International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 236-241

Detection of State Of Aggression among Adolescents: ALiterature Review

Vanishree S[1], Rakshitha S[2], Vinayak S[3], Rajath MR[4], Sahana V[5]
[1,2,3,4](Student, Department of ISE, JSS Academy of Technical Education, Bangalore, India)
[5](Assistant Professor, Department of ISE, JSS Academy of Technical Education, Bangalore, India)

Abstract: *Social media has become a huge source of online aggression. Teenage anxiety plays an importantrole in influencing their mental health. Because of the growing popularity of social media platforms, which provide confidentiality, accessibility, and the capacity to create online groups and debate, detecting hate speech and tracking it becomes a challenging issue for society, individuals, policymakers, and researchers. Various typesof aggressive behavior are trying to be detected by using several machines and deep learning approaches. Aggressive behavior is evolving over time in fast-paced social media and, generating increasing content. This paper presents a survey on aggression among adolescents. The amount of offensive speech on social media continues to rise as social media content grows. Methods for automatically detecting offensive languages are essential due to the web's huge volume. In this study, we cover the main subjects that have been researched in order to automatically recognize these types of utterances. This work presents a thorough assessment of the literature in this field, focusing on machine learning and deep learning technologies, and highlighting terminology, processing pipelines, and basic methodologies used. In this paper, we explained the valuable works of previous researchers who have dedicated their valuable time to studying aggressive behavior and detecting it using diverse methodologies.*

Keywords – *Machine Learning, Systematic Review, Deep Learning, Aggressive Behavior*

I. Introduction

The internet has made it simple for us to connect with individuals or organizations that we are interested in. The social media sites have reached a significant number of individuals in society as a result of the growth of various technologies such as high-speed internet and portable devices. The great majority of social network users are under 30 years of age. Researchers have taken advantage of the vast amounts of data available on various social networking sites and performed detailed research in a variety of fields. Sentiment Analysis is a popular field of study that utilizes a variety of data from social media. Using data from websites, a variety of studies have been conducted to determine the sentiment surrounding a given product or service.

These difficulties might range from political beliefs to religious beliefs, as well as gender, caste, and other factors. Because of this mismatch of ideas, unpleasant material is shared on social networking platforms. Hate speech and abusive content have grown commonplace on social media platforms, and they regularly generate societal upheaval. There have been reports of riots breaking out in various cities, with social media posts being the primary source of riot spread.

Hate speech is defined as the trade of verbal or nonverbal facts between users who have a high level of intolerance and anger. Aggression can come in different forms, such as user activity on social media platforms that may include unparliamentary language. It is also possible to abuse a person based on their sexual orientation, politics, race, or religious convictions. People's self-esteem is often lowered as a result of these exchanges of harsh words, which can have a harmful influence on society. The spread of offensive languages has become a global phenomenon.

Several recent studies have confirmed the correlation between increased online hate speech content and hate crimes including the election of Donald Trump in the US, the Manchester and London bombings in the UK, and terrorist acts in New Zealand. The European Union Commission has taken a number of actions, including law, to combat the negative repercussions of hate speech. In an EU hate speech rule, social media companies are required to remove hate speech information within 24 hours of being made aware of it by the European Union Commission. In contrast, identifying and removing aggressive content is a laborious, time-consuming process. As a result of these concerns and widespread hatred Other Related Concepts from the above definitions and contents analysis, it is clear that some elements are highly related to hate speech (e.g., racism, violence, gender discrimination, etc.). Moreover, we have found several previous works that have presented significant branches of hate speech.

II. Approaches For Aggressive Behavior Detection

Approaches for Aggressive Behavior Detection

Machine Learning Approaches

The idea behind Machine Learning is to create computer programs that can process data and learn on their own without being explicitly programmed. Machine learning enables computers to gain knowledge without being explicitly programmed. Data science has evolved into one of the most important disciplines in the world today. Statistical approaches are used to train algorithms to derive classifications or predictions that are then analyzed to gain insight into computer data mining. Machine learning classifiers are divided into three groups based on the way decisions are made within applications and industries. There are three types of machine learning: a) Supervised machine learning b) Unsupervised machine learning c) Semi-supervised machine learning.

Deep Learning Approaches

The deep learning technique uses neural networks layered three or more times to model the activity of the human brain by receiving knowledge from enormous quantities of data. Multiple convolution layers can assist in optimizing and tuning the accuracy of a single-layer neural network. DL is an important factor of self-driving automobiles, enabling voice commands in consumer electronics, fraud detection and many more. A computer model learns how to categorize images, words, or audio directly from these inputs, achieving cutting-edge accuracy and sometimes even surpassing human performance.

