2022 3rd Int. Conf. on Innovations in Science, Engineering and Technology (ICISET)
26-27 February 2022, Chittagong, Bangladesh

2022 International Conference on Innovations in Science, Engineering and Technology (ICISET) | 978-1-6654-8397-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICISET54810.2022.9775816

Machine Learning Models for Content Classification in Film Censorship and Rating

Syma Afsha*, Mahmudul Haque† and Hussain Nyeem‡
Department of Electrical, Electronic and Communication Engineering (EECE)
Military Institute of Science and Technology (MIST)
Mirpur Cantonment, Dhaka–1216, Bangladesh
*symaafsha.eece@gmail.com †mahmud.eece@gmail.com ‡h.nyeem@eece.mist.ac.bd

***Abstract*—Automated Film Censorship and Rating (AFCR) has recently turned out to be a major research area of Machine Learning (ML). The production and streaming services of films including movies, tv-series, animations and other audio-visual contents have been widely expanded leading to their manual censorship and rating to be a more exhausting task. Development of ML based methods has thus been emerging to designing an AFCR system. However, the initial ad-hoc efforts of developing the AFCR system demand a "complete" conceptual model of the system with its potential classes and their criteria. This paper primarily attempts to determine both the general and contextual classes of the content, and their criteria for an AFCR system. Besides, the state-of-the-art AFCR systems have been systematically reviewed to identify their underlying ML models, advantages and limitations. With a comparative analysis of the exiting ML models, we have demonstrated the effectiveness of sequential and multimodal analysis in the development of an efficient AFCR system.**

***Index Terms*—AFCR, censorship, CNN, DNN, film rating, ML, RNN.**

I. Introduction

Automated Film Censorship and Rating (AFCR) is one of the leading topics in the Machine Learning (ML) research field. Many films including movies, tv-series, animations and other audio-visual contents have inappropriate words or visual content which is unsuitable for uncongenial audiences and affects the younger generation. Primarily, it affects under-aged audiences through their exposure to visual contents on TV or other social networking websites like YouTube, Facebook *etc*. Young minds are influenced and can be intrigued to try unethical and unlawful actions in real life as a result of the films that contain activities suggestive of such negative actions, which requires the sensitive contents to be censored.

Sensitive content identification is a difficult and unsolved task, owing to the subjectivity and openness of the idea [1]. Though preventive measures are taken by broadcasters to identify sensitive contents manually, it is a slow, monotonous and inefficient process [2]. As a result, AFCR is essential for generating automatic and generalised ratings for films or any audio-visual content to determine the minimum age-group appropriate for viewing the content. Different aspects of a film that may be deemed inappropriate include violence, nudity, substance abuse, profanity, abnormal psychological substance, or other topics that are inappropriate for children or teenagers in general [3]. Based on these contents, relevant classification standards apply to the films or any audio-visual contents prepared for public display in many developed countries.

Any audio-visual content is broadly categorised into three parts: advisory, restricted, and adult or nudity. The advisory categories are further divided into General (G), Parental Guidance (PG), and Mature (M). The restricted types include Mature Accompanied (MA15+) and Restricted (R18+). All these classifications have specified minimum age for viewing any audio-visual content under that specific class. ML plays a vital role in film censorship to detect and classify sensitive and inappropriate content into a given classification standard [4], [5]. Most researchers utilised ML techniques such as Convolutional Neural Network (CNN), Support Vector Machine (SVM), Sparse Linear Discrimination (SLD), Long Short-Term Memory (LSTM), Recurrent neural networks (RNN) *etc*. in their work for detecting unsuitable contents in movies.

Despite the promises of the ad-hoc efforts made in developing ML-based AFCR system, there is an obvious need for conceptual model that interprets the classification standards for the machine to learn and classify. Thus, this paper primarily attempts to determine both the general and contextual classes of the content, and their criteria for an AFCR system. Besides, an overview of film content rating systems based on the different ML techniques is also reported to explore an appropriate and efficient AFCR system.

The rest of this paper is structured as follows. Sec. II marks out different existing approaches to develop an AFCR system followed by the criteria and rating classifications for an ideal AFCR system defined and discussed in Sec. III. Summary of the implementation is presented in Sec. IV, and the performance criteria of the prominent methods and their results are analysed are presented in Sec. V with the scopes future development. Finally, concluding remarks with the outcome and significance of this work have been summarised in Sec. VI.

