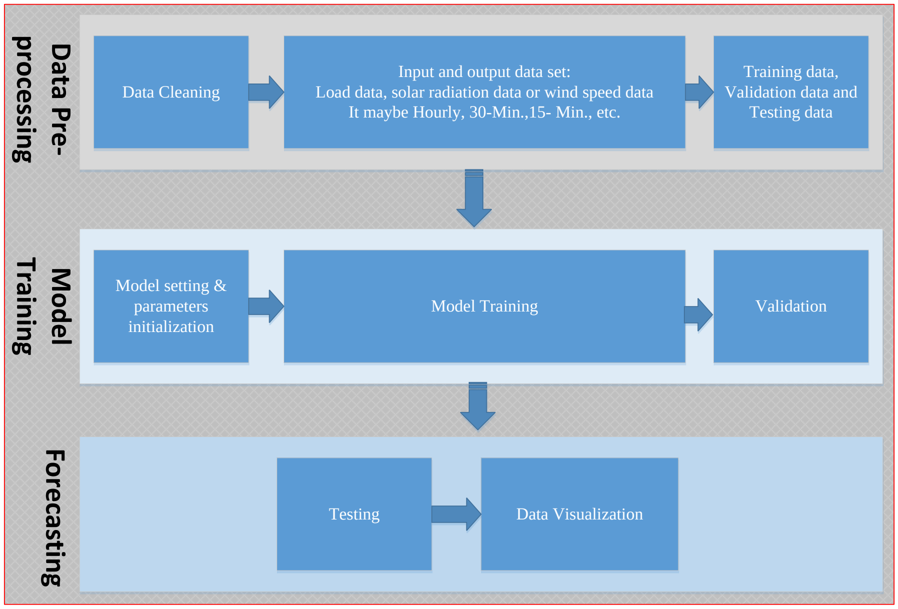


Figure 10: A generic flowchart of wind energy, solar energy, and load forecasting using DL-based methods

### 5.1. Wind Energy Forecasting

In the last decade, noticeable attention has been given to wind energy owing to a cleaner source of energy. WTs are considered the lowest carbon emitters [56]. However, the uncertainty and fluctuations (due to weather conditions) of wind energy generation bring severe issues that hinder the economic operations of the power system [18]. Hence, accurate forecasting of wind energy is of vital importance for the efficient operations of Energy Management Systems (EMSs) in the residential sector. Without accurate and reliable prediction of wind energy, maximum benefit from EMS cannot be achieved. Therefore, research community has spent much effort on developing wind energy forecasting methods, which are elaborated in detail in this section. Table 3 describes various datasets used in wind speed and energy forecasting, whereas Table 4 summarizes the efforts of the research community regarding forecasting of wind energy and speed. The majority of the wind speed datasets are collected in Asia, span two to three years, and contain fine-grained data (recording step ranges from 5 minutes to 1 hour) of wind speed, wind direction, temperature, humidity, and pressure among others. Similarly, the developed methods focus on forecasting wind speed and wind power generation with a time horizon ranging from 5 minutes to 1 hour. The key difference among the forecasting methods are in using different variations or combinations of the DL models discussed in Section 4, leading to different forecast accuracy results as listed in Table 4.

Table 3: Description of datasets used for wind speed and energy forecasting

| Ref. | Dataset Origin | Description | Total Time Period | Recording Step |
|---|---|---|---|---|
| [56] | Weather stations of Matsu and Kinmen islands, Taiwan | Authors consider 11 attributes of weather that are taken from [71]: wind speed, temperature, dew point temperature, humidity, sea pressure, station pressure, wind direction, max gust, the direction of max gust, precipitation hours, precipitation amount, and sunshine hours. (`http://eservice.cwb.gov.tw/HistoryDataQuery/index.jsp`) | Training data: January 1, 2017 to December 31, 2017; Testing data: January 1, 2018 to January 14, 2018 | Hourly |
| [70] | Wind tower of National renewable energy laboratory (NREL), National wind technology center (NWTC) | The tower is located in Boulder, Colorado, at latitude of 39.91°N, longitude of 105.23°W, and elevation of 1855 m [72]. | Training data: 2015 to 2016; Testing data: 2017 | 15 minutes |
| [73] | GEFCOM2012-WF: Publicly available dataset of seven wind farms over a 3-year period | The meteorological dataset attributes are the forecasts of zonal and meridional components of surface winds, wind speed, and wind direction [74]. | Training data: July 01, 2009 to December 31, 2010; Testing data: January 01, 2011 to June 28, 2012 | Hourly |
| [75] | Real-time data from wind farms in Bornholm Island, Denmark | The wind farms energy generation capacity is 30 MW | June 01-July 31, 2012 and November 01-December 31, 2012 (Training data 60%; Testing data 40%) | 10 minutes |
| [76] | WIND: Publicly available data from Dallas, Texas, USA | The dataset is collected through wind integration national dataset (WIND) Toolkit from 711 wind sites with total rated wind power capacity 9,987 MW [77] | January 01, 2007 to December 31, 2012 | 5 minutes |
| [78] | Wind farms in China | The dataset contain 700 samples of wind speed series data, where 1-600 samples are employed for training and testing, other 601-700 samples are exploited [79] | - | Hourly |
| [80] | Wind data from Inner Mongolia, China | The wind farm is located in the monsoon region and the annually average wind speed is 3.7 (m/s) | 10-minutes case: November 23, 2012 to November 28, 2012; hourly case: April 01, 2013 to April 30, 2013 (Training data 70%; Testing data 30%) | 10 minutes and hourly |
| [81] | Wind farms in Shandong Province, China | Monthly wind speed data; data from day 1st to 25th are used for training and data from the remaining days of each month are used for testing | January 01, 2011 to December 31, 2011 | 15 minutes |
| [82] | Wind speed data from NREL | Wind speed and energy generation data from 32,043 WTs [83] | January 01, 2004 to December 31, 2006 | Hourly |
| [84] | Wind speed data from Xinjiang Province, China | Four different datasets, each containing 750 time-series values. First 500 data values are used for training and the remaining 250 values are employed for testing | - | Hourly |
| [85] | Wind farms in Xinjiang, China | Four different datasets, each containing 700 time-series values. First 600 data values are used for training and remaining 100 values are employed for testing | - | 10 minutes |

Table 4: Summary of wind speed and energy forecasting approaches

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [56] | DBNGA | Seasonal autoregressive integrated moving average and least squares support vector regression with genetic algorithm | Taiwan | Hourly | For forecasting of wind speed, seasonal autoregressive integrated moving average (SARIMA) and least squares support vector regression for time series with genetic algorithms (LSSVRTSGA) are used. For genetic algorithm, 40 genes are used in form of binary numbers. Population size was set to 10. | The developed DBNGA shows effectiveness to compared methods in terms of forecast accuracy. [MAPE of DBNGA: 12.00 and MAPE of compared method: 13.95. RMSE of DBNGA: 0.621 and RMSE of compared method: 1.326] |
| [70] | WTD-RNN-ANFIS | WTD-ANN, WTD-SVM, WTD-RNN, ANN, SVM, RNN | - | 15 minutes | Proposed forecasting model comprises of WTD (to decompose and smooth historical time series), RNN ensemble (six RNNs with dissimilar architectures and parameter) and ANFIS (utilized as the top layer of the ensemble model). | It is verified from results that the proposed WTD-RNN-ANFIS model is superior and feasible for probabilistic wind speed prediction. [RMSE of the proposed method: 0.9678 and RMSE of compared method: 1.0045. MAE of WTD-RNN-ANFIS: 0.6516 and MAE of compared method: 0.6989] |
| [73] | BRT | Conventional unbiased forecasting methods | - | Hourly | Proposed model is based on the cost-oriented boosted regression tree method (COBRT). | The developed BRT method outperforms counterparts. [RMSE of proposed BRT: 0.1389 and RMSE of compared method: 0.1734] |
| [75] | DQR | Persistence, BELM-Normal, BELM-Beta, RBFNN | Bornholm, Denmark | 10 minutes | Proposed model is based on statistical description of the wind speed characteristics given in the frequency domain to simulate time series of output power | This work achieves higher accuracy than well-known benchmark methods of wind power forecasting. [The newly developed method outperforms by 25% and 20% the Persistence method and the RBFNN, respectively.] |
| [76] | p-WPRF, GGMM distribution | GMM | Dallas, Texas, USA | 5 minutes | Wind power forecasting is done based on probabilistic modeling, which is then used to calculate historical forecasting errors by using a continuous generalized Gaussian mixture model (GGMM). | The developed p-WPRF shows supremacy in terms of accurate and efficient wind ramp forecasting. [The performance of proposed method is improved by 21% over counterpart] |
| [78] | EWT-LSTM-Elman | ARIMA, LSTM, Elman, EWT, GRNN | China | Hourly | Proposed model consists of EWT (to decompose the raw wind speed data into several sub-layers), LSTM network (to predict the low-frequency sub-layer) and Elman neural network (to predict the high-frequency sub-layers) | The EWT-LSTM-Elman shows efficacy over counterparts. [MAPE of proposed model: 10.93 and MAPE of compared model 24.95] |

Continued on next page

**Table 4 – Continued from previous page**

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [80] | EnsemLSTM | ARIMA, SVR, ANN, KNN, GBRT | China | 10 minutes and hourly | Proposed EnsemLSTM model has six diverse LSTMs, where LSTM1 contains 1 hidden layer and 50 neurons in the hidden layer, LSTM2 has 1 hidden layer and 100 neurons in the hidden layer, LSTM3 comrpises of 1 hidden layer and 150 neurons in the hidden layer, LSTM4 is made of 2 hidden layers and [50,50] neurons in the hidden layers, LSTM5 has 2 hidden layers and [50,100] neurons in the hidden layers, and LSTM6 comprises of 2 hidden layers and [50,150] neurons in the hidden layers | The proposed EnsemLSTM has higher performance in terms of wind speed forecasting accuracy. [MAE of proposed method: 1.1410 and MAE of compared model: 1.3753. RMSE of EnsemLSTM: 1.5335 and RMSE of compared model: 1.8337] |
| [81] | Hybrid of WT and CNN | SVM and backpropagation | China | 15 minutes | The proposed hybrid approach is based on WT, CNN and ensemble technique. The weights and biases of deep CNN are trained by the back propagation rule applying stochastic gradient descent | The proposed method efficiently tackles the uncertainties, while forecasting of wind energy in all seasons and show competency in forecasting accuracy. [PINC99% for proposed method: -0.78 and PINC99% for compared method: -3.11 ] |
| [82] | LSTM-EFG | LSTM, SVR, KNN | United States | Hourly | Euclidean distance, K-Means, Spectral Clustering, Agglomerative Clustering and Birch methods are used for feature extraction. SVR, KNN, LSTM, LSTM-EFG are used as forecasting methods. | The LSTM-EFG with spectral clustering demonstrates a higher accuracy than the benchmarks. [The proposed LSTM-EFG model shows 13.10% higher performance than LSTM, 16.84% higher than KNN, and 18.30% higher than SVR.] |
| [84] | WPD-Boost-ENN-WPF | Two forecasting strategies (Recursive and MIMO) and two boosting algorithms (AdaBoost.MRT and LPBoost) | Xinjiang, China | Hourly | Mother wavelet=db3, level of decomposition=3. AdaBoost.MRT: number of example = 0.9*N (number of instances), iterations = 20, threshold = random 0 to 1. | The developed hybrid method shows effectiveness in terms of MAE over compared boosting and forecasting strategies. [MAE of the proposed method: 0.9461 and MAE of compared method 1.7492] |
| [86] | LSTMDE-HELM | ARIMA, ANN, SVR, ELM, LSTM | Inner Mongolia, China | 10 minutes and hourly | ARIMA: (p,d,q)=(2,0,1). ANN: 1 hidden layer, 10 neurons. SVR: C =18.8, $\sigma^2$=0.36. ELM: 1 hidden layer and 20 neurons. ELM = 1 hidden layer and 100 neurons, LSTMDE-HELM: LSTM1 has 1 hidden layers and 89 neurons, LSTM2 has 1 hidden layers and 135 neurons. | The proposed hybrid algorithm exploits evolutionary algorithm DE to optimize the hidden layers of LSTM and in this way, the performance of the hybrid method is enhanced over simple LSTM and other counterparts in terms of forecast accuracy. [RMSE of proposed LSTMDE-HELM: 1.5956 and RMSE of compared model: 1.6635] |

Continued on next page

**Table 4 – Continued from previous page**

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [85] | SSA-EMD-CNNSVM | SVM, CNNSVM, EMD-BP, EMD-RBF, EMD-Elman | Xinjiang, China | 10 minutes | The proposed model is based on Singular Spectrum Analysis (SAA), Empirical Mode Decomposition (EMD) and Convolutional Support Vector Machine (CNNSSVM) | The developed SSA-EMD-CNNSVM shows efficacy for 1-step to 3-step wind speed forecasting over benchmarks. [The average performance promotion in terms of MAPE, MAE, and RMSE is 42.85%, 39.21%, and 39.25%, respectively] |

The authors of [56] propose a wind speed forecasting method for efficient energy management, where they exploit DL, namely deep belief network (DBN) along with genetic algorithm (GA). GA is used for determining the DBN's parameters. They use real-time weather data from various regions of Taiwan. Both multivariate regression and time series datasets are exploited to forecast wind speed. They performed simulations to validate the effectiveness of their developed DBN and GA based forecasting model. Results demonstrate the productiveness of their developed model over counterparts. Cheng *et al.* also developed a wind energy forecasting model in a residential area [70]. The developed model consists of an RNN, an adaptive neuro-fuzzy inference system (ANFIS), and wavelet threshold denoising (WTD). WTD is used to smooth the wind speed series to capture variation trends and RNN is trained on datasets that are provided by the WTD layer. Eventually, ANFIS is considered the top layer of the ensemble model and it performs final wind speed prediction, which in turn can be used for predicting wind power generation. The developed method is then evaluated on 1-hour-ahead wind speed prediction and results affirm its superiority over counterparts.