Hybrid Approaches

Hybrid Learning is an integration of deep learning and machine learning approaches. It uses the advantages of both approaches to overcome the shortcomings in each of them, resulting in more accurate solutions.

III. Tabulation Of Acronyms

Table 1 lists the alphabetical list of acronyms used in this paper along with their full form in order to facilitate understanding of various different techniques reviewed in this paper.

Table 1 Acronyms and their full forms

Shortenedform	**Full form**	**Shortenedform**	**Full form**
ANN	Artificial Neural Network	**BERT**	Bidirectional Encoder Representations from Transformers
Bi-LSTM	Bi-directional Long Short Term Memory	**CNN**	Convolutional Neural Networks
DCNN	deep convolutional neural network	**DNN**	Deep neural networks
ELMo	Embeddings from Language Models	**GP**	Genetic Programming
LR	Logistic Regression	**LSTM**	Long Short Term Memory
MLP	multilayer perceptron	**NLP**	Natural language processing
RF	Random forest	**RNN**	recurrent neural network
SVM	Support Vector Machine	**TEC**	Twitter Emotion Corpus, Ekman

IV. Related Work

Review of related papers

Machine Learning Approaches

In 2021, Mona Khalifa et al [4] proposed an approach to determine the accuracy of the datasets using GP models and binary classification techniques. GP is an implementation of an Evolutionary Algorithm (EA) that belongs to machine learning. EAs are used to find direct answers to issues that people are unable to address. Because the training dataset is labelled, Binary Classification falls under the umbrella of Supervised Learning. And, as the name implies, it is merely a special case with only two classes. In the hybrid mutation technique, the GP achieved the best score with 77.33%, and the binary classification technique achieved the best score with 94% F1-score.

In 2021, Arathi Unni et al [5] developed a method for detecting cyberbullying comments using ensemble learning. To classify comments, various supervised ensemble classification methods are utilised like In this SVM, LR, and Perceptron models to predict the outcome. This model correctly detects cyberbullying remarks 94% of the time.

In 2020, Herodotos Herodotou et al [6] performed LR analysis for the datasets to calculate the accuracyusing the cross-validation methods. One of the most commonly used Machine Learning algorithms is logistic regression. It is used to predict a categorical dependent variable based on a set of independent factors. Hence, theresult is either discrete or categorical. The system gives probabilistic values to answer questions rather than specific answers like 0 or 1. For sarcasm, an accuracy of 93% was reported, while for offensiveness, an accuracyof 74% F1 score was reported.

In 2019, Dmitriy Levonevskiy et al [7] proposed machine learning techniques to estimate aggressiveness in Russian Texts. As for recognizing aggressive messages, they used the TEC dataset (Twitter Emotion Corpus, Ekman) with words both in English and Russian. For identifying aggressive messages, they achieved the best results using TEC(65%).

Deep Learning Approaches

In 2021, Sreekanth Madisetty et al [8] presented an analysis of the issue of aggression detection on social media. Among the deep learning methods, they employed were CNN which had five layers, LSTM, and Bi-LSTM. For social media posts, the proposed system achieved an F1-score of 0.508 while for Facebook posts, it achieved a score of 0.604.

In 2020**,** Saima Sadiq et al [9] proposed a deep neural model-based technique to detecting aggression on Twitter. In this paper, The multilayer perceptron is used to implement a deep neural network architecture. A MLP is an ANN with many layers. An MLP is a sequence of fully linked layers that make up a deep neural network. Cyber-trolls were categorized into two categories, 1 being Cyber-Aggressive (CA) and 0 being Non-Cyber-Aggressive (NCA). Out of 20,001items, 7822 falls into the cyber aggressive category, while 12,179 fall into the non-aggressive category. To improve the model's accuracy, the author used a 10-fold cross-validation method. In the best-case scenario, 92 percent of the accuracy, 90 percent precision, recall, and F1-score were achieved.

In 2020, Yanling Zhou,Yanyan Yang et al [10] described a Deep Learning-based method for detecting hate speech. Three text classification techniques are discussed here: ELMo, BERT, and CNN and each of them are applied to the detection of hate speech, with enhanced performance from fusion. The mean F1-score of final fusion was found to be 0.704 and accuracy 0.750.