II. Related Works

Film censorship requires identifying generalised criteria for each training including the recognition and classification of activities and their intensity. Recently, ML is one of the popular approaches for detecting unsuitable content [2]. Cifuentes *et al*. [6] reviewed the different strategies available in the literature for inappropriate scene detection in videos using Deep Learning (DL) approaches. Han *et al*. [7] proposed the detection of inappropriate videos based on SLD. An online YouTube video dataset was used to train models to categorise safe and unsafe videos. Again, Lyn *et al*. [2] suggested a CNN and SVM based system for film censorship. Frames of films were fed as input into the CNN and were classified into two groups as ill-suited scenes or not. Besides, Gruosso *et al*. [8] developed a content rating system to evaluate the content and report the suitability for different age-group viewers. They provided an audiovisual content rating model based on Inception v3 CNN architecture to classify and regulate disturbing contents automatically.

Wazir *et al*. [9] considered the audio signal to be a significant element in cases where the visual elements were not clear, especially, in the adult contents with dark scenes and scenes with covered bodies. They have proposed an adult content recognition model using audio features utilising RNN. They have reported the potential of LSTM cells to handle difficulties involving temporal dependencies and long-term learning, as well as vanishing gradient problems in RNN. A novel multimodal fusion approach was proposed by Moreira *et al*. [1] for sensitive scene localisation through two of the most common types of sensitive content and reported quantitative and qualitative results. They have shown that audio was vital for improving the effectiveness of space-temporal approaches as they strongly outperform still-image solutions.

Jani *et al*. [10] developed an efficient and accurate AFCR system that can detect explicit languages and inappropriate visual contents in films. Similarly, an inappropriate scene filtering application to classify images and video frames containing nudity was proposed by Garcia *et al*. [11]. In contrast, Moreira *et al*. [12] suggested an algorithm for detecting nudity using DL to extract features from detected skin regions. Emon *et al*. [13] represented different ML and DL-based algorithms like Linear Support Vector Classifier (LinearSVC), Logistic Regression, Multinomial Naïve Bayes (MNB), Random Forest (RF), Artificial Neural Network (ANN), RNN with an LSTM to detect multi-types of Bengali texts and analysis their best performances.

On ratings of the Motion Picture Association of America (MPAA), Shafaei *et al*. [14] presented a method to predict the suitability of the movie content for children and young adults based on scripts by using RNN architecture that jointly models the genre and the emotions. Aldahoul *et al*. [15] introduced a system that explored various methods and approaches, including standard features and CNN, to detect inappropriate visual content in cartoon animation. Chin *et al*. [16] suggested automatically detecting explicit content in Korean lyrics and comparing their performances using several ML models. Wazir *et al*. [17] proposed an intelligent model for profane language censorship through automated and robust detection by DCNN. Susanty *et al*. [18] developed an artificial neural network model for classifying offensive words considering the structure of the sentence to get its context.

Guedes *et al*. [19] suggested a CNN and an SVM classifier based approach to detect violent actions that involve corporal struggle in the video stream. On the other hand, Chaudhari *et al*. [20] proposed an approach to mute the audio and pixelate the lips present in presence of profane content using ML. Traoré *et al*. [21] represented a DL architecture for violence detection, combining RNN and 2- dimensional CNN (2D-CNN). CNN extracts spatial characteristics in each frame, while RNN extracts temporal characteristics. They also used optical flow to encode movement between frames for better performance. Subsequently, Martinez *et al*. [22] proposed the features that capture lexical, semantic, sentiment and abusive language characteristics to identify violence from the language used in movie scripts.

While above studies on the recent development of ML models for the AFCR system indicate an intensive research interest in the area, it also imply that development of the models are mostly on ad-hoc basis. This means that, for the varying considerations of the rating standards and their implications on the criteria make the validation of the models more challenging for a different scenario or standard. In other words, the state-of-the-art models for the AFCR system remains unknown. Thus, the need for developing a conceptual model with classification criteria for the machine to learn and classify the audio-visual contents is obvious.

