A Study of Cyberbullying Detection and Classification Techniques: A Machine Learning Approach

Srinadh Unnava
Department of CSE, GITAM School of Technology, GITAM Deemed to be University, India
sunnava@gitam.in (corresponding author)

Sankara Rao Parasana
Department of CSE, GITAM School of Technology, GITAM Deemed to be University, India
sparasan@gitam.edu

Received: 24 April 2024 | Revised: 21 May 2024 | Accepted: 25 May 2024

 DOI: https://doi.org/10.48084/etasr.7621

ABSTRACT

The popularity of online social networks has increased the prevalence of cyberbullying, making it necessary to develop efficient detection and classification methods to mitigate its negative consequences. This study offers a comprehensive comparative analysis of various machine-learning techniques to detect and classify cyberbullying. Using various datasets and platforms, this study investigates and compares the performance of various algorithms, including both conventional and cutting-edge deep learning models. To determine the best practices in various scenarios, this study includes a thorough review of feature engineering, model selection, and evaluation measures. This study also examines how feature selection and data preprocessing affect classification precision and computational effectiveness. This study provides useful information on the advantages and disadvantages of various machine learning algorithms for detecting cyberbullying through experimentation and comparative research. The results of this study can help practitioners and researchers choose the best methods for particular applications and support ongoing efforts to make the Internet safer.

Keywords-cyberbullying; machine learning; deep learning; feature creation and selection; internet; online environment

I. INTRODUCTION

Bullying is a type of cruel behavior in which a person with great social or physical influence frequently mistreats, threatens, or otherwise harms a specific target, usually a less powerful person. Bullying comes in many different flavors. Verbal bullying includes calling someone names, threatening, teasing, etc. Physical bullying includes hitting, fighting, screaming, spitting, tripping, pushing, kicking, pinching, and shoveling. Social bullying is the practice of humiliating someone in public. According to previous studies, 37% of young people in India are victims of cyberbullying and 14% of cases are chronic. The victim of cyberbullying experiences both psychological and emotional effects [1]. Social networks can be used as a medium for cybercrimes and inappropriate online behaviors, such as hacking, fraud and scams, disseminating false information, trolling, online harassment, and cyberbullying, in addition to their many positive aspects.

The prevalence of digital communication channels has increased the significant societal concern of cyberbullying. This phenomenon comprises a range of distinct manifestations of online harassment, intimidation, or victimization, frequently executed through social networks, messaging applications, and other digital mediums. Researchers, politicians, and advocacy groups around the world have shown considerable interest in the detrimental effects of cyberbullying on individuals' mental health, social well-being, and academic achievement. As a result, there is an increasing need for open and efficient methods to identify and address its widespread occurrence. The concept of cyberbullying has become prevalent in the digital age, with profound impacts on people's mental health, well-being, and sense of safety. As technology evolves, the tactics and strategies employed by cyberbullies require ongoing research and intervention efforts to address this growing problem. The timeline of cyberbullying evolution is as follows:

- 1990s - Early internet forums: Cyberbullying appears with harassment on early internet forums and chat rooms. Perpetrators use anonymity to target victims with derogatory messages and rumors.
- 2000s - Rise of social networks: Platforms like MySpace and Friendster popularize social networks, providing new avenues for cyberbullying through public shaming, impersonation, and spreading malicious content.

- 2010s **-** Anonymous messaging apps: Anonymous messaging apps such as sarahah.top and ask.fm gain popularity, facilitating cyberbullying through anonymous messages and feedback loops, leading to increased incidents of harassment and online abuse.
- Mid 2010s **-** Focus on online gaming: Cyberbullying extends to online gaming communities, where players face verbal abuse, harassment, and exclusion based on performance or identity.
- Late2010s **-** Deep fake manipulation: The emergence of deep fake methods enables cyberbullies to create and disseminate fake and manipulated content, further blurring the line between reality and fiction, exacerbating the psychological impact on victims.
- 2020s **-** AI-powered harassment: Advancements in AI enable more sophisticated cyberbullying tactics, including AI-generated harassment and targeted manipulation, posing new challenges to online safety and regulation.
- Future **-** Virtual Reality (VR) bullying: With the proliferation of VR technology, the potential for virtual harassment and immersive cyberbullying experiences becomes a looming concern, highlighting the need for proactive measures to address emerging threats in the digital landscape.