The research presented at [73] has proposed a wind speed prediction model under cost-oriented loss functions, where a cost-oriented boosted regression tree (BRT) method is developed to formulate the efficient forecasting of wind speed. Various case studies with real-time datasets are presented to verify the productivity of the presented method and a comparison with conventional unbiased forecasting methods is performed. Comparative results are evident of the effectiveness of the proposed scheme. Another study proposed a direct quantile regression (DQR) method for wind power prediction that combines the quantile regression and extreme learning machine [75]. According to [87], wind energy shows higher volatility in intra-hour resolution (i.e., 10-minutes, 15-minutes, etc.) as compared to hourly wind power. Therefore, the work [75] considers multi-step probabilistic prediction of 10-minutes wind energy. A comparative study is also presented in this work, where various well-known methods of wind energy forecasting, such as RBFNN, Smart Persistence, BELM-Normal, and BELM-Beta, are compared against the performance of newly developed forecasting method. Results show the efficacy of the newly developed 10-minutes wind power forecasting method.

The authors of [76] proposed a data-driven probabilistic wind energy ramp forecasting (p-WPRF) technique that is based on a huge amount of simulated scenarios. A publicly available dataset from [88] (containing data for a location near Dallas, Texas, USA) is exploited to affirm the effectiveness of the proposed ramp forecasting model. The authors performed simulation studies to show the efficacy of p-WPRF model and results affirm the productiveness of this work with higher accuracy and reliability.

Liu *et al.* [78] proposed a hybrid approach known as EWT-LSTM-Elman for wind speed prediction that is the combination of empirical wavelet transformation (EWT) and two types of RNNs. The EWT is exploited to decompose the raw wind speed data into multiple sub-layers and the LSTM neural network is adopted to forecast the low-frequency wind speed sub-layers. At the end, an Elman neural network (ENN) is utilized to predict the high-frequency sub-layers. Furthermore, to measure the performance of the newly proposed EWT-LSTM-Elman forecasting algorithm, eleven different forecasting algorithms are considered as benchmarks. Experimental results validate the developed algorithm in terms of high precision wind speed forecasting.

Another study at [80] presents a hybrid method for time-series wind energy forecasting, which combines the non-linear learning ensemble of DL, support vector regression machine (SVRM), LSTM, and external optimization (EO) technique. The newly developed algorithm is named as EnsemLSTM. First, unlike a single DL approach, a cluster of LSTMs is adopted to exploit and explore time-series data of wind speed. Then, non-linear regression is exploited to aggregate the forecasting of LSTMs. The top-layer of the proposed model contains SVRM. EO and final ensemble forecasting of wind speed is given by fine-tuning of the top-layer. The datasets are used from the wind farms of Inner Mongolia to perform experiments to affirm the performance of the newly developed hybrid method. In addition, the work [80] considers two case studies: forecasting of wind speed considering (i) hourly time intervals and (ii) 10-minute time intervals. A comparative study also has been taken into account, where five forecasting algorithms are employed as benchmarks, i.e., GBRT, KNN, ANN, SVR, and ARIMA. It is observed from simulations that developed EnsemLSTM has higher performance than the compared algorithms.

The work presented in [81] also tackles the wind forecasting problem and proposed a DL-based ensemble approach, where an advance point prediction model is developed based on the wavelet transform (WT) and CNN. WT decomposes wind speed data into various frequencies, while non-linear features of various frequencies, learned by CNN, are employed to enhance the forecast accuracy. To check the performance of the newly developed DL-based ensemble-based method, real-time datasets containing uncertainties are used from China. Further, the authors of [81] also implemented their proposed method for wind energy forecasting during the four seasons, i.e., summer, winter, spring, and autumn. Results from simulations demonstrated that the proposed method efficiently tackles the uncertainties and provides satisfactory performance.

Ruiguo *et al.* developed an LSTM-enhanced forget-gate (LSTM-EFG) network for wind energy forecasting [82]. The developed method replaces the *tanh* activation function with the *softsign* activation function, excludes the input-gate of traditional LSTM, and subtracts the output of the forget-gate in the way of the all-1 matrix. It utilizes the results as the input of the data update. In this way, the convergence speed is enhanced by the newly developed model LSTM-EFG. In addition, this model also exploits the feature extraction method that is hybridized with cluster methods in order to select the data having similar characteristics. Extensive experiments have been performed in the study and results show that the LSTM-EFG achieves minimum MSE value compared to well-established methods such as LSTM, SVR, and KNN.

The study presented in [84] proposes a hybrid algorithm to forecast the big multi-step wind speed. ENN, wavelet packet decomposition (WPD), wavelet packet filter (WPF), and boosting algorithms are exploited to enhance the forecast accuracy. Furthermore, this study utilizes four time-series datasets to affirm the performance of the newly proposed WPD-Boost-ENN-WPF algorithm. Experimental results show the efficacy of the proposed forecasting algorithm over counterparts in terms of big multi-step wind speed prediction.

In [86], Hu *et al.* present a hybrid algorithm, namely LSTMDE-HELM, for long-term wind speed forecasting, where they perform hybridization by combining the best features of hysteretic extreme learning machine (HELM), LSTM, non-linear combined mechanism, and differential evolution (DE) algorithm. The working of the developed hybrid method is as follows: firstly, a biological neural system property named hysteresis in the activation function of ELM is used to enhance its efficiency. Afterward, DE is adopted to optimize the number of hidden layers in the LSTM to maintain a balance between learning performance and complexity of the LSTM (as there is no clear mechanism in order to set the hidden layers of LSTM). Finally, the prediction results of each predictor in the developed hybrid algorithm are aggregated by the non-linear combined mechanism, which is the combination of LSTM and DE. Furthermore, extensive experiments are performed to affirm the effectiveness of the newly developed hybrid forecasting method. For this purpose, they have exploited real-time wind speed data of Inner Mongolia and China. A comparative study has been performed to show the efficacy of the LSTMDE-HELM model. Results indicate the higher performance over the compared algorithms, namely, LSTM, ELM, SVR, ANN, and ARIMA, in terms of forecast accuracy.

Another work [85] presents a hybrid algorithm, termed SSA-EMD-CNNSVM, which combines the best features of EMD, singular spectrum analysis (SSA), and CNNSVM for multi-step wind speed forecasting. In the newly developed hybrid algorithm, the SSA is employed to mitigate the noise and it extracts trends in the actual wind speed data. The EMD is employed to explore the fluctuation features from wind data and decompose time-series wind speed to multiple sub-layers. CNNSVM is utilized to forecast the wind speed sub-layers. Furthermore, to examine the forecasting efficiency of the newly developed hybrid algorithm, several benchmarks are taken into account and experiments are performed. According to experimental results, the proposed SSA-EMD-CNNSVM forecasting method has satisfactory performance over counterparts for 1-step-3-step wind speed forecasting with the MAPE = 42.85%, MAE = 39.21%, and RMSE = 39.25% average performance promotion.

The use of DNNs appears to be one of the most commonly used wind energy prediction techniques. When DNNs are combined with optimization techniques for tuning the large number of network parameters, the accuracy of the overall system can be greatly improved. Hence, a significant growth can be seen in research on the aforementioned hybrid techniques, which aim to complement the predictive stage with the optimization of parameter sets to allow higher degrees of precision. Under various conditions, such as limited data access or lack of weather stations near to the wind farms being tested, these hybrid models have made it possible to refine conventional statistical methods based on historical data and to offer solutions to climate variability issues for real wind farms. Figures 11 and 12 demonstrate this observation, where hybrid approaches outperform other conventional approaches. To make the comparison fair, we show the error values as reported in the original papers over that same two datasets, one taken from wind farms in China [79] and one from NREL National Wind Technology Center (NWTC), Boulder, Colorado [72]. An important factor to note is the evaluation process, of which the RMSE or the MAE are the most common means of evaluating the accuracy of the models in place.

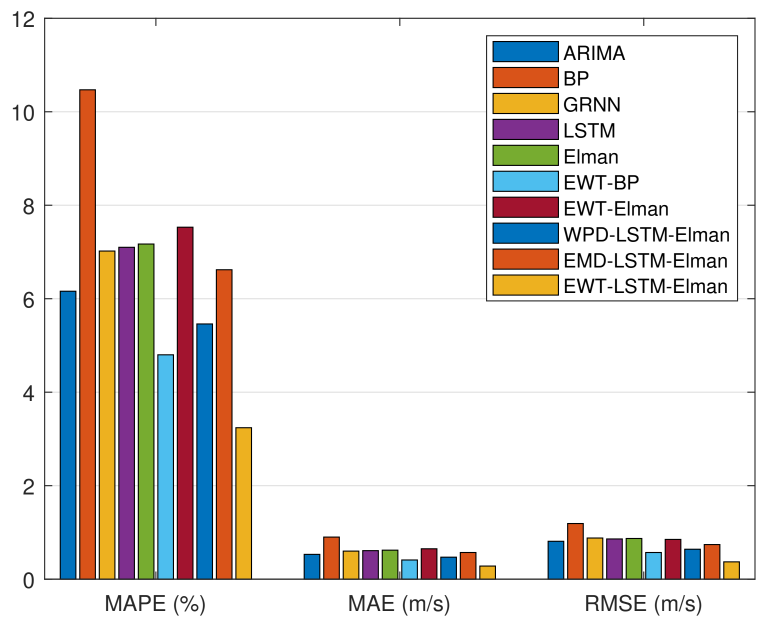


Figure 11: Comparison of different wind forecasting methods that were implemented on the same dataset, taken from wind farms in China [79]

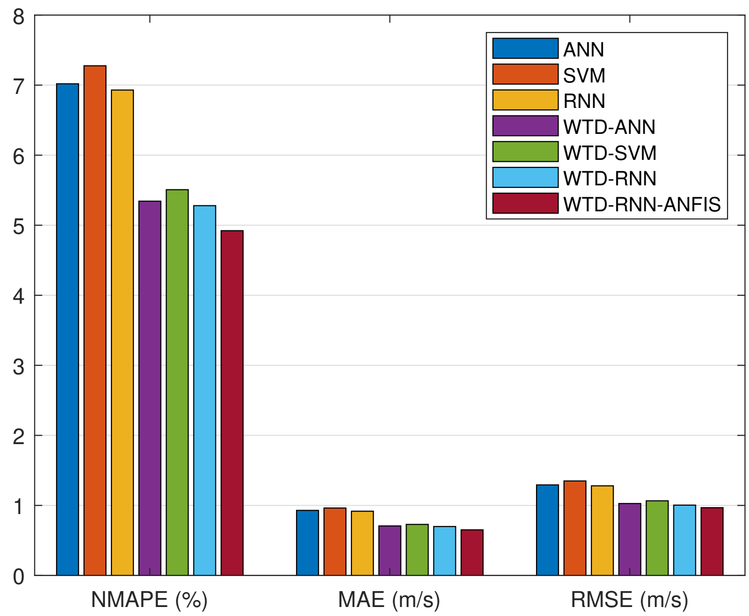


Figure 12: Comparison of different wind energy forecasting methods that were implemented on the same dataset, taken from NREL National Wind Technology Center (NWTC), Boulder, Colorado [72]

### *5.2. Solar Energy Forecasting*

Electricity demand is rising day by day due to the growing number of the population, which also generates a massive amount of greenhouse gases. Hence, people and organizations are moving towards sustainable sources of energy such as solar panels. However, because of the intermittent nature of solar power, the forecasting of solar energy needs to be accurate. Solar panel power generation may be forecasted on a 1-hour, 2-hour, 10-hour, or 1-day basis. State-of-the-art solar irradiance and energy forecasting studies have been included in this section that are critically analyzed in terms of methodologies, pros and cons. Table 5 describes various datasets used in solar irradiance and energy forecasting, whereas Table 6 summarizes the efforts of the research community regarding forecasting of solar irradiance and energy. Unlike the wind speed datasets, the solar energy ones cover several locations worldwide (e.g., Europe, US, Asia) and primarily record hourly data spanning several months to years. Similarly, the proposed approaches forecast solar power generation in hourly steps, typically up to 24 hours ahead. The majority of methods use a hybrid approach combining DNN, RNN, or LSTM as these methods work well in identifying temporal correlations among the data with varying degrees of success rates (see Table 5).