III. Redefining the AFCR Criteria

AFCR is a system to rate films based on age demographics to prevent uncongenial audiences from accessing inappropriate content. Motion Picture Content Rating (MPCR) System is one of the film rating systems to rate films. The stated goal of the MPAA rating system is to offer parents some advanced information about movies to decide what movies they want their children to see or not. Taking components that have been considered for rating a film by the MPCR [5] into consideration, the components that can be considered for AFCR has been illustrated in Fig. 1.

Based on the presence and intensity of these types of contents, films are classified and the classification is justified.

A. Classification

Rating classifications are planned to classify movies to determine the suitability of films based on respective age groups [4] hence executed as follows.

1) Rated G: General Audiences. All ages audiences are allowed.

2) Rated PG: Parental Guidance Suggested. Some content may be inappropriate for underage audiences.

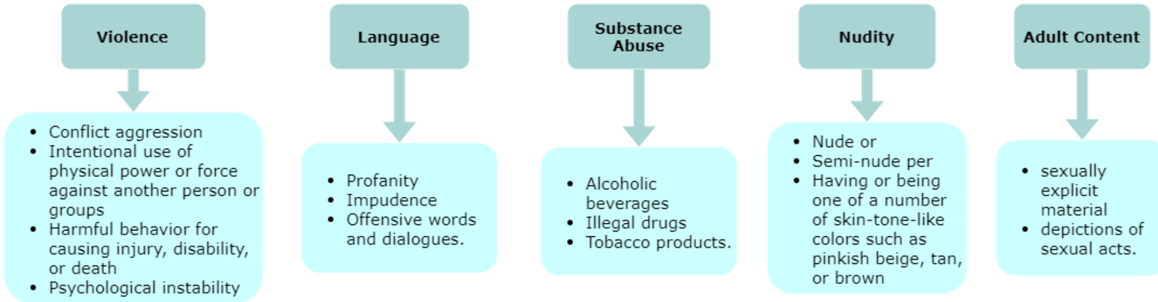
3) Rated PG-13: Parents Strongly Awareness. Some content may be inappropriate for underage audiences under 13.

4) Rated R: Restricted. Underage audiences under 17 require escorts with parents or adult guardians.

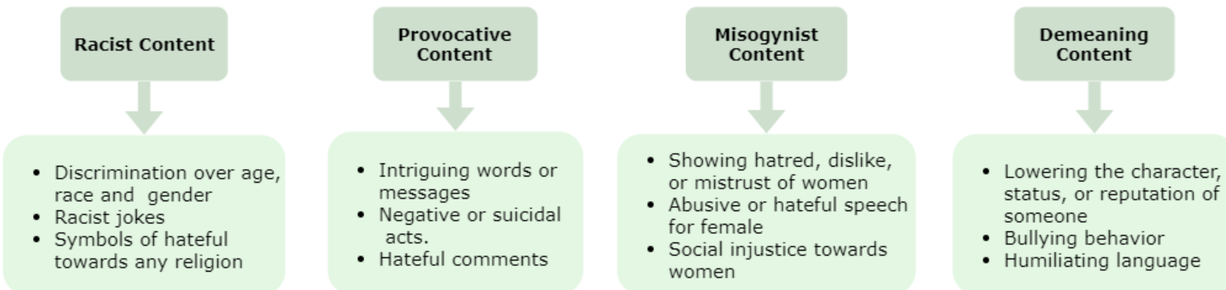
5) Rated NC-17: Adults Only. No One 17 and under is allowed

B. Considerations for Classification

Based on the type of content and the intensity of the content, these classifications are applied.



(a)



(b)

Fig. 1: Proposed classes and their criteria for AFCR: (a) general and (b) contextual
.

1) Violence may be minimal in G-rated films and maybe intense in PG-rated movies. Depictions of intense violence are permitted under the PG-13 rating, but the violence that is both realistic and extreme will generally require at least an R rating and violence acceptable for NC-17.
2) Language beyond polite conversation is permitted in G-rated films, but no more important words are present. Profanity and impudence may be present in PG-rated films, and the use of one of the harsher sexually derived words may be present in a PG-13 rating film. Multiple occurrences of sexual context will usually incur an R rating and are acceptable for NC-17.
3) Drug use content may be present in PG-13 and above. Drug abuse will generally require an R rating and is acceptable for NC-17.
4) Nudity is restricted to PG and constitutes more than brief nudity will require at least a PG-13 rating. Sexually oriented nudity will generally need an R rating and is acceptable for NC-17.
5) Adult content is only acceptable for NC-17.