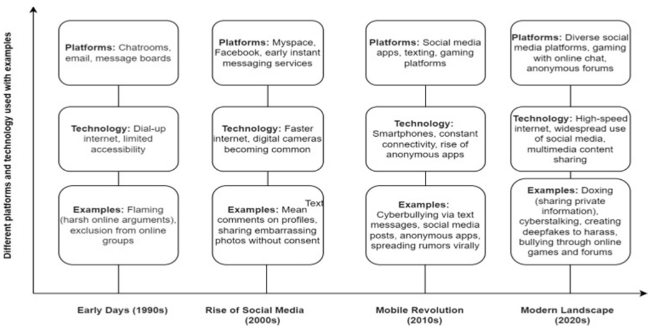

Fig. 1. Evolution of cyberbullying.

The utilization of Machine Learning (ML) methods has become crucial in the fight against cyberbullying, due to their ability to evaluate large amounts of textual and multimedia data to detect patterns indicating harassing conduct [2]. This study aims to assess the effectiveness of various ML-based methods in identifying and categorizing cyberbullying events. This study also aims to gain insight into strengths, limitations, and potential areas for improvement in cyberbullying detection systems by examining and comparing different approaches, algorithms, and datasets used in previous studies.

Section II of this paper presents an overview of the literature review on the available research. Section III presents a comprehensive study of the issues associated with detecting cyberbullying. Section IV provides an overview of the categorization of methods employed in cyberbullying. Section V provides an explanation of the different algorithms employed. Section VI outlines the various research obstacles and potential avenues for further research. SectionVII provides the concluding statement.

II. LITERATURE SURVEY

In [3], a method was proposed for automated cyberbullying detection using the psychological attributes of Twitter users. This method involved collecting, extracting features, and classifying tweets. The final dataset included 5453 tweets. WEKA 3.8 was used for cross-validation and Random Forest (RF) and J48 were used for classification. In [4], a pre-trained BERT model was used to detect cyberbullying. This model generates iterative and job-specific embeddings using the transformer deep neural network. The model consists of 12 layers, with a classifier layer organizing the embeddings. Compared to previous models, it successfully identified cyberbullying instances, providing reliable results. In [5], the TF-IDF vectorizer, Naive Bayes (NB), and SVM models were used to classify tweets in GitHub and Kaggle datasets. The SVM model outperformed NB in accuracy, as it achieved 71.25%. In [6], a supervised learning strategy was introduced to detect cyberbullying. Data were preprocessed, features were extracted using TF-IDF and sentiment analysis methods, and classifiers used n-gram language models. In [7], a deep learning-based approach, Optimized Twitter Cyberbullying Detection (OCDD) is introduced. It uses word vectors as input for CNNs for classification. GloVe approach generates word embeddings, and meta-heuristic optimization is worn for optimal classification. In [8], the authors managed to defend against weak assaults. We can detect cyberbullying using a variety of ML algorithms, some perform better than others and guide us to the optimum method.

In [9], various ML techniques were used to detect cyberbullying in tweets, including NB, KNN, Decision Tree (DT), RF, and Support Vector Machine (SVM), evaluating their accuracy using the Natural Language Toolkit. In [10] an ML method was introduced to identify cyberbullying texts on Twitter using NB and SVM. The study found favorable results when a language model with more n-grams was used, outperforming the NB classifier. In [11], a combination of Natural Language Processing (NLP) and ML was used to recognize aggressive or insulting language in both English and Hinglish. This study focused on the design of techniques to effectively detect abusive and bullying online comments. In [12], an integrated model was proposed that combined feature extraction and classification from social media text datasets. The feature extraction engine extracted CB detection, considering context, user feedback, and psychological traits, while the classification engine classified outputs and then subjected them to an assessment system that could reward or punish them. In [13], a text mining method was proposed using ML algorithms to identify bullying text, with SVM outperforming Bernoulli NB with an overall classification accuracy of 87.14%. In [14], problems were classified into two parts: assessing toxicity and identifying different forms. The proposed ensemble approach achieved the highest accuracy, with an F1-score of 0.828 for harmful/non-toxic classification and 0.872 for toxicity prediction. In [15], a neural network framework was proposed to analyze cyberbullying datasets, comparing 11 classification techniques and examining the impact of feature extraction and NLP on performance.