Gensler *et al.* proposed a solar energy forecasting approach by employing DL in [54]. Twenty-one PV panels are considered for generating energy, and day-ahead forecasting is made. In their work, a MLP [89], a type of feed-forward ANN, is employed as a base architecture consisting of several layers (recall Section 4). The results of the MLP forecast are compared with other models, such as ANN and physical models.

The study presented in [90] proposes a statistical approach for short-term Spatio-temporal forecasting of solar power. This paper forecasts power for a very short-time period (1-6 hrs). For this study, distributed power plants are exploited along with their Spatio-temporal dependencies in order to improve prediction accuracy. In addition, their model's computational complexity is low, making it simple to use, and is considered a suitable solution for industrial applications. The simulation results support (in terms of accuracy) the proposed model over current models. The work [91] designs an RNN-based prediction model for solar irradiance. Authors have used a version of RNN known as a gated recurrent unit (GRU) and LSTM [92]. Extensive simulations are carried out to check the efficiency of the proposed model in terms of precise solar irradiance prediction. It is validated through results that the GRU and LSTM are better suited to predict time-series irradiance as compared to simple RNN.

The research at [93] presents a solar forecasting method using numerical weather prediction (NWP) and CNNs. A Gaussian process is employed to transfer the incoming values of PV power into the main grid and train the CNN. The developed CNN can also map outputs of $6 \times 6$ to $31 \times 31$ based on the transposed conversion operation. Experiments are performed to validate the developed CNN model and adequate accuracy is achieved in comparison with benchmark models, i.e., ridge regression, persistent method, and FC NN.

In [94], Subhadip *et al.* present a deep NN, known as SolarisNet, for solar energy prediction. They employ limited weather parameters, i.e., maximum temperature, minimum temperature, and hourly solar radiation. Simulations were conducted to test the performance of the developed SolarisNet model, and data is used from India's meteorological department. Findings from simulations present a higher performance of the proposed model relative to ANN [95, 96], SVR [97], and Gaussian process regression [98].

Another solar power prediction approach is proposed, employing Deep RNN (DRNN), in [99]. The proposed method uses real-time data from the National Resources of Canada [100]. Results from simulations are compared with current forecasting approaches that show the efficacy of developed method. The authors of [101] propose a new hybrid adaptive learning model (ALM) for solar intensity prediction over the short and long term. A time-varying multiple linear model is built to deal with the linear and dynamic properties of data. A GA back-propagation NN (GABPNN) is then implemented in order to learn the non-linear relationship of data. The proposed hybrid ALM is capable of capturing the linear, nonlinear, and temporal relationship in data. Results from simulations confirm that the developed forecasting model shows efficiency over several benchmarks in both long and short-term solar intensity forecasting.

Abdel *et al.* designed a novel PV energy forecasting model in [102] employing deep LSTM-RNN. They also consider the temporal changes during prediction model building. This study analyzes five various LSTM models with different architectures in order to check their effectiveness. They consider several commonly used prediction models for comparison purposes, including ANNs, multiple linear regression (MLR), and bagged regression tree (BRT). Another research develops a high-precision deep CNN model called 'SolarNet' for solar radiation prediction [103]. Experiments are carried out to verify the performance of the proposed forecasting model. From the results, it is confirmed that the SolarNet model shows efficiency, in terms of accurate prediction, over counterparts.

The research proposed in [104] constructs two forecasting methods, based on DNNs, to forecast daily solar and wind energy. The Kaggle dataset is used for the research and model preparation. Additionally, this research proposes DNN ensembles in order to enhance single DNN predictions by reducing variance and is illustrated by experiments showing the randomness in DNN training elements resulting in efficient and stable DNN ensembles. Another forecasting method for wind and solar energy is provided in [105]. The proposed method takes into account the gradient boosting algorithm and feature engineering technique that extracts the knowledge from the NWP grid. They also present a comparative analysis of the proposed method and the approach, which has only one NWP point for a particular location. The simulation results are evident that the forecast accuracy for solar and wind energy is increased (in terms of MAPE) by 16.09% and 12.85%, respectively. Another solar power forecasting method, based on ML, is built in [106]. They also conduct a comparative study with multiple regression approaches to demonstrate their technique's effectiveness. It is affirmed from simulation results that their proposed method forecasts with 27% higher accuracy than the current forecasting approaches. The study presented in [107] developed a DL-based hybrid algorithm for short-term solar irradiance prediction. The hybrid method combines GRU network with an attention mechanism, where an Inception NN (INN) is developed for feature extraction from original data. The proposed inception-based hybrid GRU approach is tested on the dataset taken from [108], and results show higher performance over single LSTM and GRU in terms of forecast accuracy.

Each prediction model has its own pros and cons in predicting solar irradiance and PV power generation; thus, it is difficult to determine which is the best among all the models. However, the following findings are suggested from the studies examined in this paper. For a single model, many studies demonstrate that LSTM has higher efficiency over RNN under all circumstances because the LSTM has intrinsic memory to resolve vanishing gradient issues arising in the RNN. In addition, multiple studies examined reveal that the hybrid models perform better than the standalone ones in the prediction of solar irradiance. This is evident in the Figures 13 and 14, which compare existing approaches using the same datasets, as reported in the original papers. However, in terms of computational or training time, GRU exhibits more efficiency compared to LSTM. Overall, taking into account training time and estimation accuracy, the

GRU model yields a satisfactory result for the forecasting of PV power and solar irradiance.

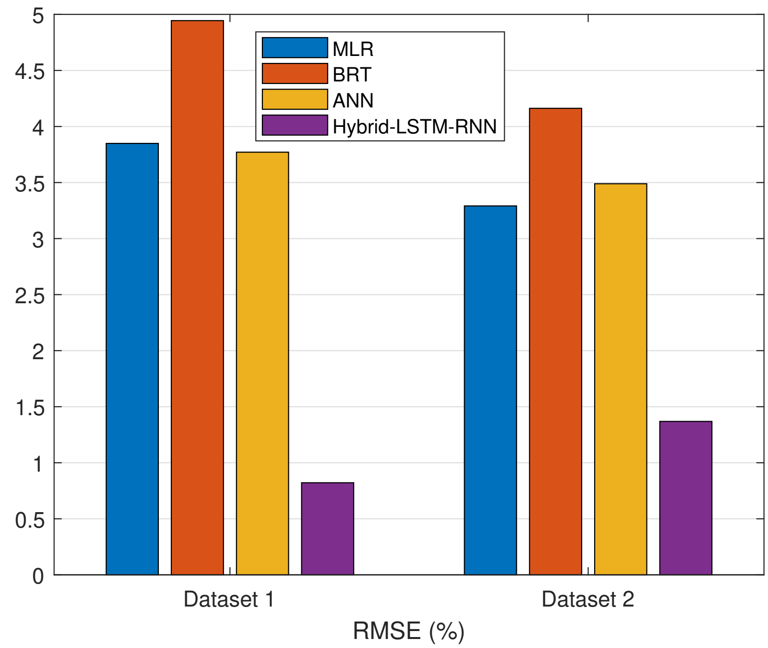


Figure 13: Comparison of different solar energy forecasting methods that were implemented on two datasets (taken from Solar farms in Aswan and Cairo, Egypt [102])

### *5.3. Electric Load and Consumption Forecasting*

Load forecasting for buildings/homes, industrial areas, and the commercial sector plays a significant role in the modern era of the smart grid. An accurate load/demand forecasting for energy consumers is a challenging task because of their stochastic behavior regarding electricity consumption. However, a lot of research studies have focused to tackle this issue and this section critically analyses these studies along with their benefits and drawbacks. Table 7 describes various datasets used for forecasting electric load and consumption, whereas various studies on the forecasting of electricity load and electricity consumption are summarized in Table 8. The majority of datasets contain hourly load data spanning several months along with time information (e.g., month, day of week) and temperature, which are considered strong predictors of electric load for both commercial and residential consumers. Based on this data, the surveyed approaches employ a wide range of DL algorithms to make hourly forecasts for the next few hours to few days, and offer different degrees of forecasting performance as listed in Table 8.

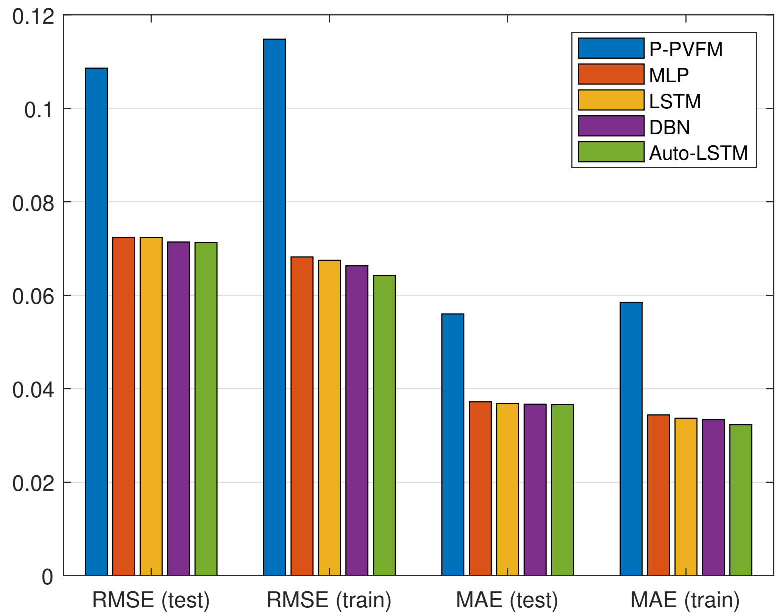


Figure 14: Comparison of different solar energy forecasting methods that were implemented on the dataset (taken from [109])

Table 5: Description of datasets used for solar irradiance and energy forecasting

| **Ref.** | **Dataset Origin** | **Description** | **Total Time Period** | **Recording Step** |
|---|---|---|---|---|
| [54] | German Solar Farm, Germany | Data from 21 photovoltaic facilities, with nominal power ranging between 100kW and 8500kW [109] | Training data of 490 days; Validation data of 250 days; Testing data of 250 days | Hourly |
| [90] | Two datasets from mid-west and south region of France | $1^{st}$ dataset comes from 9 power plants with peak power ranging between 45kWc and 5MWc; $2^{nd}$ dataset comes from 185 power plants with peak power ranging between 32kWc and 58kWc | Training data of 15 months; Testing data of 5-months | 15 minutes |
| [91] | Publicly available global horizontal solar radiation data | The dataset contains data for 10-years measured by a French meteorological organization | Jan 1998 to Dec 2007 | Hourly |
| [93] | American meteorological society | The dataset published within the context of a contest [112] | Training data: Jan 01, 1994 to Dec 31, 2006; Testing data: Jan 01, 2007 to Dec 31, 2007 | Hourly |
| [94] | Kalyani meteorological site, Bengal, India | No additional information provided | Training data: 80%; Testing data: 20% | Hourly |