Based on these criteria and contents, we propose a conceptual model of an AFCR system as illustrated in Fig. 2. We would now analyse the existing prominent AFCR models in light of the proposed framework in the following section.

IV. Prominent ML-based AFCR Models

Several ML-based computer vision models were developed for the AFCR system. ML has enabled automated learning through the retention of data with a varying set of criteria and considerations of an AFCR system. Besides, the type of

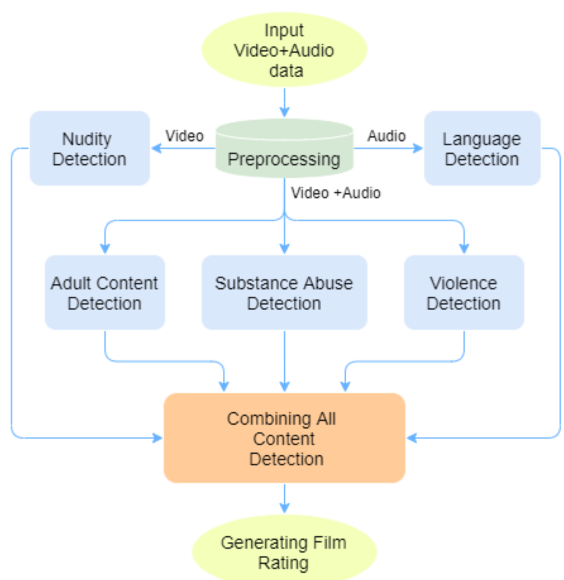


Fig. 2: Conceptual model of the proposed AFCR system.

dataset used and the extracted features also play crucial roles in defining the architecture of different ML models, which are illustrated in Table I.

We observe that none of the AFCR models completely addresses all content types of classification for AFCR having different considerations and criteria. Hence, the models that addressed each criterion separately or multiple criteria together have been considered for further evaluation based on the