III. CHALLENGES IN CYBERBULLYING DETECTION

Since online communication is constantly changing and identifying harmful behavior might require subtle nuances, there are various obstacles to detecting cyberbullying. The following are some of these main obstacles:

- Dynamic nature of language: The way people communicate online is dynamic and ever-changing. It is difficult to develop a complete set of guidelines to identify cyberbullying because new phrases, acronyms, and slang are constantly being created.
- Contextual understanding: Cyberbullying frequently involves implicit threats, sarcasm, and context, which makes it challenging for automated algorithms to understand the intended meaning. Distinguishing conduct requires an understanding of the context.
- Multi-modal content: Various media, including text, photos, videos, and memes, can be used to cyberbully. To identify dangerous content in multimodal formats, advanced algorithms are needed to analyze many types of data.
- False positives and negatives: It can be difficult to strike a compromise between reducing false positives that misidentify harmless information as cyberbullying and false negatives that miss real instances of cyberbullying. Overly strict filtering could stifle appropriate dialogues, while insufficient protection could let dangerous content pass unreported.
- Anonymity and pseudonymity: Cyberbullies frequently use pseudonyms or anonymity, making it difficult to find and identify them. Due to their anonymity, some people may feel more confident in acting badly without worrying about the consequences.
- Cultural and linguistic variations: Customs and expressions differ between cultures and languages. A thorough understanding of language and cultural quirks is necessary to modify cyberbullying detection methods to consider these variances.
- Adaptability to new platforms: Systems for detecting cyberbullying must rapidly evolve when new online communities and avenues for communication open up. It can be difficult to create universal solutions because every platform can have a different set of difficulties.
- Legal and ethical concerns: Finding a middle ground between protecting user privacy and keeping an eye out for cyberbullying presents moral and legal challenges. Finding the perfect amount of intervention without violating a person's rights is a difficult task.
- User perception and reporting bias: Users may report cyberbullying in different ways, depending on how they define the term. Training data utilized by detection systems can be affected by reporting biases, which can affect the systems' effectiveness.

In [16, 17] the importance of studying session-based cyberbullying detection during a social media session was highlighted. Some of the key challenges identified in session-based cyberbullying detection are the lack of comprehensive datasets to capture social media sessions and diverse forms of cyberbullying, the definition of session boundaries, and the development of efficient algorithms for real-time detection. Addressing the dynamic nature of cyberbullying, addressing context and sarcasm in text, and ensuring privacy and ethical considerations are also essential. In [18, 19] the challenges of detecting cyberbullying were discussed in conjunction with the available data sources, features, and classification techniques. NLP and ML were highlighted as popular approaches to identify bullying keywords. NLP extracted sentiment analysis, linguistic patterns, and contextual information, while ML used SVM and NB for classification. In [20], a deep transfer learning model was proposed to detect image-based cyberbullying on social networks, achieving an accuracy of 89%. This model was effective in textual posts and has the potential to mitigate cyberbullying concerns. However, this study focused on image-based cyberbullying detection, neglecting textual cyberbullying. This study lacks detailed information on the dataset used, discusses potential challenges in real-world implementation, and does not address ethical considerations such as privacy concerns and false positives/negatives. The model's accuracy may vary depending on hyperparameter settings. In [21], the Graph Convolutional Network (GCN) model demonstrated high efficiency in detecting intricate instances of cyberbullying that specifically target particular attributes of victims. This model surpassed existing baseline methods and successfully addressed the issue of class imbalance in datasets. In [22, 23], botnet detection was used to detect and prevent various malicious activities. In [24], an image edge-preserving classification method was used. Similar ML methods were used in [25, 26] to detect target information, which is very crucial to determine present and future attacks.

IV. CLASSIFICATION OF CYBERBULLYING TECHNIQUES

Cyberbullying can be categorized into many different techniques, depending on the nature and method of harassment or abuse, as shown in Figure 2. A concise overview of several prevalent categorizations includes the following.