Continued on next page

**Table 5 – Continued from previous page**

| Ref. | Dataset Origin | Description | Total Time Period | Recording Step |
|---|---|---|---|---|
| [99] | Solar farms in Canada | The data consists of global horizontal and global tilted irradiance along with the corresponding time [113] | Training data: 70%; Validation data: 10%; Testing data: 20% | Hourly |
| [101] | UMASS Trace Repository | Solar intensity measured in watts/$m^2$; Dataset attributes used: temperature, wind speed, humidity, precipitation, and dew point [114] | Training data: Jan 01, 2015 to Dec 31, 2016; Testing data: Jan 01, 2017 to Feb 28, 2017 | 5 minutes |
| [102] | Solar farms in Aswan and Cairo, Egypt | The data locations have subtropical desert low-latitude arid hot climate | Training data: 70%; Testing data: 30% | Hourly |
| [103] | Solar sites in Tainan, Taiwan | Data collected through computer monitoring system of PV sites; radiometer is used to capture at least one record/minute | Training data: Jan 01, 2015 to April 31, 2015; Testing data: May 01, 2015 to June 31, 2015 | Hourly |
| [104] | Publicly available Kaggle dataset | Contains solar radiation of 98 stations of Oklahoma's Mesonet network [115] | Training data: Jan 01, 1994 to Dec 31, 2005; Validation data: Jan 01, 2006 to Dec 31, 2006; Testing data: Jan 01, 2007 to Dec 31, 2007 | Hourly |
| [105] | Solar farms in Porto, Portugal | No additional information available | April 28, 2013 to June 28, 2016 | Hourly |
| [106] | US National weather service (NWS) | Solar radiation data of small city-size regions throughout the US, with several metrics per hour [116] | Jan 01, 2010 to Oct 31, 2010 | Hourly |
| [107] | National Renewable Energy Laboratory, USA | Solar radiation data of various places in Nevada, USA [108] | Jan 01, 2001 to Dec 31, 2005 | 30 minutes |

Table 6: Summary of solar irradiance and energy forecasting approaches

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [54] | Auto-LSTM | ANN, LSTM, MLP, DBN, and DNN | Germany | Hourly | MLP consists of multiple FC layers of neurons and a back propagation algorithm. For Auto-LSTM, $n = 2$ previous samples are used to predict a new value. Furthermore, tanh activation function is used except for output layer, where a Rectified Linear Unit (ReLU) activation function is used. | The developed hybrid approach demonstrates higher forecast accuracy; however, the efficiency of the DBN is closer to the proposed method. [RMSE of newly developed approach: 0.0713, compared approach: 0.0714] |
| [90] | Spatio-Temporal model | Autoregressive and random forest | France | 15 minutes | This work applies a spatio-temporal model to the stationarized series and addresses the problem of high dimension data by using Lasso regularization. | The developed statistical forecasting method indicates high performance over counterparts in terms of computation complexity and accuracy. [The performance improvement of nRMSE is 20% over counterparts] |

Continued on next page

**Table 6 – Continued from previous page**

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [91] | LSTM | Naive, RNN, and GRU | France | Hourly | A special Recurrent Neural Network variations Long Short-Term Memories and Gated Recurrent Unit models are used. | The LSTM-based prediction technique reveals superiority over comparative approaches. [nRMSE of newly developed approach: 0.2115, compared approach: 0.2198] |
| [93] | Gaussian process regression based CNN | NN, and Ridge regression, Persistence | Oklahoma, USA | Hourly | Input of the network contains the values of the 87 features on a 6 by 6 grid, and the output of the network is the forecasts on a 31 by 31 grid. Three types convolution operations are considered: regular convolution with 3×3 filters, transposed convolution with varied sizes of filters, and regular convolution with one 1×1 filter. | The newly developed method shows efficacy in terms of minimum MAE [MAE of the proposed method: 212642 and compared method: 4399526] |
| [94] | DNN namely 'SolarisNet' | Gaussian process regression, SVR, and ANN | India | Hourly | A 6-layer deep neural network is considered. Input layer consists of 1x3 neurons and direct connection activation function. Non-linearity augmentation layer has 2x2x3 neurons and tan sigmoid function. Dimensionality embedding layer has 1x2 neurons and log sigmoid activation function is used. Network is trained by Levenberg-Marquardt (LM) back propagation technique | The SolarisNet prediction model performs efficiently in terms of high accuracy. [SolarisNet RMSE: 1.7661 and compared model RMSE: 2.7930] |
| [99] | Deep RNN | LSTM, SVR, and FNN | Canada | Hourly | A deep recurrent neural network is considered for prediction of the solar irradiance and LSTM neuron was introduced to solve the exploding gradient problem. | The results from simulations confirm that the proposed deep RNN outperforms counterparts; performance is measured as RMSE. [The RMSE of proposed model: 0.068 and compared method: 0.18] |
| [101] | ALHM: hybrid of GABPNN and multiple linear model | SVM and ANN | - | Hourly and 5 minutes | An adaptive learning hybrid model using integration of the time-varying multiple linear model and a genetic algorithm back propagation three-layer neural network is used. | Experiments validate that the hybrid model can accurately predict the energy produced from solar panels. [The MAPE of ALHM: 13.68 and compared method: 20.39] |
| [102] | Hybrid LSTM-RNN | multiple linear regression, bagged regression trees, and ANN | Aswan and Cairo, Egypt | Hourly | Considered LSTM network comprises a one-input visible layer, a hidden layer with four LSTM blocks (neurons), and an output layer that gives the predicted power. Sigmoid activation function is used for the LSTM blocks and We network was trained for 20, 50, and 100 epochs with a batch size of 1. | The proposed hybrid model provides a very small error rate as opposed to compared methods. [The RMSE of LSTM-RNN: 82.15 and compared method: 384.89] |

Continued on next page

**Table 6 – Continued from previous page**

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [103] | Deep CNN | LSTM, MLP, decision tree, SVM, random forest | Tainan, Taiwan | Hourly | Proposed network comprises of three 1D convolution layers and three pooling layers. Sigmoid activation function is used. However, the rectified linear unit (ReLU) is employed as an activation function of the convolution and output layers to reduce the chance of gradient vanishing. | The developed deep CNN reveals effectiveness in terms of minimum error rate. [The average MAE of deep CNN: 112.26 and compared method: 143.27] |
| [104] | DNN ensemble model | SVR | Oklahoma, USA | Hourly | Architecture comprises of two initial convolutional layers, two FC layers and a final linear readout layer. Non-symmetric ReLUs in the hidden layer and Glorot–Bengio weight initialization heuristic are used to dilate the Glorot–Bengio uniform intervals by a factor of 1.5. | The newly proposed DNN employs minibatch preparation, weight initialization, and dropout regularization to intrude independent randomness; simulation results support the robustness and higher accuracy of the DNN ensemble model. [The average MAE of DNN ensemble: 209.09 and compared method: 222.52] |
| [105] | Gradient boosting trees | Quantile Regression Forests | Porto, Portugal | Hourly | Proposed model is based on the gradient boosting trees algorithm | First work to propose a method to use domain knowledge to extract features from NWP grid; this knowledge can increase the forecast accuracy over existing methods. [The newly developed methods indicates forecast improvement 16.09% over current methods] |
| [106] | SVM-RBF | Linear regression and past-predicts future models | USA | Hourly | Models are based on multiple regression techniques for generating prediction models, including linear least squares and support vector machines using multiple kernel functions | The SVM-RBF forecasting model denotes higher accuracy. [The accuracy is enhanced using the proposed model by 27% over compared methods] |
| [107] | Inception-based hybrid GRU | LSTM and GRU | USA | 5, 10, 20, and 30 minutes | The proposed hybrid model uses INN for feature extraction and RNN for model training. Then, a two-layer GRU structure predicts solar irradiance and an attention mechanism deals with GRU output by assigning various weights. Finally, hidden neurons are discarded by dropout layer and FC NN is used to show results. | The proposed hybrid inception-based GRU shows higher accuracy over counterparts. [The MAPE and MAE of proposed method: 5.80 and 26.49, LSTM: 6.01 and 26.95, and GRU: 6.13 and 27.28] |

The technique proposed in [110] developed a short-term load forecasting method by exploiting a DBN. The hourly load data of North Macedonia from 2008 to 2014 is used for the modeling. The authors compare the obtained results not only with the actual hourly data of North Macedonia but also with another neural network, namely MLP. Results demonstrate efficacy in terms of reduced MAPE. Another work [111] also exploits a DBN model for power load forecasting on the basis of historical data. It considers real-time time-series historical load data of South Africa for demand forecasting. In addition, weather parameters, like wind speed, temperature, etc. are also taken into account to check their impacts and to improve the forecast accuracy of the proposed model. Simulations have been performed to validate the model, while the temperature impact on forecast error is also analyzed. Results show the effectiveness of the developed model.

Robinson *et al.* [117] developed a power demand forecasting model for commercial consumers using ML techniques. They developed a gradient boosting regression (GBR) based model to forecast the power demands of commercial buildings. In addition, they perform experiments on various datasets that are obtained from different locations of the United States. First, they exploit the data of New York city and the same forecasting model is implemented on the data of Atlanta city. Results validate the performance of the newly developed model. Another paper [118] considers a load forecasting problem in residential areas as well as in commercial buildings. A deep RNN is employed for medium to long term energy consumption forecasting. The datasets from commercial buildings of Salt Lake city, USA are exploited to perform simulations and a 3-layer MLP forecasting model is implemented to examine the efficiency of the developed forecasting model. Simulation results show the effectiveness of the proposed deep RNN based model over MLP for load demand prediction of commercial buildings. However, 3-layer MLP shows efficacy in the forecasting of the residential load.

The research work presented in [119] tackles the load forecasting problem of residential areas. Usually, volatility and uncertainty in household demand forecasting are considered the key issues. Traditional techniques are used to solve these issues in various ways such as customer classification, load aggregation, and spectral analysis. However, this paper adopts a mechanism to learn directly from uncertainties and develops a new forecasting algorithm, termed pooling-based deep RNN (PDRNN). It utilizes the load profiles of several consumers as a pool of inputs, enabling the model to address the over-fitting problem. Furthermore, it is claimed that it is the first attempt to develop a DL application for residential consumers. Extensive simulations have been performed and data of 920 smart-metered consumers from Ireland are exploited. Additionally, to check the performance of the newly developed model, authors have performed a comparison with other benchmarks, i.e., ARIMA, SVR, and classical deep-RNN. A comparative study shows the efficacy of the PDRNN forecasting model.

Another research work [120] also adapts DL based methods for load forecasting. Specifically, a hybrid forecasting method is developed by combining the best features of CNN and K-means clustering. They used a large dataset obtained from the power grid, which is clustered into subsets using the K-means algorithm, and the obtained subsets are used to train the CNN. The authors also performed simulations for both seasons (summer and winter) to validate the productiveness of the proposed hybrid model and a comparative study is also taken into account, where several forecasting algorithms employing linear regression, linear regression+L-means, SVR, and CNN are considered. Results affirm the effectiveness of their hybrid CNN-K-means forecasting algorithm in terms of higher accuracy.

Xueheng *et al.* proposed a hybrid power demand forecasting algorithm that combines EMD and DBN [121]. To forecast the power demand, first, the historical load demand series are decomposed into multiple intrinsic mode functions (IMFs) and then a DBN containing two RBMs is opted to model each IMF. Eventually, the prediction results of all IMFs are combined by either weighted or unbiased summation to attain an aggregated output for power demand. Furthermore, they performed experiments to show the legitimacy of their proposed forecasting method by employing the datasets from the Australian Energy Market Operator (AEMO) [122]. They utilized nine other forecasting methods as benchmarks for comparative purpose, i.e., persistence, SVR, ANN, DBN, random forest, EDBN, EMD-SVR, EMD-ANN, and EMD-RF.

The study presented in [123] proposes a load and price forecasting method to balance electricity load demand and supply. For this purpose, a hybrid algorithm is developed on the bases of a multi-stage forecast engine (MFE) and dual-tree complex wavelet transform (DCWT). First, the signals enter the wavelet transform and then are filtered by a novel feature selection. Subsequently, the signals are forecasted by MSFE in 3 steps and then an intelligent algorithm is opted to enhance the forecast accuracy. Eventually, an improved fusion algorithm collects the outputs of MSFE. To check the effectiveness of their proposed forecasting method, extensive simulations have been performed using the datasets from the energy department of Australia and England. Various forecasting methods, like ARIMA, SVR, RBFNN, WT+RBFNN are also employed for comparative study.