TABLE I: SUMMARY OF PROMINENT ML-BASED AFCR MODELS

Model	Datasets	Content-types	Features	Limitations	Performance
SVM and RNN [14]	Movie script dataset [22]	Violence, nudity and adult content	Detects genre and the emotions to predict MPAA rating, works 6% better than other ML methods compared.	Did not use visual features.	F1-score: 78%
Image-based linear regression [10]	COIL-20-PRoC and collection of English words	Language and adult content	Considers video and language, accepts videos from phone's storage and assigns a rating.	Accuracy of the model not determined.	vulgar factor: ≥ 5
CNN [23]	Pornography2k [24] and UCF101 [25]	Nudity and adult content	Near real time detection of private parts in video by binary classification for 50,000 annotated objects	Temporal nature of the dataset not utilised.	**Pornography-2k dataset:** FP: 1.66%, FN: 2.34%, **UCF101 dataset:** FP: 3.41%
SLD [7]	online video [26]	Nudity, Adult Content	Classification method, fastest train and test data processing time, dependent on sample size, white noise and regressive coefficient.	Temporal nature of the dataset not utilised.	accuracy: 98.31%
RNN [9]	NDPI video dataset [27]	Adult content	Mel-Frequency Cepstrum Coefficients (MFFCs) used to extract features of voice files converting from videos using binary classification.	Long training time, temporal nature of the dataset not utilised.	accuracy: 86.50%, precision: 89.00%, recall: 84.76%, f-score: 86.83%
Model [16]	South Korean Broadcasting System (KBS)	Language	Filtering method implemented that used custom profanity dictionary, outperformed competing models they have previously designed.	Model was trained with only song data.	precision: 0.96, recall: 0.97, f1 score: 0.96
CNN Inception V3 model [8]	Violent Scenes Dataset (VSD) [28]	Violence	Automatically classify and censor violent scenes using two classifiers.	Did not test other content, high false discovery rate.	**binary classification:** precision: 88.8% on R class, accuracy 90.9%, **multiple classifications:** precision: 81% on R class, accuracy : 76.1%. false discovery rate of PG13 class: 33.2%
MFCC, PROS, HOG, TRoF [1]	Pornography-2k dataset	Nudity, adult contents, language	Adult and violent content localisation, quantitative and qualitative results are reported with a low memory footprint and runtime.	All sensitive contents are not addressed.	**Audio:** accuracy: 76.31% (PROS), 79.72% (MFCC), F2 score: 77.22% (PROS), 80.98% (MFCC); **Video:** accuracy: 87.25% (HOG), 86.47% (TRoF), F2 score: 89.65% (HOG), 89.89% (TRoF); **Multimodal:** accuracy: 90.72% (all), F2 score: 91.93% (all)
RNN and SVM [22]	Movie-DiC	Violence and language	Detects violent content in movie scripts detecting abusive language using multiple classifications.	Multimodal approaches and sentiment-related features.not used.	**GRU (16):** precision: 60.9%, recall: 60%, f-score: 60.4%
Nudity Detection Algorithm [11]	Local image and video dataset	Nudity	Nudity filtering application is highly dependent on skin colour, using binary classification.	Temporal nature of the dataset not utilised.	precision: 90.33%, accuracy: 80.23%
RANDOM FOREST-64 [12]	AIIA-PID4 [29]	Nudity	Detects faces as false positives in portrait photos using binary classification.	Temporal nature of the dataset not utilised, varying skin regions not tested.	accuracy: 96.96%, f1-score: 94.94%
RNN [13]	Custom data (Popular media services)	Language	Detects of multi-type abusive Bengali language using multiple classifications,performs best to detect offensive language.	Cannot be used for detecting other contents.	accuracy: 82.20%
ANN [18]	Internet videos	Language	Informal grammar and abbreviation can be adequately tackled using binary classification.	Cannot be used for detecting other content.	accuracy: 96.8%
2D-CNN and RNN [21]	Hockey dataset, Violent Flow dataset	Violence	Optical flow used to assess movement across images for reliable performance, combining 2D-CNN with bi-directional RNN for binary classification.	Cannot be used for detecting other content.	**Hockey dataset:** accuracy 99%, **Violence Flow dataset:** accuracy: 93.75%, **Real-time Violence Situation:** accuracy: 96.74%
Fine-tuned ResNet50_V2 and SVM [2]	Collection web-videos	Nudity and adult content	Applies CNN as a feature extractor and SVM as a classifier for more reliable performance using binary classification.	Hyperparameters did not tune practically and systematically, temporal nature of the dataset not utilised.	accuracy: 92.80%, precision: 93%, recall: 93%, f1 score: 93%
Deep CNN, ResNet and EfficientNet [15]	NPDI dataset, and NPDI videos	Nudity and adult content	Binary classification done on feature vectors extracted from various CNNs using SVM to detects inappropriate content with complex background.	Problem of small-scale porn regions inside the frames not considered, temporal nature of the dataset not utilised.	accuracy: 87.87%, f1 score: 87.87%, AUC (94.40%), FNR: 10.87%, FPR: 13.35%
CNN and SVM [19]	Hockey, Crowd and Movies datasets	Violence	Fast detection of violent content in real-time videos, with high accuracy	Violence not identified temporally and spatially.	accuracy: 97.50% (Hockey dataset), 99.80% (Movies dataset), and 93.40% (Crowd dataset)
LSVM, HOG and OpenCV [20]	iBUG 300-W dataset	Language	Audio extracted from video and converted to text to check profane words using binary classification, lips pixalated and audio silenced on detection	Temporal nature of the dataset not utilised, situation with multiple people in the frame not addressed.	accuracy: 82.35%
Resnet50 [17]	Local datasets	Language	High performance of binary profanity classification, demonstrated the usability of CNN spectral image of audio data.	Temporal nature of the dataset not utilised.	ER: 1.24, F1 score: 98.54%

similar performance metrics they have shown to validate the performance of their work. In cases, the works that do not address AFCR directly but address detection or classification of different content types have also been considered. Upon comparison of the metrics, a comparative analysis can be incorporated to determine the state-of-the-art AFCR method.