- Direct cyberbullying refers to the act of attacking the victim directly through means such as messages, posts, comments, or emails. It includes derogatory remarks, menacing statements, dissemination of false information, or sharing humiliating images or videos.
- Cyberstalking refers to the act of persistently monitoring the online activity of victims, inundating them with an excessive number of communications, or issuing threats. In addition, they may collect personal information to intimidate or exert control over the victim.
- Flaming is the act of engaging in heated online disputes or disagreements using angry and offensive language. It happens frequently on public forums, chat rooms, or social networks.

- Exclusion refers to the deliberate act of intentionally excluding an individual from online groups, chats, or social circles. As a result, the victim may experience a sense of isolation and loneliness.
- Impersonation is the creation of counterfeit profiles or accounts to mimic the victim and disseminate false information, defaming their character or harming their interpersonal connections.
- Outing and trickery encompass the act of divulging private or humiliating facts about the target without their permission or deceiving them into disclosing confidential information that is subsequently exploited against them.
- Cyberbullying by proxy refers to a situation where the bully recruits others to engage in harassment or to spread rumors and gossip about the victim. This might intensify the effects and increase the difficulty for the victim to break out of the cycle of abuse.
- Catfishing is the act of fabricating a false persona to deceive someone into forming an online relationship or acquaintance. It has the potential to cause emotional manipulation, exploitation, and psychological damage.

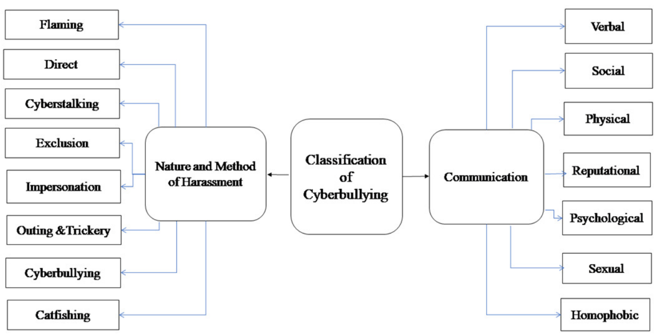

Fig. 2. Classification of cyberbullying based on nature and method of harassment.

Cyberbullying can be categorized into many forms based on the methods and media employed to harass or intimidate the target.

- Verbal cyberbullying refers to the act of using language to inflict harm on the target, typically through means such as text messages, emails, or internet comments. This could involve name-calling, insults, or threats.
- Physical cyberbullying does not include direct contact but refers to forms of online harassment, such as unauthorized access to someone's accounts causing damage to their gadgets or altering their online presence.
- Reputational cyberbullying involves perpetrators intentionally damaging the victim's social standing or credibility by disseminating false information, creating fake profiles, or posting incriminating content [27].
- Social cyberbullying refers to the act of engaging in harmful behavior on social media platforms. This includes activities such as distributing false information, publishing humiliating images or videos, and deliberately excluding someone from online groups or conversations [28].
- Psychological cyberbullying is a form of online harassment that seeks to cause emotional damage to the target using tactics such as gaslighting, manipulation, or coercion. This can result in the victim experiencing anxiety, sadness, or other psychological consequences.
- Sexual cyberbullying refers to the act of sending explicit messages, photographs, or videos without obtaining agreement or participating in online sexual harassment. This behavior can cause people to experience emotions such as humiliation, embarrassment, and trauma.
- Homophobic, racist, or religious cyberbullying refers to the act of singling out someone and subjecting him to hate speech, slurs, or discriminatory activities online due to their sexual orientation, race, ethnicity, or religious beliefs.