Gabriel *et al.* also tackled the load forecasting problem in [124] and proposed a load forecasting framework that built a wavenet ensemble for short term power demand forecasting. Firstly, data are transformed and normalized to remove trends, then an optimal time window is constructed and a subset of features is selected. Subsequently, the bootstrapping, cross-validation, simple mean, and median algorithms are employed for the ensemble aggregation of the wavenet learners. Finally, forecasted values are realized via a one-step-ahead strategy. The authors have considered different forecasting methods, such as MLP, single wavenet, and regression tree, for experiments and compared them with the proposed algorithm. In addition, they used real-time datasets from Global Energy Forecasting Competition, Italy to perform experiments.

Another energy demand forecasting problem for the residential community is taken into account by Mujeeb *et al.* in [125]. They proposed a hybrid forecasting algorithm, namely deep LSTM (DLSTM) that combines the best qualities of LSTM and DNN. The proposed DLSTM uses the automatic feature learning mechanism from DNN and all other forecasting steps are performed by LSTM. To evaluate the newly proposed algorithm, they perform experiments by using the datasets of New York city. They forecast day-ahead and week-ahead power demand. Furthermore, MAPE and RMSE are computed to check the performance of proposed and benchmark algorithms.

The authors of [126] also consider the load forecasting problem and propose a solution for residential areas. An adaptive circular conditional expectation (ACCE) technique is developed based on circular analysis to define the subresiduals operation schedules. Next, an adaptive linear model (LM) is opted to forecast the residual component demand by exploiting the results of the ACCE process at each time step. Finally, the forecast performance is evaluated as the normalized mean absolute error (NMAE) and a comparison is performed with auto-regressive model (AR) [127] and auto-regressive with exogenous input (ARX) [128] forecasting model to validate the ACCE method.

The Inception Time forecasting model, an ensemble of deep CNN, can be used for time-series forecasting. The fundamental building block of the inception model is known as an inception module, which comprises of bottleneck, convolutional, max pooling, and depth concatenation layers. The concept of inception module is adopted from image processing in which network architectures like AlexNet, GoogleNet, etc., are used for image classification or recognition. Recurrent Inspection CNN (RICNN) model is proposed for short-term electricity load forecasting in [66]. In RICNN model, RNN is combined with 1-dimensional CNN network to learn the spatial and temporal representations of electricity load. The RNN learns the long-term and short-term temporal dependencies present in the electricity load time-series data. Then, the CNN learns the low-level (spatially adjacent local) and high-level (valleys and peaks) features of the electricity load time-series. The electricity consumption dataset of 3 large electricity distribution complexes of Korea electric power corporation (KEPCO) is utilized for building the RICNN model. This model outperforms the benchmark model MLP in terms of MAPE.

Ahmad *et al.* proposes a short-term load forecasting (STLF) method for industrial areas [129]. The primary objective of this work is to enhance forecast accuracy along with high convergence speed. For this purpose, the authors proposed a hybrid ANN that employees the mutual information (MI) for features selection, while enhanced differential evolution (DE) is exploited for error minimization. Consequently, execution time was reduced by 52.38% and 95.5% accuracy was recorded in simulation results, as compared to bi-level forecast strategy.

Many of these deep learning-based forecasting algorithms have successfully addressed the forecasting analysis and have outperformed the forecasting challenges of ML and NNs. There are a variety of issues related to the forecast study of the form of load, period, temperature, seasons, customer behavior, and holidays. For example, the prediction of household load use for individuals varies depending on the extent of the use of appliances. However, most of the studies in this paper show that hybrid approaches outperform the standalone or conventional models in terms of performance and accuracy. Indicatively, we have shown such comparison of some standalone models with a hybrid approach for load forecasting in Figure 15.

## 6. Current Challenges and Future Research Directions

DL-based approaches have been considered beneficial means of enhancing the efficiency of smart microgrids to provide potential strategic solutions for precise power generation forecasting from RESs and load demand forecasting. In this

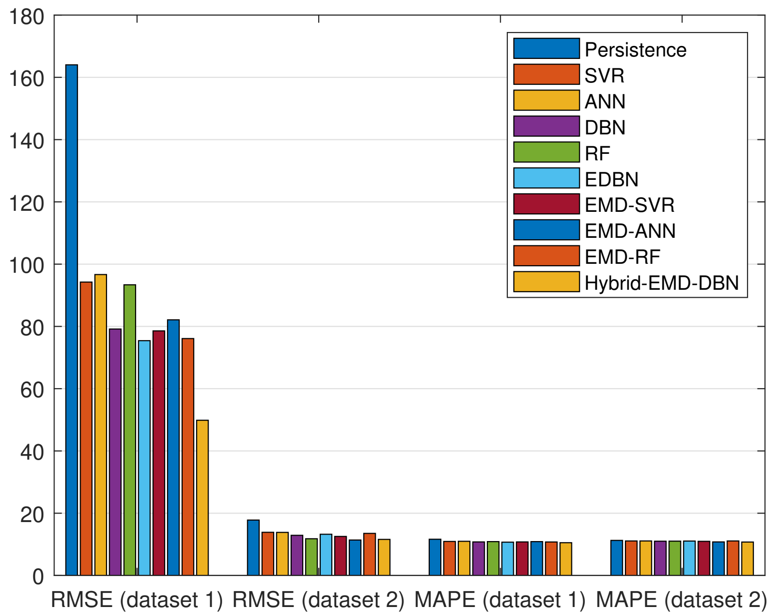


Figure 15: Comparison of different load forecasting methods that were implemented on two datasets (taken from Australian Energy Market Operator (AEMO) [130])

Table 7: Description of datasets used for energy consumption and load forecasting

| **Ref.** | **Dataset Origin** | **Description** | **Total Time Period** | **Recording Step** |
|---|---|---|---|---|
| [110] | Electricity Transmission System Operator (MEPSO) of North Macedonia | Dataset consists of hourly load demand along with hourly temperature [131] | 2008-2014 | Hourly |
| [111] | South Africa | Energy data is taken from a substation of South African utility 88/11 kV, 80 MVA [132]; temperature data is also collected separately | August 2012 to May 2016 | Hourly |
| [117] | Commercial buildings in New York and Atlanta, United States | Data collected from New York City Mayor's Office of Sustainability based on Local Law 84 Data Disclosures and contains 13223 rows of data [133] | 2015 | Hourly |
| [119] | Energy regulation commission of Ireland | Dataset contains records of 5000 consumers (having smart meters); current study used data of 920 smart metered consumers [134] | July 01, 2009 to Dec 31, 2010 | 30 minutes |
| [121] | Australian energy market operator, Australia | Dataset includes data from 5 cities: NSW, Tasmania, Queensland, South-Australia, and Victoria; the study used 4 months in 2013, one from each season [130] | 2013; Testing: first 3 weeks of each month; Training: last week of each month | Hourly |

Continued on next page

**Table 7 – Continued from previous page**

| Ref. | Dataset Origin | Description | Total Time Period | Recording Step |
|---|---|---|---|---|
| [118] | The University of Utah, Salk Lake City, UT, USA | Data collected for a public safety building, which is a net-zero, LEED platinum, having area of 175,000 Sq ft. | Training data: May 18, 2015 to May 18, 2016; Testing data: May 19, 2016 to Aug 08, 2016 | Hourly |
| [124] | Italy | The two datasets are publicly available and taken from Italy and Global Energy Forecasting Competition; both datasets consists of 8760 records for one year | Jan 01, 2015 to Dec 31, 2015 | Hourly |
| [125] | New England (dataset ISO-NE) and New York (dataset NYISO) | ISO-NE contains data for 8 years and NYISO presents data for 13 years; both datasets are publicly available | ISO-NE dataset: Jan 2011 to Mar 2018; NYISO dataset: Jan 2006 to Sept 2018 | Hourly |
| [126] | Single house located in Montreal | Hourly load data combined with hourly outside temperature | One year | Hourly |
| [66] | Three different areas of South Kora | Dataset includes real-time data collected by sensors from three different areas of South Korea, i.e., Incheon, Gwangju, and Shihwa | 503 days for Incheon, 517 days for Gwangju, and 530 days for Shihwa | 30 minutes |

Table 8: Summary of energy consumption and load forecasting approaches

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [110] | DBN | MLP | North Macedonia | Hourly | A multi-layer feed forward perceptron (MLP) is considered and a back propagation algorithm is used for training. Each pair of layers of the neural network is pre-trained by using restricted Boltzmann machine (RBM). | The authors validate the performance of the developed model through MAPE and their model shows supremacy over counterparts. [MAPE of the proposed model is minimized by 8.6% over counterparts] |
| [111] | DBN | - | South Africa | Hourly | First, unsupervised learning is used and, to reduce the set of features, DBN has been trained by contrastive divergence. In the second step, supervised training is used to train an appended layer to pre-trained network. | They did not compare their model with any benchmark; however, the obtained errors were around 4% |
| [117] | GBR | Linear regression, ET regressor, RF regressor | New York City and Atlanta, USA | Hourly | Proposed model is based on gradient boosting regression method. | Experiments show that the developed model attained higher accuracy; however, they performed experiments only on datasets of commercial buildings. [MAE of the proposed method: 0.24 and MAE of compared method: 0.45] |
| [119] | PDRNN | ARIMA, SVR, classical deep-RNN | Ireland | 30 minutes | Proposed method uses load profiles pooling and then deep-RNN. | Results from experiments demonstrate the effectiveness of the proposed model over counterparts in terms of RMSE to ARIMA, SVR, and classical deep-RNN by 19.5%, 13.1%, and 6.5%, respectively |

Continued on next page

**Table 8 – Continued from previous page**

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [129] | AFC-STLF | Bilevel and MI-based ANN | USA | Hourly | Forecasting module consists of ANN with 24 ANs, 1 hidden layer having 5 ANs. | This work achieved high forecast accuracy and execution time is reduced by 52.38% to compared approaches |
| [120] | CNN + K-means | Linear regression, linear regression+L-means, SVR, CNN | USA | Hourly | Data was divided into training and testing subsets by using K-Means clustering. The proposed CNN consists of Filter: 1*3, Pooling: 1*2, Layer Number: 2, and Parameter estimation algorithm: AdomOptimizer. | The developed model shows efficacy as higher accuracy. [MAPE of proposed model: 3.055 and MAPE of benchmark method: 3.95] |
| [121] | EMD + DBN Hybrid | EMD-ANN, EMD-RF, EMD-SVR, EDBN, random forest, DBN, ANN, SVR, and Persistence | Australia | Hourly | ANN and EMD-ANN: size of NN is determined by the size of input vector. DBN: 2 RBMs are stacked for pre-training with the size of [100 100]. Iterations for back propagation = 500. RF and EMD based RF: decision trees = 500 | Experimental analysis reveals EMD-based hybrid method outperforms the corresponding single structure models for time-series load prediction. [MAPE of the proposed method: 0.9187 and MAPE of compared method: 1.6580. RMSE of the proposed method: 118.49 and RMSE of base method: 181.61] |
| [118] | Deep RNN | 3-layer MLP | Salt Lake City, USA | Hourly | Layer 1 is provided with input at one hour resolution. Layer 2 is the first LSTM layer and acts as an encoder. Layer 3 is used as decoder. Layer 4 is used to concatenate the output of layer 3 with the original input vector. Finally, layers 5 and 6 comprise a multi-layered perceptron neural network. | The proposed model shows efficiency only for commercial load forecasting; the compared algorithm MLP shows efficiency over deep RNN for residential load forecasting. [MAPE of proposed mode: 0.77 and MAPE of compared model: 0.948] |
| [123] | DCWT and MFE | ARIMA, SVR, RBFNN, WT+RBFNN | Australia, England | Hourly | The proposed multistage hybrid forecast model consists of ANN, RBFNN, and SVM, where ANN is based on the back-propagation NN and RBFNN comprises of three layers. | The proposed hybrid algorithm shows efficacy in term of forecast accuracy. [NMAPE of the proposed approach: 7.63 and NMAPE of compared method: 10.43; NRMSE of newly developed model: 6.73 and NRMSE of benchmark method: 9.54] |
| [124] | Enhanced wavenet ensemble | MLP, single wavenet, regression tree | Italy | Hourly | Cross-validation like, Bootstrapping, constructive selection, inputs decimation, median, mode, simple mean, and stacked generalization algorithms are used for the ensemble aggregation of wavenet learners. After ensembling, one-step-ahead forecasting strategy is used for predictions. | Experimental analysis shows productiveness of the wavenet ensemble-based load forecasting method. [The performance of the proposed method is increased by 13% over counterparts] |
| [125] | Deep LSTM | LSTM, DNN, ELM, ANN, Nonlinear Autoregressive network with exogenous variables (NARX) | New York City, USA | Hourly | DLSTM comprises five layers: 1 input layer, 2 LSTM layers, 1 FC layer, and the regression output layer. The number of hidden units in LSTM layer 1 and 2 is 250 and 200 respectively. | They exploited real-time data and their proposed DLSTM shows efficacy in terms of convergence rate and highest accuracy. [MAE of deep LSTM: 2.9 and MAE of benchmark method: 9.7; NRMSE of the proposed method: 0.087 and MAPE of compared method: 0.2] |

Continued on next page

Table 8 – Continued from previous page

| Ref. | Method(s) | Compared Method(s) | Location | Horizon | Model Description | Outcome/observation(s) |
|---|---|---|---|---|---|---|
| [126] | Adaptive ACCE | AR, ARX | Canada | Hourly | Proposed model is based on an Adaptive Circular Conditional Expectation (ACCE) method. | The newly developed algorithm shows effectiveness in terms of higher accuracy. The performance is measured in NMAE. [The newly developed method improves forecasting accuracy by 23% over benchmark models] |
| [66] | RICNN | MLP | South Korea | 30 minutes | In the proposed inception-based hybrid model, a CNN captures local significant relationship and RNN handles a variable length of sequential data. Then, an inception module with four 1-D convolution of various sizes is included between the last LSTM layer and first FC layer to make forecasting on the basis of past information as well as predicted future information. | The newly inception-based RICNN approach demonstrates higher performance in terms of higher accuracy. The performance is measured in MAPE. [The MAPE of RICNN for 7 days training is: 7.832 and compared method: 11.260. The MAPE of RICNN for 3 days training is: 8.086 and compared method: 10.002] |

section, this study outlines the research challenges/directions of DL methods applied for precise wind, solar, and power demand forecasting.