A number of prominent ML-based AFCR models having different neural network architectures, the datasets, content-types, and performance metrics of the models have been investigated. Classification criteria, promises and limitations, and the performance metrics and their values for each of the considered models are analysed and summarised in Table. I. The table represents the unique features of the widely used approaches for detecting and classifying different components that are considered for AFCR using ML. Along with the features, the performance of the models and their limitations have also been illustrated.

V. Performance of the AFCR Models

In this section, the comparative study of the performance and effectiveness of various ML-models is presented. An AFCR model depends directly on the accurate detection and classification of the identified content types: violence, profanity, substance abuse, nudity, and adult content. It is necessary to determine the intensity of these contents in a film to generate their respective rating. As we observed in Table. I, the consideration of the performance metrics were not the same, and most of the metrics can be represented directly or derived into accuracy and F-score. To learn the merit of those models, we have re-evaluated the performance and transferred their performance on a similar ground for their fair comparison of the performance. Particularly, we have used false positive (FP), false negative (FN), precision (P) and recall (R) values and hence plotted Fig. 3 using the F-score (F) values using equation (1) and the accuracies are shown in the Table I.

$$F = 2 \times \frac{P * R}{P + R} \times 100\% \quad (1)$$

We observe from Fig. 3 that the DL-based models generally offer higher accuracy. The CNN-based architectures have been widely used for nudity, adult content, and violence detection or classification. From the work of Aldahoul *et al*. [15], it is observed that even after using CNN, despite high accuracy, the models had a high rate of false-positive and false-negative outcomes, which denote the inconsistency in the accuracy for activity recognition and classification, *e.g*. adult content, violent detection *etc*., since the temporal nature of the video dataset has not been considered. Besides, these approaches have been used for binary classifications, and the intensity of the contents cannot be determined this way effectively.

The relatively new RNN based architectures also help analyse the sequential properties and give better results when combined with other ML/DL architectures. As seen from Daniel *et al*. [1], the multimodal approach can increase the efficiency of the architecture to determine the intensity of the contents in films to execute AFCR. Again, the shortage