In 2020, Pradeep Kumar Roy et al [11] based on the DCNN and LSTM model, an approach to the detection of hate speech was proposed. The deepest type of convolutional neural networks used for image and video pattern recognition is the CNN or DCNN. Traditional artificial neural networks have developed into DCNNs, which use a three-dimensional neural pattern inspired by animal visual brain. RNNs such as the LSTM can learn long-term relationships. In order to prevent the problem of long-term reliance, LSTMs are specifically designed. They collected the datasets from Twitter and found the F1-score of the dataset to be 0.59

In 2019, Ana-Sabina Uban et al [13] proposed Detection of Abusive Language Online using CNN for sentence classification and NLP for tasks. Images are typically analysed using a CNN, which is a deep neural network for analyzing structured data sets. CNNs have become instrumental in many visual applications such as image categorization and natural language processing for text tagging, as well as computer vision and computer graphics. Three different datasets were used namely aggressive language dataset, offensive language dataset and sentiment dataset. The f1-score of the aggressive language dataset is 65.24 and the f1-score of the offensive language dataset is 57.24.

Hybrid Approaches

In 2021, Kirti Kumari et al [16] proposed a technique for automatically detecting cyber-aggression in online social networks is based on LSTM autoencoding followed by a machine learning classifier. LSTMs were created expressly to avoid the issue of long-term dependency. They collected the datasets from Facebook and Twitter and automated those data sets, they applied classifier algorithms and found the f1-score of the Facebook dataset to be 0.81 and the f1-score twitter dataset to be 0.71.

In 2021, Girish V P et al [17] conducted a project which aimed at detecting hate speech and categorize them into various classes using machine learning and lexicon-based approach. To find the most effective classifier, the results of several classifiers were compared. They created a webpage using flask where a word or a sentence could be typed and after selection of classifier the result was displayed. The Support Vector Machine Classifier had the best accuracy and F1-score.

In 2021, Cach N. Dang et al [18] used hybrid methods to analyze sentiment. They combined LSTM networks, CNN, and SVM and reviewed various datasets from different domains. The hybrid model increased the accuracy compared to individual models on the datasets.

In 2021, György Kovács et al [19] used the cross-validation method, CNN-LSTM, and FastText to train and evaluate the models. Cross-validation is a technique for assessing the model's efficiency by training it on a

portion of input data and testing it on another subset of input data that has never been seen before. The use of CNNs is often advantageous for such tasks as geographic data analysis, computer vision, image recognition, signal processing, natural language processing, and several other range of applications. The weighted-f1score of the CNN-LSTM model is 0.8063 and macro-f1 is 0.7486.

Comparative Study

Table 2 Machine Learning Approach

Paper	**Year**	**Platform**	**Features and algorithms**	**Precision**	**Recall**	**Accuracy**	**F1-Score**
[4]	2021	Twitter	GP and binary classification	-	-	-	Generic programming P=77.33% Binary classification =94%
[5]	2021	Kaggle, YouTube and Twitter	Perceptron, LR, and SVM	-	-	94%	-
[6]	2020	Twitter	LR and cross-validation	-	-	71%	-
[7]	2019	Russian Twitter corpus, TEC (Twitter Emotion Corpus, Ekman)	MachineLearning	-	-	65%	-

Table 3 Deep Learning Approach

Paper	**Year**	**Platform**	**Features and algorithms**	**Precision**	**Recall**	**Accuracy**	**F1-Score**
[8]	2021	Facebook, SocialMedia	CNN, LSTM and Bi-LSTM	-	-	-	Social media posts= 0.508 and Facebook posts= 0.604
[9]	2020	Twitter	DNN	92%	90%	90%	90%
[10]	2020	Twitter	ELMo, BERT, and CNN	-	-	75%	70.4%
[11]	2020	Twitter dataset	DCNN and LSTM	0.67	0.53	92-95%	0.59
[13]	2019	Facebook andTwitter	CNN and NLP	64.54%	75.35%	57.52%	65.24%
[14]	2018	Facebook Pages and Twitter	CNN	57%	59%	73.2%	58%

Table 3 Hybrid Approach

Paper	**Year**	**Platform**	**Features and algorithms**	**Precision**	**Recall**	**Accuracy**	**F1-Score**
[16]	2021	Facebook and Twitter	LSTM and ML classifier	-	-	-	Facebook=0.81 Twitter=0.71.