### *6.1. Serving DL with a Huge Amount of Data*

A superior performance can be achieved by DL approaches only when huge and high quality data is available [135]. The quantity and quality of historical data have significant importance during training of large and complex architecture, as DL models have numerous parameters to be learned and configured. This challenge still remains open in EMSs, because unfortunately, unlike other research domains like image processing, natural language processing, and computer vision, good-quality labeled datasets are still lacking for energy management along with load/energy forecasting applications. The key reason behind this is that utility companies and service providers keep real-time and historical data confidential because of various security and privacy concerns. Since the data is usually gathered through sensors, several over issues also exist, such as duplication, mislabeling, and temporary loss of data streams. Hence, there is exigent need of integrated technologies for building intelligent systems in smart microgrids such as combining DL and Internet of Things (IoT) technologies for data collection as well as a streamlining platform for data processing. Blockchain enabled IoT technologies can also help with advanced DL applications in smart grid area.

### *6.2. Higher Computational Cost and Complexity*

ML and DL based approaches entirely rely on historical data, and based on this data, they perform forecasting. A strong dependency on big data, however, demands a large number of storage devices. In addition, high processing is another major challenge, when utilizing approaches focused on DL [136]. Unnecessary features and duplication of data are a main cause of high computation cost and complexity. The higher processing time is required to train redundant data as opposed to train clean data. ML-based approaches and different classification methods can be used to eliminate redundancy from data and speed up the training cycle, while enhancing classification and regression accuracy. Hence, for building reliable, accurate, and low-cost forecasting system, researchers can take the benefits from todays' computing technologies, such as in-database processing, in-memory processing, and parallel processing. Overall, reducing the computational complexity is a fundamental direction for further research.

### *6.3. Spatiotemporal Forecasting*

Probabilistic forecasting of load demand and power generations from RESs plays an important role for optimizing operations of future smart microgrid. It is observed from current literature that a lot of forecasting studies related to energy generation from PV and wind turbines mainly use on-site information and propose solution for single wind or solar farm [137, 138]. Nonetheless, energy farms are geographically distributed and form a network in a distribution system. Regarding load forecasting, most of the current works develop DL-based prediction models only for a single home; however, utility companies are expecting load prediction for a smart community or smart city from researchers [136]. Spatiotemporal prediction approaches are considered more accurate and feasible for future smart microgrid than the single-location techniques [139]. Hence, the development of novel DL models that deal with spatiotemporal dynamics of solar and wind energy along with load demand will enhance the performance of future smart grids.

### *6.4. ANN Accuracy for Long Term Prediction*

ANNs are more efficient and effective means for short-term wind speed and wind power forecasting than physical and statistical forecasting techniques [34]. However, in the case of long-term prediction, the requirement of historical data increases and consequently, ANN accuracy decreases. This weakness needs special attention and ANN-based techniques need to be made accurate for long term predictions, as well.

### *6.5. Heterogeneous Users*

Heterogeneous users and their variant skill levels is another issue that urges the research community to implement ML in a way that is beneficent and understandable for expert as well as novice users. For instance, several papers discussed above only focus on either residential or commercial consumers. In addition, ML models should be capable to support big and small heterogeneous data and remain equally efficient for small and big data [140].

### *6.6. Mobility due to Emerging Applications*

Thanks to the emerging Information and Communication Technologies (ICT), which are making us capable to compliment the traditional energy portfolios with RESs, while at the same time, electrification of energy is occurring at the load side such as integration of Unmanned Aerial Vehicles (UAVs), Electric Vehicles (EVs) and Internet of Shipping [141]. It is to be noted that owing to the mobile nature of above-mentioned technologies, prediction of demands or loads is becoming more challenging. Hence, more sophisticated DL-based prediction schemes are required that consider the mobility models too. Similarly, due to the emerging concepts of Vehicles to Grid (V2G) and expected billions of IoT devices with some having capability of wireless energy harvesting, source side power prediction will become more challenging.

### *6.7. Federated Learning*

The data that is gathered for load forecasting or distributed RESs, is typically obtained in private settings, which is why, it is prone to privacy concerns. Moreover, excessive transmission of data towards a central cloud or data center via wireless communication links requires expensive communication equipment cost and may lead to high latency. This makes it impractical to transmit all the data to a centralized location for training DL models. To overcome the above-mentioned problems, it is important to devise new DL schemes, which can be trained locally at the distributed devices on the bases of the data gathered and collaboratively building a common regional learning platform, a process termed as Federated Learning.

### *6.8. Uncertainty Quantification*

Uncertainty quantification helps in several important decisions today. Forecasting made without uncertainty quantification cannot be reliable and trustworthy [142]. In order to comprehend the DL working, it is necessary to first understand uncertainty quantification. For instance, the DL methodology starts with the collection of more appropriate datasets, selection of an appropriate DL model based on performance goals, training the model by employing a labeled dataset, and optimization of various learning parameters that will help in achieving satisfactory performance. There exist multiple uncertainties involved in the DL steps, which need to be quantified. For instance, they include selection/collection of training data, accuracy and completeness of training data, comprehending the DL models along with their performance bounds and limitations, as well as uncertainties based on operational data [142, 143]. The primary objective of uncertainty quantification is to disclose reliable confidence scores for forecasting results that are generated by DL approaches and what the DL method has not learned properly. In the energy management and forecasting area, the uncertainty quantification has attracted noticeable attention from research community in last couple of years. Current studies show its applications and advantages, i.e., energy management application in smart grid [144], and uncertainty quantification in wind power forecasting [139, 145]. Hence, this area still remains open for future work in order to enhance the reliability and accuracy of DL models.

### 6.9. Growing and Pruning DL Models

Growing and pruning are novel approaches that can be employed to enhance the accuracy and reduce computational complexity of DL models. In this approach, first, a DL architecture is designed with least necessary hidden layers and neurons. Then, new layers and neurons are built in the architecture by applying the growing approach. On the contrary, by employing the pruning approach, a number of neurons along with hidden layers are removed from the DL architecture. Both the growing and pruning approaches-based architectures repeat three key operations until acceptable performance is achieved [146]: i) training the model, ii) changing weights based on growing or pruning criteria, and iii) retraining the model. In the last couple of years, the field of growing and pruning in DL models has earned huge attention from research community and several studies have discussed its effectiveness in various research domains, including speech emotion recognition [147], self care activities [146], and health services enhancement [148]. Hence, the implementation of growing and pruning approaches for DL models in energy management and forecasting area are still an open direction for researchers and industry.

### 6.10. Forecasting of Ocean, Bio, and other Renewable Energies

It is observed from current literature that DL methods are commonly adopted for day-ahead and real-time forecasting from solar and wind energy sources. However, there exist several sources of renewable energy other than solar and wind, for instance, hydro energy, geothermal energy, ocean energy, and bio energy [27]. Although ML- and DL-based method can be applied in these energy sources, their applications for energy prediction are scarce. For example, ML and DL approaches have been employed for geothermal map generation [149], site location modeling for geo thermal [150], scheduling of hydropower plant [151], sea-level variation forecasting for ocean energy [152], output voltage forecasting in geothermal energy [153], and density prediction in bio energy [154]. However, all of the aforementioned works are 6 to 26 years old, and fairly outdated. Therefore, forecasting of energy from geothermal, bio, and other RESs by single and hybrid DL approaches is an unexplored area with a potentially significant research value.

## 7. Conclusion

The intermittent nature of renewable energy sources leads to unreliable energy generation from renewable energy sources, which ultimately necessitate research regarding renewable energy forecasting. Reliable forecasting of solar and wind power can help in improving the quality of service and efficient power management. ML- and DL-based forecasting techniques are considered effective and efficient methodologies for energy forecasting that utilize historical data. In this survey, we performed comprehensive state-of-the-art literature review regarding energy and load forecasting using DL-based techniques. The scope of a set of forecasting models is reviewed in terms of energy types (i.e., wind energy and solar energy) building types (i.e., commercial and non-commercial buildings), and temporal granularities of forecasting (i.e., 5-minutes, 10-minutes, 15-minutes, 30-minutes, and hourly). Furthermore, the properties of the datasets that are used for training and testing forecasting models are also investigated, including data types (i.e., benchmark data, real-time data, and simulation data), dataset features (i.e., data origin, features related to indoor environmental conditions and outdoor weather conditions), dataset recording step (i.e., 10-minutes, 15-minutes, 30-minutes, and hourly), and dataset sizes (i.e., total time duration). The performance levels of studied models are also summarized in terms of forecast accuracy (MAPE, nMAPE, MAE, and RMSE). Each DL-based forecasting model has its own advantages and disadvantages in predicting wind energy, solar energy and load forecasting, thus, it is difficult to determine which is the best among all the models. However, our findings suggest that for all the forecasting applications under consideration, hybrid DL algorithms achieve a high level of performance in terms of prediction accuracy. Moreover, hybrid DL schemes exhibit more tolerance to data incompleteness as compared to pure DNN-based DL. Despite the many advances in DL-based forecasting, a large set of challenges remain unresolved that motivate interesting future research directions, including DL with huge amount of data, lowering computational cost and complexity, spatiotemporal forecasting, mobility due to emerging applications, uncertainty quantification, and use of pruned DL models in smart microgrids.

## References


[1] S. Aslam, An optimal home energy management scheme considering grid connected microgrids with day-ahead weather forecasting using artificial neural network, Ph.D. thesis, Masters Thesis, COMSATS University Islamabad, Islamabad, Pakistan (2018).

[2] M. S. H. Nizami, J. Hossain, E. Fernandez, Multi-agent based transactive energy management systems for residential buildings with distributed energy resources, IEEE Transactions on Industrial Informatics.

[3] L. Park, Y. Jang, S. Cho, J. Kim, Residential demand response for renewable energy resources in smart grid systems, IEEE Transactions on Industrial Informatics 13 (6) (2017) 3165–3173.

[4] M. H. K. Tushar, A. W. Zeineddine, C. Assi, Demand-side management by regulating charging and discharging of the ev, ess, and utilizing renewable energy, IEEE Transactions on Industrial Informatics 14 (1) (2017) 117–126.

[5] R. H. M. Zargar, M. H. Y. Moghaddam, Development of a markov-chain-based solar generation model for smart micro-grid energy management system, IEEE Transactions on Sustainable Energy.

[6] L. Ruan, Y. Yan, S. Guo, F. Wen, X. Qiu, Priority-based residential energy management with collaborative edge and cloud computing, IEEE Transactions on Industrial Informatics.

[7] M. Dicorato, G. Forte, M. Trovato, C. B. Muñoz, G. Coppola, An integrated dc microgrid solution for electric vehicle fleet management, IEEE Transactions on Industry Applications 55 (6) (2019) 7347–7355.

[8] L. Abdallah, T. El-Shennawy, Reducing carbon dioxide emissions from electricity sector using smart electric grid applications, Journal of Engineering 2013.