V. ALGORITHMS USED TO DETECT AND CLASSIFY CYBERBULLYING

In today's digital environment, cyberbullying is an increasing concern, and numerous methods have been proposed to detect and address it. In [7], the Optimized Twitter Cyberbullying Detection (OCDD) approach was introduced, using a meta-heuristic optimization algorithm and DL to detect cyberbullying on a Twitter dataset, achieving an accuracy of 81.7%. Human intelligence labeled training data and GloVe-generated word embeddings were used for classification. In [29], a DL method was proposed to detect cyberbullying, achieving an accuracy of 84.3%. Despite potential limitations due to code-switching, this study helped to understand the capabilities of DL in identifying cyberbullying and emphasized the need for language use in detection strategies. In [30], a DL-based approach was proposed to detect cyberbullying, achieving 93.97% accuracy [24]. However, this study faced limitations such as skewed data, language dynamics, false positive rates, and limited generalizability. In [31], a DL architecture was introduced to identify cyberbullying instances within Roman Urdu micro-texts. A slang-phrase dictionary was developed, domain-specific stop words were removed, and unstructured data were processed. This study used CNN, RNN-BiLSTM, and RNN-LSTM models, with RNN-BiLSTM achieving an F1-score of 0.67 and 85% validation accuracy in the aggressiveness class. In [32], a sentiment classification model was presented to identify cyberbullying. The proposed model used a CNN for local characteristics, an attention mechanism for character streams, and BiGRU for global context, achieving 91.07% accuracy. In [15], 11 classification approaches, including ML and neural networks, were compared on real-world cyberbullying datasets. Bidirectional neural networks and attention models were accurate, with Bi-LSTM and Bi-GRU being the top two. Limitations included generalizability and data bias. In [20], DL and transfer learning models were combined to detect cyberbullying images on social networks, achieving 89% accuracy. The transfer learning models VGG16 and InceptionV3 achieved 89% accuracy, while the DL 2D-CNN achieved 69.60%. Advantages and drawbacks of existing methods and research gaps

Ref	Advantages	Drawbacks	Research Gaps
[7]	81.7% accuracy using optimization algorithm and DL Twitter cyberbullying detection. CNN was used in text mining, cyberbullying detection was novel.	Requires significant training data and computational resources.	Needs strategies to minimize the requirement for extensive training data and computational resources.
[15]	High accuracy, good results with LR and TF-IDF, GloVe worked well with NN.	Data bias and limited generalization.	Comparing approaches and attributes to increase the accuracy of cyberbullying detection.
[20]	Achieved 89% accuracy, successfully identified most image-based cyberbullying posts.	Limits in text image pairings in cyberbullying postings and textual cyberbullying detection.	Improve the model for identifying text-image combinations and textual cyberbullying.
[29]	Used a stacked embedding method that included BERT and GloVe, improved upon classic ML algorithms, such as LR and SVM, achieved 84.3% accuracy.	Possible areas for development in the domain of code swapping.	Needs methods to enhance the model's functionality in code-swapping scenarios.
[30]	Compared to conventional algorithms, it provides a smarter method to identify cyberbullying. Achieved an accuracy of 93.97% using a CNN algorithm.	High false positive rates, restricted generalizability, biased data, and challenges in capturing contextual knowledge and language dynamics.	Issues with skewed data, ways to enhance the model's contextual comprehension and language dynamics, strategies to decrease false positive rates, and ways to improve the model's generalizability.
[31]	Extensive preprocessing on micro-text data and evaluated models' variety of effectiveness and efficiency.	Limited dataset.	More extensive dataset collection.
[32]	Integrated a CNN layer for local features, a Bi-GRU layer for global context, and an attention mechanism layer for weighting representative words to attain 91.07% accuracy.	The dataset was limited and biased, was computationally intensive, and had mediocre performance on new data.	Decrease computational effort, collect a more complete dataset, fix dataset bias, and improve the model work better with new data.
[33]	89.5% accuracy, considered bilingual data, identified text that contained cyberbullying in several languages.	Lack of contextual awareness, scope limitations, and language barriers.	Increase precision in identifying texts that include cyberbullying in several languages.
[34]	Higher precision using feature extraction techniques and evaluation of characteristics such as malevolent intent, recurring patterns, and abusive language.	Data bias and limited generalization.	Mitigate data bias and enhance the model's generalization.
[35]	Included a wide range of essential features, improved cyberbullying detection.	Limited dataset and dataset bias.	More complete dataset, remove dataset bias, better model explainability.