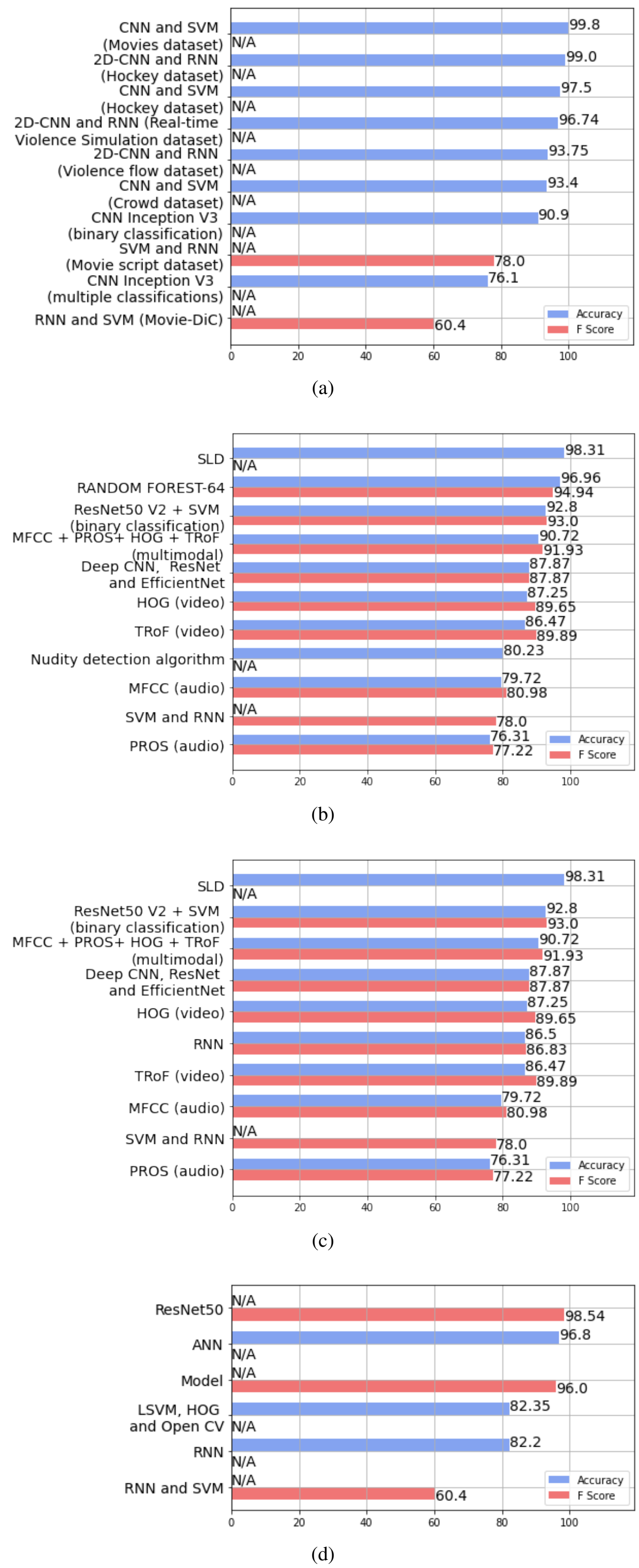


Fig. 3: Performance of the state-of-the-art models for the content types: (a) violence, (b) nudity, (c) adult content, and (d) language respectively.

of datasets and approaches for substance abuse detection in addition to the necessity of the work itself is open to ML-based research.

VI. Conclusion

Film Censorship and Rating is a necessary and significant protocol ensured to maintain the film industry's standard and prevent the exposure of an uncongenial audience to inappropriate content beyond the amount they are supposed to be exposed to, based on their age group, ethnicity and various other factors. This is a time-consuming and tiresome task and is often prone to mistakes due to fatigue, bias, *etc*. An ML-based AFCR system can effectively censor and rate films in the shortest possible time and the most effective way avoiding many humane limitations. Our work reviewed and compared the state-of-the-art methods and datasets, set the criteria, and identified the content types to be considered for AFCR using ML. We have observed promising scopes of further research and development in this field. Especially due to improved computational capacity and suitable datasets, concepts like sequential analysis and multimodal analysis have introduced enormous opportunities for future research works in this field.

References

[1] D. Moreira, S. Avila, M. Perez, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, "Multimodal data fusion for sensitive scene localization," *Information Fusion*, vol. 45, pp. 307–323, 2019.

[2] H. S. Lyn, S. Mansor, N. AlDahoul, and H. A. Karim, "No. 5 convolutional neural network-based transfer learning and classification of visual contents for film censorship," *Journal of Engineering Technology and Applied Physics*, vol. 2, no. 2, pp. 28–35, 2020.

[3] G. Gajula, A. Hundiwale, S. Mujumdar, and L. Saritha, "A machine learning based adult content detection using support vector machine," in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2020, pp. 181–185.

[4] M. P. Association and N. A. of Theatre Owners. (2020) Classification and rating rules. Accessed on: 11 August, 2021. [Online]. Available: https://www.filmratings.com/Content/Downloads/rating_rules.pdf

[5] A. K. Pour, W. C. Seng, S. Palaiahnakote, H. Tahaei, and N. B. Anuar, "A survey on video content rating: taxonomy, challenges and open issues," *Multimedia Tools and Applications*, pp. 1–25, 2021.

[6] J. Cifuentes, A. L. S. Orozco, and L. J. G. Villalba, "A survey of artificial intelligence strategies for automatic detection of sexually explicit videos," *Multimedia Tools and Applications*, pp. 1–18, 2021.

[7] W. Han and M. Ansingkar, "Discovery of elsagate: Detection of sparse inappropriate content from kids videos," in *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*. IEEE, 2020, pp. 46–47.

[8] M. Gruosso, N. Capece, U. Erra, and N. Lopardo, "A deep learning approach for the motion picture content rating," in *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2019, pp. 137–142.

[9] A. S. B. Wazir, H. A. Karim, M. H. L. Abdullah, and S. Mansor, "Acoustic pornography recognition using recurrent neural network," in *2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2019, pp. 144–148.

[10] K. Jani, M. Chaudhuri, H. Patel, and M. Shah, "Machine learning in films: an approach towards automation in film censoring," *Journal of Data, Information and Management*, vol. 2, no. 1, pp. 55–64, 2020.