[9] J. Ahmed, Z. Salam, An accurate method for mppt to detect the partial shading occurrence in a pv system, IEEE transactions on industrial informatics 13 (5) (2017) 2151–2161.

[10] A. Woyte, V. Van Thong, R. Belmans, J. Nijs, Voltage fluctuations on distribution level introduced by photovoltaic systems, IEEE Transactions on energy conversion 21 (1) (2006) 202–209.

[11] Renewable energy trends, https://www.enerdata.net/system/files/publications/global-energy-trends-2018-edition-enerdata.pdf (February 2020).

[12] D. Y. Yamashita, I. Vechiu, J.-P. Gaubert, A review of hierarchical control for building microgrids, Renewable and Sustainable Energy Reviews 118 (2020) 109523.

[13] X. Wang, P. Guo, X. Huang, A review of wind power forecasting models, Energy procedia 12 (2011) 770–778.

[14] B. Yildiz, J. I. Bilbao, A. B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renewable and Sustainable Energy Reviews 73 (2017) 1104–1122.

[15] B.-M. Hodge, C. B. Martinez-Anido, Q. Wang, E. Chartan, A. Florita, J. Kiviluoma, The combined value of wind and solar power forecasting improvements and electricity storage, Applied Energy 214 (2018) 1–15.

[16] K. Wang, X. Qi, H. Liu, A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network, Applied Energy 251 (2019) 113315.

[17] S. Aslam, H. Herodotou, N. Ayub, S. M. Mohsin, Deep learning based techniques to enhance the performance of microgrids: A review, in: 2019 International Conference on Frontiers of Information Technology (FIT), IEEE, 2019, pp. 116–1165.

[18] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, A. R. Varkonyi-Koczy, State of the art of machine learning models in energy systems, a systematic review, Energies 12 (7) (2019) 1301.

[19] A. Mostafaeipour, M. B. Fakhrzad, S. Gharaat, M. Jahangiri, J. A. Dhanraj, S. S. Band, A. Issakhov, A. Mosavi, Machine learning for prediction of energy in wheat production, Agriculture 10 (11) (2020) 517.

[20] A. Abdelaziz, M. Elhoseny, A. S. Salama, A. Riad, A machine learning model for improving healthcare services on cloud computing environment, Measurement 119 (2018) 117–128.

[21] S. Tuli, S. Tuli, R. Tuli, S. S. Gill, Predicting the growth and trend of covid-19 pandemic using machine learning and cloud computing, Internet of Things (2020) 100222.

[22] S. Shamshirband, M. Fathi, A. Dehzangi, A. T. Chronopoulos, H. Alinejad-Rokny, A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues, Journal of Biomedical Informatics (2020) 103627.

[23] Q. Mao, F. Hu, Q. Hao, Deep learning for intelligent wireless networks: A comprehensive survey, IEEE Communications Surveys & Tutorials 20 (4) (2018) 2595–2621.

[24] F. Fotovatikhah, M. Herrera, S. Shamshirband, K.-w. Chau, S. Faizollahzadeh Ardabili, M. J. Piran, Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work, Engineering Applications of Computational Fluid Mechanics 12 (1) (2018) 411–437.

[25] S. Faizollahzadeh Ardabili, B. Najafi, S. Shamshirband, B. Minaei Bidgoli, R. C. Deo, K.-w. Chau, Computational intelligence approach for modeling hydrogen production: A review, Engineering Applications of Computational Fluid Mechanics 12 (1) (2018) 438–458.

[26] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, A review of deep learning for renewable energy forecasting, Energy Conversion and Management 198 (2019) 111799.

[27] S. K. Jha, J. Bilalovic, A. Jha, N. Patel, H. Zhang, Renewable energy: Present research and future scope of artificial intelligence, Renewable and Sustainable Energy Reviews 77 (2017) 297–317.

[28] I. Abubakar, S. Khalid, M. Mustafa, H. Shareef, M. Mustapha, Application of load monitoring in appliances' energy management–a review, Renewable and Sustainable Energy Reviews 67 (2017) 235–245.

[29] K. Amasyali, N. M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews 81 (2018) 1192–1205.

[30] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, X. Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renewable and Sustainable Energy Reviews 82 (2018) 1027–1047.

[31] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, Renewable Energy 105 (2017) 569–582.

[32] M. F. Zia, E. Elbouchikhi, M. Benbouzid, Microgrids energy management systems: A critical review on methods, solutions, and prospects, Applied energy 222 (2018) 1033–1055.

[33] S. Shamshirband, T. Rabczuk, K.-W. Chau, A survey of deep learning techniques: application in wind and solar energy resources, IEEE Access 7 (2019) 164650–164666.

[34] A. P. Marugán, F. P. G. Márquez, J. M. P. Perez, D. Ruiz-Hernández, A survey of artificial neural network in wind energy systems, Applied energy 228 (2018) 1822–1836.

[35] S. Mohanty, P. K. Patra, S. S. Sahoo, A. Mohanty, Forecasting of solar energy with application for a growing economy like india: Survey and implication, Renewable and Sustainable Energy Reviews 78 (2017) 539–553.

[36] S. N. Fallah, R. C. Deo, M. Shojafar, M. Conti, S. Shamshirband, Computational intelligence approaches for energy load forecasting in smart energy management grids: state of the art, future challenges, and research directions, Energies 11 (3) (2018) 596.

[37] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, J. Massana, A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings, IEEE Communications Surveys & Tutorials 16 (3) (2014) 1460–1495.

[38] M. Manbachi, M. Ordonez, Intelligent agent-based energy management system for islanded ac/dc microgrids, IEEE Transactions on Industrial Informatics.

[39] M. Ahrarinouri, M. Rastegar, A. R. Seifi, Multi-agent reinforcement learning for energy management in residential buildings, IEEE Transactions on Industrial Informatics.

[40] A. T. Eseye, M. Lehtonen, Short-term forecasting of heat demand of buildings for efficient and optimal energy management based on integrated machine learning models, IEEE Transactions on Industrial Informatics.

[41] M. Aslam, J.-M. Lee, H.-S. Kim, S.-J. Lee, S. Hong, Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study, Energies 13 (1) (2020) 147.

[42] M. H. Hassoun, et al., Fundamentals of artificial neural networks, MIT press, 1995.

[43] G. Daniel, Principles of artificial neural networks, Vol. 7, World Scientific, 2013.

[44] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: A survey, Heliyon 4 (11) (2018) e00938.

[45] M. M. Mijwel, Artificial neural networks advantages and disadvantages, Retrieved from LinkedIn: https://www. linkedin. com/pulse/artificial-neuralnet works-advantages-disadvantages-maad-m-mijwel.

[46] J. V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, Journal of clinical epidemiology 49 (11) (1996) 1225–1231.

[47] Y. Bengio, et al., Learning deep architectures for ai, Foundations and trends® in Machine Learning 2 (1) (2009) 1–127.

[48] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.

[49] W. Zhang, K. Itoh, J. Tanida, Y. Ichioka, Parallel distributed processing model with local space-invariant interconnections and its optical architecture, Applied optics 29 (32) (1990) 4790–4797.

[50] M. Matsugu, K. Mori, Y. Mitari, Y. Kaneda, Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Networks 16 (5-6) (2003) 555–559.

[51] Basic overview of convolutional Neural Network (CNN), `https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17` (December 2020).

[52] Y. Bengio, et al., Learning deep architectures for ai, Foundations and trends® in Machine Learning 2 (1) (2009) 1–127.

[53] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.

[54] A. Gensler, J. Henze, B. Sick, N. Raabe, Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks, in: 2016 IEEE international conference on systems, man, and cybernetics (SMC), IEEE, 2016, pp. 002858–002865.

[55] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (7) (2006) 1527–1554.

[56] K.-P. Lin, P.-F. Pai, Y.-J. Ting, Deep belief networks with genetic algorithms in forecasting wind speed, IEEE Access 7 (2019) 99244–99253.

[57] D. H. Ackley, G. E. Hinton, T. J. Sejnowski, A learning algorithm for boltzmann machines, Cognitive science 9 (1) (1985) 147–169.

[58] C.-T. Lin, C. Lee, Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems, Prentice-Hall, Inc., 1996.

[59] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE transactions on neural networks 5 (2) (1994) 157–166.

[60] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, et al., Gradient flow in recurrent nets: the difficulty of learning long-term dependencies (2001).

[61] Recurrent Neural Networks Tutorial, `http://www.wildml.com/2015/09/recurrentneural-networks-tutorial-part-1-introduction-to-rnns/` (June 2019).

[62] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on lstm recurrent neural network, IEEE Transactions on Smart Grid 10 (1) (2017) 841–851.

[63] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[64] F. A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with lstm, in: 9th International Conference on Artificial Neural Networks (ICANN), IET, 1999, pp. 850–855.

[65] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, C. Kang, Probabilistic individual load forecasting using pinball loss guided lstm, Applied Energy 235 (2019) 10–20.

[66] J. Kim, J. Moon, E. Hwang, P. Kang, Recurrent inception convolution neural network for multi short-term load forecasting, Energy and Buildings 194 (2019) 328–341.

[67] W. Hu, Z. Chen, B. Bak-Jensen, Optimal operation strategy of battery energy storage system to real-time electricity price in denmark, in: IEEE PES general meeting, IEEE, 2010, pp. 1–7.

[68] S. Aslam, N. Javaid, F. A. Khan, A. Alamri, A. Almogren, W. Abdul, Towards efficient energy management and power trading in a residential area via integrating a grid-connected microgrid, Sustainability 10 (4) (2018) 1245.

[69] S. Aslam, Z. Iqbal, N. Javaid, Z. A. Khan, K. Aurangzeb, S. I. Haider, Towards efficient energy management of smart buildings exploiting heuristic optimization with real time and critical peak pricing schemes, Energies 10 (12) (2017) 2065.

[70] L. Cheng, H. Zang, T. Ding, R. Sun, M. Wang, Z. Wei, G. Sun, Ensemble recurrent neural network based probabilistic wind speed forecasting approach, Energies 11 (8) (2018) 1958.

[71] J. Wang, Y. Song, F. Liu, R. Hou, Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models, Renewable and Sustainable Energy Reviews 60 (2016) 960–981.

[72] NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data), `https://www.osti.gov/biblio/1052222-nrel-national-wind-technology-centernwtc-m2-tower-boulder-colorado-data` (May 2020).

[73] G. Li, H.-D. Chiang, Toward cost-oriented forecasting of wind power generation, IEEE Transactions on Smart Grid 9 (4) (2016) 2508–2517.

[74] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012 (2014).

[75] C. Wan, J. Lin, J. Wang, Y. Song, Z. Y. Dong, Direct quantile regression for nonparametric probabilistic forecasting of wind power generation, IEEE Transactions on Power Systems 32 (4) (2016) 2767–2778.

[76] M. Cui, J. Zhang, Q. Wang, V. Krishnan, B.-M. Hodge, A data-driven methodology for probabilistic wind power ramp forecasting, IEEE Transactions on Smart Grid 10 (2) (2017) 1326–1338.

[77] C. Draxl, A. Clifton, B.-M. Hodge, J. McCaa, The wind integration national dataset (wind) toolkit, Applied Energy 151 (2015) 355–366.

[78] H. Liu, X.-w. Mi, Y.-f. Li, Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and elman neural network, Energy conversion and management 156 (2018) 498–514.

[79] J. Wang, S. Xiong, A hybrid forecasting model based on outlier detection and fuzzy time series–a case study on hainan wind farm of china, Energy 76 (2014) 526–541.

[80] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, K.-D. Lu, Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization, Energy conversion and management 165 (2018) 681–695.

[81] H.-z. Wang, G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, Y.-t. Liu, Deep learning based ensemble approach for probabilistic wind power forecasting, Applied energy 188 (2017) 56–70.

[82] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, Z. Zhang, Lstm-efg for wind power forecasting based on sequential correlation features, Future Generation Computer Systems 93 (2019) 33–42.

[83] National Renewable Energy Laboratory (NREL), `https://pvwatts.nrel.gov/` (November 2019).

[84] Y. Li, H. Shi, F. Han, Z. Duan, H. Liu, Smart wind speed forecasting approach using various boosting algorithms, big multi-step forecasting strategy, Renewable energy 135 (2019) 540–553.

[85] X. Mi, H. Liu, Y. Li, Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine, Energy conversion and management 180 (2019) 196–205.

[86] Y.-L. Hu, L. Chen, A nonlinear hybrid wind speed forecasting model using lstm network, hysteretic elm and differential evolution algorithm, Energy Conversion and Management 173 (2018) 123–142.