In [33], a multilingual DL framework was proposed to detect cyberbullying, overcoming linguistic limitations and contextual knowledge issues. CNN-BiLSTM outperformed other models due to its ability to learn global features and long-term dependencies, which makes it crucial to protect users on social networks. In [34], a system was proposed to detect and prevent cyberbullying on social media networks, using supervised ML methods and classifying cyberbullying into classes such as bigotry, sexuality, physical injury, and profanity. The accuracy of the system was improved by feature extraction techniques. In [35], an innovative method using BERT was proposed to detect cyberbullying in social networks, which improved the efficacy of detecting and categorizing cyberbullying.

VI. RESEARCH GAPS AND FUTURE DIRECTION

The field of cyberbullying has a multitude of complex research challenges that hinder successful identification, prevention, and intervention strategies. Several significant challenges are present:

- Underreporting and Lack of Data: Cyberbullying often goes unreported due to fear, humiliation, or uncertainty about authorities' reactions, leading to a significant lack of data in research databases.
- Dynamic Nature of Cyberbullying: Cyberbullying evolves rapidly, making traditional ML models difficult to adapt. Developing real-time algorithms to dynamically learn from emerging cyberbullying patterns presents a significant challenge.
- Anonymity and Pseudonymity: Online anonymity and pseudonymity hinder cyberbullying identification and accountability, as explicit attribution of abusive conduct is lacking, posing challenges in implementing targeted responses.
- Data Collection and Annotation: Cyberbullying detection faces challenges in acquiring labeled datasets for ML models, as accurate labeling requires human judgment and time investment. Additionally, ensuring data quality and dependability is a challenge.
- Class Imbalance: Cyberbullying datasets often have imbalanced classes, over-representing non-bullying events. This can lead to biased model performance, favoring majority correctness over minority cyberbullying.
- Context Sensitivity: Cyberbullying detection requires understanding contextual factors, as seemingly harmless language can indicate it in different environments. However, integrating contextual information into ML models is challenging.
- Multimodal Analysis: Cyberbullying, a complex issue that involves textual, visual, and audiovisual forms, presents a challenge in developing efficient multimodal analysis systems capable of capturing and interpreting information across various modalities.
- Privacy and Ethical Concerns: Cyberbullying detection methods often involve sensitive user data, balancing accuracy with privacy and ethical concerns, making it challenging to ensure that technologies respect user privacy rights.

VII. CONCLUSION

The examination of cyberbullying detection and classification strategies through the application of ML methods underscores the dynamic nature of efforts to address online harassment. After conducting a thorough study, it is clear that ML techniques have promising opportunities to precisely detect and classify cyberbullying incidents. However, the efficacy of these methods fluctuates depending on variables such as dataset quality, feature selection, and algorithm efficiency. Although certain approaches exhibit improved performance in terms of precision, recall, and overall accuracy, there is still potential for improvement and investigation of hybrid or ensemble methods to further enhance detection capabilities. Furthermore, current research efforts should prioritize the resolution of obstacles, such as identifying cyberbullying that is contingent on the context and adjusting to the changing dynamics of online communication. To summarize, although ML provides useful tools for detecting and classifying cyberbullying, there is still a significant amount of work to significantly optimize these strategies. Ongoing studies and collaborative efforts are vital to advance the discipline and ameliorate the detrimental consequences associated with cyberbullying in the era of digital technology.

REFERENCES

[1] M. Arif, "A Systematic Review of Machine Learning Algorithms in Cyberbullying Detection: Future Directions and Challenges," *Journal of Information Security and Cybercrimes Research*, vol. 4, no. 1, pp. 01–26, Jun. 2021, https://doi.org/10.26735/GBTV9013.

[2] D. Van Bruwaene, Q. Huang, and D. Inkpen, "A multi-platform dataset for detecting cyberbullying in social media," *Language Resources and Evaluation*, vol. 54, no. 4, pp. 851–874, Dec. 2020, https://doi.org/10.1007/s10579-020-09488-3.

[3] V. Balakrishnan, S. Khan, and H. R. Arabnia, "Improving cyberbullying detection using Twitter users' psychological features and machine learning," *Computers & Security*, vol. 90, Mar. 2020, Art. no. 101710, https://doi.org/10.1016/j.cose.2019.101710.