[11] M. B. Garcia, T. F. Revano, B. G. M. Habal, J. O. Contreras, and J. B. R. Enriquez, "A pornographic image and video filtering application using optimized nudity recognition and detection algorithm," in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*. IEEE, 2018, pp. 1–5.

[12] D. C. Moreira and J. M. Fechine, "A machine learning-based forensic discriminator of pornographic and bikini images," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[13] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A deep learning approach to detect abusive bengali text," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*. IEEE, 2019, pp. 1–5.

[14] M. Shafaei, N. S. Samghabadi, S. Kar, and T. Solorio, "Rating for parents: Predicting children suitability rating for movies based on language of the movies," *arXiv preprint arXiv:1908.07819*, 2019.

[15] N. Aldahoul, H. A. Karim, M. H. L. Abdullah, A. S. B. Wazir, M. F. A. Fauzi, M. J. T. Tan, S. Mansor, and H. S. Lyn, "An evaluation of traditional and cnn-based feature descriptors for cartoon pornography detection," *IEEE Access*, vol. 9, pp. 39 910–39 925, 2021.

[16] H. Chin, J. Kim, Y. Kim, J. Shin, and M. Y. Yi, "Explicit content detection in music lyrics using machine learning," in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2018, pp. 517–521.

[17] A. S. B. Wazir, H. A. Karim, M. H. L. Abdullah, S. Mansor, N. AlDahoul, M. F. A. Fauzi, and J. See, "Spectrogram-based classification of spoken foul language using deep cnn," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2020, pp. 1–6.

[18] M. Susanty, A. F. Rahman, M. D. Normansyah, A. Irawan *et al.*, "Offensive language detection using artificial neural network," in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*. IEEE, 2019, pp. 350–353.

[19] A. R. M. Guedes and G. C. Chávez, "Real-time violence detection in videos using dynamic images," in *2020 XLVI Latin American Computing Conference (CLEI)*. IEEE, 2020, pp. 503–511.

[20] A. Chaudhari, P. Davda, M. Dand, and S. Dholay, "Profanity detection and removal in videos using machine learning," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2021, pp. 572–576.

[21] A. Traoré and M. A. Akhloufi, "Violence detection in videos using deep recurrent and convolutional neural networks," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020, pp. 154–159.

[22] V. R. Martinez, K. Somandepalli, K. Singla, A. Ramakrishna, Y. T. Uhls, and S. Narayanan, "Violence rating prediction from movie scripts," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 671–678.

[23] J. Mallmann, A. O. Santin, E. K. Viegas, R. R. dos Santos, and J. Geremias, "Ppcensor: Architecture for real-time pornography detection in video streaming," *Future Generation Computer Systems*, vol. 112, pp. 945–955, 2020.

[24] H. M. Heyman, F. Senejoux, I. Seibert, T. Klimkait, V. J. Maharaj, and J. J. M. Meyer, "Identification of anti-hiv active dicaffeoylquinic- and tricaffeoylquinic acids in helichrysum populifolium by nmr-based metabolomic guided fractionation," *Fitoterapia*, vol. 103, pp. 155–164, 2015.

[25] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human actions classes from videos in the wild," *arXiv preprint arXiv:1212.0402*, 2012.

[26] Youtube-8m: A large and diverse labeled video dataset for video understanding research. Accessed on: 11 August, 2021. [Online]. Available: https://research.google.com/youtube8m/

[27] A. P. B. Lopes, S. E. de Avila, A. N. Peixoto, R. S. Oliveira, M. d. M. Coelho, and A. d. A. Araújo, "Nude detection in video using bag-of-visual-features," in *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*. IEEE, 2009, pp. 224–231.

[28] C.-H. Demarty, C. Penet, M. Soleymani, and G. Gravier, "Vsd, a public dataset for the detection of violent scenes in movies: design, annotation, analysis and evaluation," *Multimedia Tools and Applications*, vol. 74, no. 17, pp. 7379–7404, 2015.

[29] S. Karavarsamis, N. Ntarmos, K. Blekas, and I. Pitas, "Detecting pornographic images by localizing skin rois," *International Journal of Digital Crime and Forensics (IJDCF)*, vol. 5, no. 1, pp. 39–53, 2013.