[87] P. Sorensen, N. A. Cutululis, A. Vigueras-Rodríguez, L. E. Jensen, J. Hjerrild, M. H. Donovan, H. Madsen, Power fluctuations from large wind farms, IEEE Transactions on Power Systems 22 (3) (2007) 958–965.

[88] C. Draxl, A. Clifton, B.-M. Hodge, J. McCaa, The wind integration national dataset (wind) toolkit, Applied Energy 151 (2015) 355–366.

[89] K. Colchester, H. Hagras, D. Alghazzawi, G. Aldabbagh, A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms, Journal of Artificial Intelligence and Soft Computing Research 7 (1) (2017) 47–64.

[90] X. G. Agoua, R. Girard, G. Kariniotakis, Short-term spatio-temporal forecasting of photovoltaic power production, IEEE Transactions on Sustainable Energy 9 (2) (2017) 538–546.

[91] M. C. Sorkun, C. Paoli, Ö. D. Incel, Time series forecasting on solar irradiation using deep learning, in: 2017 10th International Conference on Electrical and Electronics Engineering (ELECO), IEEE, 2017, pp. 151–155.

[92] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PloS one 12 (7).

[93] K. Chen, Z. He, K. Chen, J. Hu, J. He, Solar energy forecasting with numerical weather predictions on a grid and convolutional networks, in: 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), IEEE, 2017, pp. 1–5.

[94] S. Dey, S. Pratiher, S. Banerjee, C. K. Mukherjee, Solarisnet: A deep regression network for solar radiation prediction, arXiv preprint arXiv:1711.08413.

[95] B. Amrouche, X. Le Pivert, Artificial neural network based daily local forecasting for global solar radiation, Applied energy 130 (2014) 333–341.

[96] A. K. Yadav, S. Chandel, Solar radiation prediction using artificial neural network techniques: A review, Renewable and sustainable energy reviews 33 (2014) 772–781.

[97] J.-L. Chen, H.-B. Liu, W. Wu, D.-T. Xie, Estimation of monthly solar radiation from measured temperatures using support vector machines–a case study, Renewable Energy 36 (1) (2011) 413–420.

[98] S. Salcedo-Sanz, C. Casanova-Mateo, J. Muñoz-Marí, G. Camps-Valls, Prediction of daily global solar irradiation using temporal gaussian processes, IEEE Geoscience and Remote Sensing Letters 11 (11) (2014) 1936–1940.

[99] A. Alzahrani, P. Shamsi, M. Ferdowsi, C. Dagli, Solar irradiance forecasting using deep recurrent neural networks, in: 2017 IEEE 6th international conference on renewable energy research and applications (ICRERA), IEEE, 2017, pp. 988–994.

[100] High-resolution solar radiation datasets, `http://www.nrcan.gc.ca/energy/renewable-electricity/solar-photovoltaic/18409` (June 2019).

[101] Y. Wang, Y. Shen, S. Mao, G. Cao, R. M. Nelms, Adaptive learning hybrid model for solar intensity forecasting, IEEE Transactions on Industrial Informatics 14 (4) (2018) 1635–1645.

[102] M. Abdel-Nasser, K. Mahmoud, Accurate photovoltaic power forecasting models using deep lstm-rnn, Neural Computing and Applications 31 (7) (2019) 2727–2740.

[103] P.-H. Kuo, C.-J. Huang, A green energy application in energy management systems by an artificial intelligence-based solar radiation forecasting model, Energies 11 (4) (2018) 819.

[104] D. Díaz-Vico, A. Torres-Barrán, A. Omari, J. R. Dorronsoro, Deep neural networks for wind and solar energy prediction, Neural Processing Letters 46 (3) (2017) 829–844.

[105] J. R. Andrade, R. J. Bessa, Improving renewable energy forecasting with a grid of numerical weather predictions, IEEE Transactions on Sustainable Energy 8 (4) (2017) 1571–1580.

[106] N. Sharma, P. Sharma, D. Irwin, P. Shenoy, Predicting solar generation from weather forecasts using machine learning, in: 2011 IEEE international conference on smart grid communications (SmartGridComm), IEEE, 2011, pp. 528–533.

[107] K. Yan, H. Shen, L. Wang, H. Zhou, M. Xu, Y. Mo, Short-term solar irradiance forecasting based on a hybrid deep learning methodology, Information 11 (1) (2020) 32.

[108] T. Stoffel, A. Andreas, University of nevada (unlv): Las vegas, nevada (data), Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2006).

[109] A. Gensler, J. Henze, N. Raabe, V. Pankraz, Germansolarfarm data set (2016).

[110] A. Dedinec, S. Filiposka, A. Dedinec, L. Kocarev, Deep belief network based electricity load forecasting: An analysis of macedonian case, Energy 115 (2016) 1688–1700.

[111] S. Motepe, A. N. Hasan, B. Twala, R. Stopforth, Power distribution networks load forecasting using deep belief networks: The south african case, in: 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), IEEE, 2019, pp. 507–512.

[112] A. McGovern, D. J. Gagne, J. Basara, T. M. Hamill, D. Margolin, Solar energy prediction: An international contest to initiate interdisciplinary research on compelling meteorological problems, Bulletin of the American Meteorological Society 96 (8) (2015) 1388–1395.

[113] High-resolution solar radiation datasets, `http://www.nrcan.gc.ca/energy/renewable-electricity/solar-photovoltaic/18409` (June 2017).

[114] The UMass Trace Repository, `http://traces.cs.umass.edu/` (January 2018).

[115] 2013–2014 solar energy prediction contest, `https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest` (August 2014).

[116] Historical solar radiation data of NWS, `http://www.weather.gov` (September 2019).

[117] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta, M. A. Brown, R. M. Pendyala, Machine learning approaches for estimating commercial building energy consumption, Applied energy 208 (2017) 889–904.

[118] A. Rahman, V. Srikumar, A. D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Applied energy 212 (2018) 372–385.

[119] H. Shi, M. Xu, R. Li, Deep learning for household load forecasting—a novel pooling deep rnn, IEEE Transactions on Smart Grid 9 (5) (2017) 5271–5280.

[120] X. Dong, L. Qian, L. Huang, Short-term load forecasting in smart grid: A combined cnn and k-means clustering approach, in: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2017, pp. 119–125.

[121] X. Qiu, Y. Ren, P. N. Suganthan, G. A. Amaratunga, Empirical mode decomposition based ensemble deep learning for load demand time series forecasting, Applied Soft Computing 54 (2017) 246–255.

[122] AEMO, Australian Energy Market Operator, https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-planning-data (March 2020).

[123] N. Ghadimi, A. Akbarimajd, H. Shayeghi, O. Abedinia, A new prediction model based on multi-block forecast engine in smart grid, Journal of Ambient Intelligence and Humanized Computing 9 (6) (2018) 1873–1888.

[124] G. T. Ribeiro, V. C. Mariani, L. dos Santos Coelho, Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting, Engineering Applications of Artificial Intelligence 82 (2019) 272–281.

[125] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, M. K. Afzal, Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities, Sustainability 11 (4) (2019) 987.

[126] F. Amara, K. Agbossou, Y. Dubé, S. Kelouwani, A. Cardenas, S. S. Hosseini, A residual load modeling approach for household short-term load forecasting application, Energy and Buildings 187 (2019) 132–143.

[127] W.-R. Wu, P.-C. Chen, Adaptive ar modeling in white gaussian noise, IEEE Transactions on Signal Processing 45 (5) (1997) 1184–1192.

[128] C. Bennett, R. A. Stewart, J. Lu, Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks, Energies 7 (5) (2014) 2938–2960.

[129] A. Ahmad, N. Javaid, M. Guizani, N. Alrajeh, Z. A. Khan, An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid, IEEE Transactions on Industrial Informatics 13 (5) (2016) 2587–2596.

[130] AEMO, Australian Energy Market Operator 2013, http://www.aemo.com.au (June 2013).

[131] MEPSO (Electricity Transmission System Operator of Macedonia), http://mepso.com.mk/en-us/Default.aspx (June 2019).

[132] South African electrification programme", Global network on energy for sustainable development, http://energyaccess.gnesd.org/cases/22-south-african-electrificationprogramme.html (June 2019).

[133] New York City Mayor's Office of Sustainability. Local Law 84 Data Disclosures, http://www.nyc.gov/html/gbee/html/plan/ll84_scores.shtml (April 2018).

[134] Electricity smart metering customer behaviour trials (cbt) findings report, Tech. rep., Commission for Energy Regulation, https://www.cru.ie/wp-content/uploads/2011/07/cer11080ai.pdf (2011).

[135] K. Zhou, C. Fu, S. Yang, Big data driven smart energy management: From big data to big insights, Renewable and Sustainable Energy Reviews 56 (2016) 215–225.

[136] T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting 32 (3) (2016) 914–938.

[137] Y. Zhang, J. Wang, X. Wang, Review on probabilistic forecasting of wind power generation, Renewable and Sustainable Energy Reviews 32 (2014) 255–270.

[138] D. W. Van der Meer, J. Widén, J. Munkhammar, Review on probabilistic forecasting of photovoltaic power production and electricity consumption, Renewable and Sustainable Energy Reviews 81 (2018) 1484–1512.

[139] H. Quan, A. Khosravi, D. Yang, D. Srinivasan, A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids, IEEE Transactions on Neural Networks and Learning Systems.

[140] S. Schelter, F. Biessmann, T. Januschowski, D. Salinas, S. Seufert, G. Szarvas, M. Vartak, S. Madden, H. Miao, A. Deshpande, et al., On challenges in machine learning model management., IEEE Data Eng. Bull. 41 (4) (2018) 5–15.

[141] S. Aslam, M. P. Michaelides, H. Herodotou, Internet of ships: A survey on architectures, emerging applications, and challenges, IEEE Internet of Things Journal.

[142] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, A. Khosravi, U. R. Acharya, V. Makarenkov, et al., A review of uncertainty quantification in deep learning: Techniques, applications and challenges, arXiv preprint arXiv:2011.06225.

[143] E. Begoli, T. Bhattacharya, D. Kusnezov, The need for uncertainty quantification in machine-assisted medical decision making, Nature Machine Intelligence 1 (1) (2019) 20–23.

[144] A. K. Raz, P. C. Wood, L. Mockus, D. A. DeLaurentis, System of systems uncertainty quantification using machine learning techniques with smart grid application, Systems Engineering 23 (6) (2020) 770–782.

[145] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, G. Zhang, Deep uncertainty quantification: A machine learning approach for weather forecasting, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2087–2095.

[146] K. Akyol, Comparing of deep neural networks and extreme learning machines based on growing and pruning approach, Expert Systems with Applications 140 (2020) 112875.

[147] M. Sánchez-Gutiérrez, P. P. González-Pérez, Discriminative neural network pruning in a multiclass environment: A case study in spoken emotion recognition, Speech Communication.

[148] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio, S. K. Antani, Iteratively pruned deep learning ensembles for covid-19 detection in chest x-rays, arXiv preprint arXiv:2004.08379.

[149] S. A. Kalogirou, G. A. Florides, P. D. Pouloupatis, I. Panayides, J. Joseph-Stylianou, Z. Zomeni, Artificial neural networks for the generation of geothermal maps of ground temperature at various depths by considering land configuration, Energy 48 (1) (2012) 233–240.

[150] I.-A. Yeo, J.-J. Yee, A proposal for a site location planning model of environmentally friendly urban energy supply plants using an environment and energy geographical information system (e-gis) database (db) and an artificial neural network (ann), Applied Energy 119 (2014) 99–117.

[151] R.-H. Liang, Y.-Y. Hsu, Scheduling of hydroelectric generations using artificial neural networks, IEE Proceedings-Generation, Transmission and Distribution 141 (5) (1994) 452–458.

[152] O. Makarynskyy, D. Makarynska, M. Kuhn, W. Featherstone, Predicting sea level variations with artificial neural networks at hillarys boat harbour, western australia, Estuarine, Coastal and Shelf Science 61 (2) (2004) 351–360.

[153] A. Garg, V. Vijayaraghavan, S. Mahapatra, K. Tai, C. Wong, Performance evaluation of microbial fuel cell by artificial intelligence methods, Expert systems with applications 41 (4) (2014) 1389–1399.

[154] H. Yang, Z. Ring, Y. Briker, N. McLean, W. Friesen, C. Fairbridge, Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by lc and gc–ms, Fuel 81 (1) (2002) 65–74.