[17]	2021	Facebook, Instagram and Twitter	Lexicon Based Approach – Vader, Text Blob Machine Learning Approach – SVM, LR	-	-	31.5% 28.3% 40.6% 41.40%	31.2% 24.6% 38.2% 35.3%
[18]	2021	Twitter	LSTM, CNN, and SVM	84.0%	91.0%	93.4%	93.4%
[19]	2021	Twitter and Facebook	CNN-LSTM	-	-	-	0.8063

Figure 1 shows the different techniques reviewed in this paper, CNN algorithm is used in most of the papers, followed by LSTM and SVM algorithms.

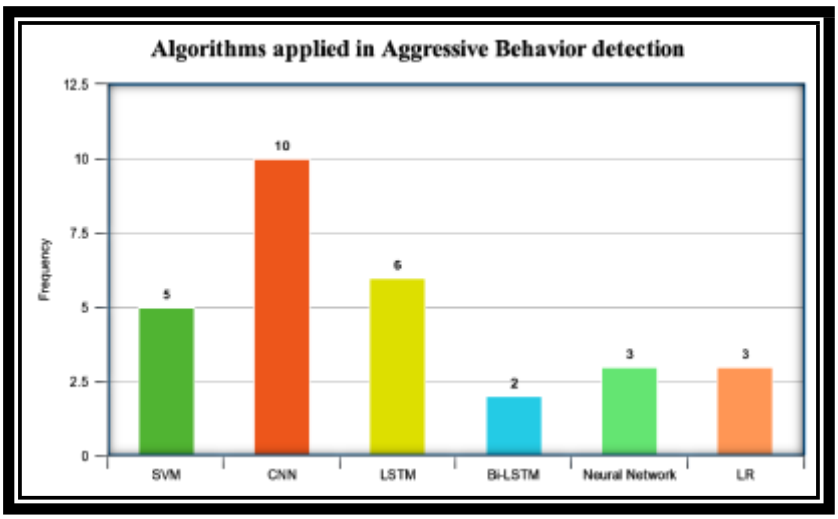

Fig 1. Algorithms applied in aggressive behavior detection among the papers

V. Limitation

This paper presents a comprehensive research study on the automatic identification of offensive language and cyberbullying. The majority of the datasets utilized are unbalanced, which has an impact on the classifiers' accuracy. Significantly among the younger generation, language changes with time. New vocabulary enters the linguistic culture on a constant basis. As a result, researchers are encouraged to develop flexible algorithms for detecting new lingo and abbreviations associated with cyberbullying on social media networks. In this detailed research study, we analyse that the performance of the classifiers is affected due to imbalance datasets and to balance the dataset, data resampling can be used. Different learning algorithms produce varying results, confirming the importance of determining an appropriate learning algorithm, one that is tailored to the problem domain in which the text classification is being done. Offensive speech has different categories according to gender, sexual identity, nationality, historical events , religious beliefs and so on, all this types of words cannot be identified using a particular dataset.

VI. Conclusion

The paper describes a survey for detecting aggressive communications using automated text classification approaches, we have reviewed existing literature to detect aggressive behavior by using deep learning approaches. This paper looked at three approaches to identify cyberbullying communications - deep learning approaches, machine learning approaches, and hybrid approaches. The findings of this study are important because they will be used as a baseline against which future research in different automatic text categorization algorithms for automatic offensive speech identification will be compared. A number of different sorts of discriminative features that have been utilized to detect cyberbullying on online social networking sites have also been discussed. Furthermore, the most effective deep learning algorithms for categorizing cyberbullying texts in online social networking sites were determined. To build extremely effective and reliable cyberbullying

detection models, a significant amount of research is necessary. We hope that the current study will provide important information and guide to new paths in the field of recognizing violent human behavior, including cyberbullying detection on social networking sites. Offensive speech isn't just confined to texts; other forms of engagement, such as picture and video detection, can also be used to focus on the future.

References

[1] Mohiyaddeen, M., & Siddiqui, S. (2021). "Automatic Hate Speech Detection: A Literature Review". Available at SSRN 3835825, DOI: https://dx.doi.org/10.2139/ssrn.38873833

[2] Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). "Hate speech: A systematized review". Sage Open, 10(4), 2158244020973022, DOI: https://doi.org/10.1177/2158244020973022

[3] Saroar Jahan, M., & Oussalah, M. (2021). "A systematic review of Hate Speech automatic detection using Natural Language Processing". arXiv e-prints, arXiv-2106, DOI: https://doi.org/10.48550/ arXiv.2106.00742

[4] Aljero, M. K. A., & Dimililer, N. (2021). "Genetic Programming Approach to Detect Hate Speech in Social Media".Ieee Access, 9, 115115-115125, DOI: https://doi.org/10.1109/ACCESS.2021.310 4535

[5] Arathi Unni, Ranimol K R, Linda Sebastian, Rajalakshmi S, Sissy Siby, 2021, "Detecting the Presence of Cyberbullying using Machine Learning," International Journal Of Engineering Research & Technology (IJERT) NCREIS – 2021 (Volume 09 – Issue 13).