[4] J. Yadav, D. Kumar, and D. Chauhan, "Cyberbullying Detection using Pre-Trained BERT Model," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, Jul. 2020, pp. 1096–1100, https://doi.org/10.1109/ICESC48915.2020.9155700.

[5] R. R. Dalvi, S. Baliram Chavan, and A. Halbe, "Detecting A Twitter Cyberbullying Using Machine Learning," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, May 2020, pp. 297–301, https://doi.org/10.1109/ICICCS48265.2020.9120893.

[6] J. Hani, M. Nashaat, M. Ahmed, Z. Emad, E. Amer, and A. Mohammed, "Social Media Cyberbullying Detection using Machine Learning," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019, https://doi.org/10.14569/IJACSA.2019.0100587.

[7] M. A. Al-Ajlan and M. Ykhlef, "Optimized Twitter Cyberbullying Detection based on Deep Learning," in *2018 21st Saudi Computer Society National Computer Conference (NCC)*, Apr. 2018, pp. 1–5, https://doi.org/10.1109/NCG.2018.8593146.

[8] N. Mehendale, K. Rajpara, K. Shah, and C. Phadtare, "A Review on Cyberbullying Detection Using Machine Learning." May 2022, https://doi.org/10.2139/ssrn.4116153.

[9] M. Patidar, M. Lathi, M. Jain, M. Dhakad, and Y. Barge, "Cyber Bullying Detection for Twitter Using ML Classification Algorithms," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, no. 11, pp. 24–29, Nov. 2021, https://doi.org/10.22214/ijraset.2021.38701.

[10] J. O. Atoum, "Cyberbullying Detection Through Sentiment Analysis," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, Dec. 2020, pp. 292–297, https://doi.org/10.1109/CSCI51800.2020.00056.

[11] N. Mehendale, K. Shah, C. Phadtare, and K. Rajpara, "Cyber Bullying Detection for Hindi-English Language Using Machine Learning." May 2022, https://doi.org/10.2139/ssrn.4116143.

[12] N. Yuvaraj *et al.*, "Nature-Inspired-Based Approach for Automated Cyberbullying Classification on Multimedia Social Networking," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–12, Feb. 2021, https://doi.org/10.1155/2021/6644652.

[13] S. Andleeb, R. Ahmed, Z. Ahmed, and M. Kanwal, "Identification and Classification of Cybercrimes using Text Mining Technique," in *2019 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, Dec. 2019, pp. 227–2275, https://doi.org/10.1109/FIT47737.2019.00050.

[14] M. Ibrahim, M. Torki, and N. El-Makky, "Imbalanced Toxic Comments Classification Using Data Augmentation and Deep Learning," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, Dec. 2018, pp. 875–878, https://doi.org/10.1109/ICMLA.2018.00141.

[15] C. Raj, A. Agarwal, G. Bharathy, B. Narayan, and M. Prasad, "Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques," *Electronics*, vol. 10, no. 22, Jan. 2021, Art. no. 2810, https://doi.org/10.3390/electronics10222810.

[16] L. Cheng, Y. N. Silva, D. Hall, and H. Liu, "Session-Based Cyberbullying Detection: Problems and Challenges," *IEEE Internet Computing*, vol. 25, no. 2, pp. 66–72, Mar. 2021, https://doi.org/10.1109/MIC.2020.3032930.

[17] S. R. Basha, J. K. Rani, and J. J. C. P. Yadav, "A Novel Summarization-based Approach for Feature Reduction Enhancing Text Classification Accuracy," *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 5001–5005, Dec. 2019, https://doi.org/10.48084/etasr.3173.

[18] W. N. Hamiza Wan Ali, M. Mohd, and F. Fauzi, "Cyberbullying Detection: An Overview," in *2018 Cyber Resilience Conference (CRC)*, Putrajaya, Malaysia, Nov. 2018, https://doi.org/10.1109/CR.2018.8626869.

[19] S. R. Basha and J. K. Rani, "A Comparative Approach of Dimensionality Reduction Techniques in Text Classification," *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 4974–4979, Dec. 2019, https://doi.org/10.48084/etasr.3146.

[20] P. K. Roy and F. U. Mali, "Cyberbullying detection using deep transfer learning," *Complex & Intelligent Systems*, vol. 8, no. 6, pp. 5449–5467, Dec. 2022, https://doi.org/10.1007/s40747-022-00772-z.

[21] J. Wang, K. Fu, and C. T. Lu, "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection," in *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, Dec. 2019, pp. 1699–1708, https://doi.org/10.1109/BigData50022.2020.9378065.

[22] M. A. Haq and M. A. R. Khan, "Dnnbot: Deep neural network-based botnet detection and classification," *Computers, Materials and Continua*, vol. 71, no. 1, pp. 1729–1750, 2022, https://doi.org/10.32604/cmc.2022.020938.

[23] M. A. Haq, "DBoTPM: A Deep Neural Network-Based Botnet Prediction Model," *Electronics*, vol. 12, no. 5, Jan. 2023, Art. no. 1159, https://doi.org/10.3390/electronics12051159.

[24] S. Merugu, A. Tiwari, and S. K. Sharma, "Spatial–Spectral Image Classification with Edge Preserving Method," *Journal of the Indian Society of Remote Sensing*, vol. 49, no. 3, pp. 703–711, Mar. 2021, https://doi.org/10.1007/s12524-020-01265-7.

[25] S. Merugu, K. Jain, A. Mittal, and B. Raman, "Sub-scene Target Detection and Recognition Using Deep Learning Convolution Neural Networks," in *ICDSMLA 2019*, 2019, pp. 1082–1101, https://doi.org/10.1007/978-981-15-1420-3_119.

[26] S. Merugu, M. C. S. Reddy, E. Goyal, and L. Piplani, "Text Message Classification Using Supervised Machine Learning Algorithms," in

ICCCE 2018, 2018, pp. 141–150, https://doi.org/10.1007/978-981-13-0212-1_15.

[27] S. S. Alzahrani, "Data Mining Regarding Cyberbullying in the Arabic Language on Instagram Using KNIME and Orange Tools," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9364–9371, Oct. 2022, https://doi.org/10.48084/etasr.5184.

[28] K. Wang, Q. Xiong, C. Wu, M. Gao, and Y. Yu, "Multi-modal cyberbullying detection on social networks," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, Jul. 2020, pp. 1–8, https://doi.org/10.1109/IJCNN48605.2020.9206663.

[29] T. Mahlangu and C. Tu, "Deep Learning Cyberbullying Detection Using Stacked Embbedings Approach," in *2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, Johannesburg, South Africa, Nov. 2019, pp. 45–49, https://doi.org/10.1109/ISCMI47871.2019.9004292.

[30] V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2019, pp. 604–607, https://doi.org/10.1109/ICACCS.2019.8728378.

[31] A. Dewani, M. A. Memon, and S. Bhatti, "Cyberbullying detection: advanced preprocessing techniques & deep learning architecture for Roman Urdu data," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, Art. no. 160, https://doi.org/10.1186/s40537-021-00550-7.

[32] Y. Luo, X. Zhang, J. Hua, and W. Shen, "Multi-featured Cyberbullying Detection Based on Deep Learning," in *2021 16th International Conference on Computer Science & Education (ICCSE)*, Lancaster, United Kingdom, Aug. 2021, pp. 746–751, https://doi.org/10.1109/ICCSE51940.2021.9569270.

[33] M. Raj, S. Singh, K. Solanki, and R. Selvanambi, "An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques," *SN Computer Science*, vol. 3, no. 5, Jul. 2022, Art. no. 401, https://doi.org/10.1007/s42979-022-01308-5.

[34] A. Perera and P. Fernando, "Accurate Cyberbullying Detection and Prevention on Social Media," *Procedia Computer Science*, vol. 181, pp. 605–611, Jan. 2021, https://doi.org/10.1016/j.procs.2021.01.207.

[35] A. Desai, S. Kalaskar, O. Kumbhar, and R. Dhumal, "Cyber Bullying Detection on Social Media using Machine Learning," *ITM Web of Conferences*, vol. 40, 2021, Art. no. 03038, https://doi.org/10.1051/itmconf/20214003038.