[6] Herodotou, H., Chatzakou, D., & Kourtellis, N. (2020, December). "A Streaming Machine Learning Framework for Online Aggression Detection on Twitter". In 2020 IEEE International Conference on Big Data (Big Data) (pp. 5056- 5067). IEEE, DOI: https://doi.org/10.1109/BigData50022.2020 .9377980

[7] Levonevskiy, D., Malov, D., & Vatamaniuk, I. (2019, August). "Estimating Aggressiveness of Russian Texts by Means of Machine Learning". In International Conference on Speech and Computer (pp. 270-279). Springer, Cham.

[8] Madisetty, S., & Desarkar, M. S. (2018, August). "Aggression detection in social media using deep neural networks". In Proceedings of the first workshop on trolling, aggression and cyberbullying (TRAC-2018) (pp. 120-127). https://aclanthology.org/W18-4415

[9] Sadiq, S., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., & On, B. W. (2021). "Aggression detection through deep neural model on twitter". Future Generation Computer Systems, 114, 120-129.

[10] Zhou, Y., Yang, Y., Liu, H., Liu, X., & Savage, N. (2020). "Deep learning based fusion approach for hate speech detection". IEEE Access, 8, 128923-128929, DOI: https://doi.org/10.1109/ACCESS .2020.3009244

[11] Roy, P. K., Tripathy, A. K., Das, T. K., & Gao, X. Z. (2020). "A Framework for Hate Speech Detection Using Deep Convolutional Neural Network". IEEE Access, 8, 204951-204962, DOI:https://doi.org/10.1109/ACCESS.2020.3037073

[12] KopparthiHarika, M., Mounika, I. C., Anuradha, T., & Sharon, P. (2020). "Hate Speech Detection in Tweets using Machine Learning Algorithm". International Journal of Engineering Applied Sciences and Technology, 4(12), 558-561.

[13] Uban, A. S., & Dinu, L. P. (2019, June). "On transfer learning for detecting abusive language online". In International Work-Conference on Artificial Neural Networks (pp. 688-700). Springer, Cham.

[14] Singh, V., Varshney, A., Akhtar, S. S., Vijay, D., & Shrivastava, M. (2018, October). "Aggression detection on social media text using deep neural networks". In Proceedings of the 2nd Workshop on Abusive Language Online (ALW2) (pp. 43-50), DOI: http://dx.doi.org/10.18653/v1/W18-5106

[15] Yenala, H., Jhanwar, A., Chinnakotla, M. K., & Goyal, J. (2018). "Deep learning for detecting inappropriate content in text". International Journal of Data Science and Analytics, 6(4), 273-286, DOI: https://doi.org/10.1007/s41060-017- 0088-4

[16] Kumari, K., Singh, J. P., Dwivedi, Y. K., & Rana, N. P. (2021). "Bilingual Cyber-aggression detection on social mediausing LSTM autoencoder". Soft Computing, 1-14, DOI: https://doi.org/10.1007/s00500-021-05817-y

[17] Girish V P, Namratha Bhat, Bhavani B S, Abhin K V, Mrs. Sahana V (2021). "Detection and classification of hate speech". Vol. 5, Issue 11, ISSN no. 2455-2143. DOI: doi.org/10.33564/IJEAST.2021.V05I11.032

[18] Dang, C. N., Moreno-García, M. N., & De la Prieta, F. (2021). "Hybrid Deep Learning Models for Sentiment Analysis". Complexity, 2021, DOI: https://doi.org/10.1155/2021/9986920

[19] Kovács, G., Alonso, P., & Saini, R. (2021). "Challenges of Hate Speech Detection in Social Media". SN Computer Science, 2(2), 1-15, DOI: https://doi.org/10.1007/s42979-021-00457-3

[20] Tommasel, A., Rodriguez, J. M., & Godoy, D. (2018, August). "Textual aggression detection through deep learning".In Proceedings of the first workshop on trolling, aggression and cyberbullying (TRAC-2018) (pp. 177-187).