SN Computer Science (2023) 4:859
https://doi.org/10.1007/s42979-023-02301-2

ORIGINAL RESEARCH

A Survey of Cyberbullying Detection and Performance: Its Impact in Social Media Using Artificial Intelligence

Khateeja Ambareen[1] · **S. Meenakshi Sundaram**[1]

Received: 19 May 2023 / Accepted: 4 September 2023

Abstract
Recently, cyberbullying has become one of the most important topics on social media. Online social media users have recognised this as a severe problem, and in recent years, effective detection models have been developed. This has taken on substantial importance. The numerous forms of cyberbullying on social media are highlighted by this poll. Currently, research is being done to identify cyberbullying using AI approaches. We talk about various machine learning and natural language processing (NLP) methods that are used to identify cyberbullying. Additionally, the difficulties and potential directions for future research in the area of AI detection of cyberbullying have been discussed. Attacks on victims of cyberbullying have surged by 40% in 2020's pandemic season. 20% of the increase in juvenile suicides is attributable to cyberbullying. Attacks involving cyberbullying are expected to reach an all-time high in 2025, according to 60% of experts. 38% of respondents report daily exposure to cyberbullying on social media platforms. Even though many people are aware of cyberattacks, cyberbullying has begun to rise alarmingly. By keeping track of the signs of cyberbullying before it occurs, internet service providers can develop more precise classifications for the behaviour to prevent it. Large data sets can also be processed using deep learning techniques.

Keywords Cyberbullying · Machine learning techniques · Cybercrime detection · Social media · Algorithms · Twitter

Introduction

Cyberbullying has recently been recognized as a national health concern by social media users, and creating a detection model in the modern period has been shown to be scientifically beneficial [1]. On social media, users publish a variety of content, including documents, images, and videos, and engage in online communication. People connect with social media through cellphones or computers. The most popular social media platforms are Facebook, Twitter, Instagram, TikTok, and others. Social media is now utilised for a wide range of objectives, including education and business [2]. Social media is causing a lot of new jobs to be created, which is good for the world economy [1]. Social networking has a lot of benefits, but it also has some drawbacks. In order to hurt the reputations and sentiments of others, malicious users of this medium engage in dishonest and deceptive behavior. Recently, cyberbullying has become one of the most serious phenomena in internet media [3]. Cyberbullying and cyber-harassment are terms used to describe bullying or harassment that occurs online. Cyberbullying and cyberharassment are both forms of online bullying. Due to the widespread adoption of digital and technological advancements in the modern world, cyberbullying has become more common among adolescents [4].

Online social networks (OSNs) play a significant role in facilitating social connection, but they also foster antisocial behavior like trolling, cyberbullying, and hate speech. Cyberbullying is the practise of expressing hostile or hateful remarks through the use of short message services or media platforms found on the Internet. Natural language processing (NLP) for automatic detection is hence the first step towards stopping cyberbullying [5].

This article is part of the topical collection "Advances in Computational Approaches for Image Processing, Wireless Networks, Cloud Applications and Network Security" guest edited by P. Raviraj, Maode Ma and Roopashree H R.

✉ Khateeja Ambareen
khateeja.ambareen@gmail.com

S. Meenakshi Sundaram
1965drsms@gmail.com

[1] Department of CSE, GSSSIETW, Affiliated to VTU, Belagavi, Mysore, India

Published online: 08 November 2023

Cyberbullying in Social Media Platforms

Most people are aware that cyberbullying occurs when someone uses a variety of communication channels, including text messages, emails, social media, and other online platforms, to publish or send hurtful words or comments, including images. Young children and teenagers are commonly the targets of cyberbullying because they are more open to new technology, such as the Internet [5]. A sort of harassment in which one person insults or offends another is described as cyberbullying [6]. Cyberbullying is the willful, persistent, and hostile use of technology to harass or damage another person.

Unlike more conventional types of bullying, cyberbullying can happen every day of the week, at any hour of the day or night. It occurs in a variety of ways, including sending text messages, spreading rumors, and posting embarrassing videos and images on social media [7]. Cyberbullying is different from traditional bullying in that it occurs online, where the message will either be delivered to the victim directly or posted in open forums where anyone can view it. Although these cyberbullies can be anyone, they frequently know their victims [8]. They may be a friend or classmate at times. Modern communication channels offer a wide range of cyberbullying techniques. The following forms of cyberbullying, for instance, have been documented [9–12].

Cyberbullying Types

1. *Flaming*: starting a battle online.
2. *Harassment*: the victim of harassment receives nasty and insulting communications on a regular basis this is the type of cyberbullying that most publications are seeking to stop.
3. *Cyberstalking*: the victim receives offensive or intimidating messages that make them feel endangered.
4. *Masquerade*: the bully presents a false persona.
5. *Trolling*: making offensive comments on social media with the intent of offending other users.
6. *Discredit*: spreading unfavorable rumors about someone else and divulging private information about them in public.
7. *Absence*: the Phenomena of excluding someone from a group of individuals
8. *Petty crime*: the act of taking someone's identity online and fabricating a false profile to deceive others.
9. *Dissing*: the phenomena in which information or opinions about another person are spread with the intention of harming that person's reputation or popularity.
10. *Trickery*: trickery is the act of having someone trust you with their secrets and personal information, only for them to utilize that confidence to disclose that information online to the public.
11. *Frapping*: frapping is the practise of someone using your social media accounts to pretend to be you, publish messages under your name, and lead people astray.

The Impact of Cyberbullying

The pandemic's impacts have led to a 40% spike in cyberbullying attacks.

Youth suicides are on the rise, and 20% of that increase is due to cyberbullying. Attacks involving cyberbullying are expected to reach an all-time high in 2025, according to 60% of experts (Source: Statista.com). Cyberbullying affects more than 38% of persons who use online media platforms every day. In response to the cyberbullying attack, about 25% of kids self-harm as a coping mechanism for the humiliation. If they experience cyberbullying, adolescents between the ages of 12 and 18 are more likely to experience social and health problems in the future in the Journal of Adolescence. Cyberbullying is spreading alarmingly despite the fact that many individuals are aware of cyberattacks. 64 percent of victims who get hostile instant messages claim to have personally interacted with the sender. Online teenagers reported receiving inappropriate forwarding of private messages in nearly one in six cases (15%) (Source: Pew Research Centre). The cyberbullying incidences in India are depicted in Fig. 1. In the year 2019, there were 542 reported cases of cyberstalking, and in the following year, there was a sharp rise to 739 cases, for a total of 1000 cases. The percentage of parents whose children have experienced bullying is shown in Fig. 2 by age. In India, children between the ages of 14 and 18 encounter the largest percentage of cyberbullying incidents—59.9%. The suicide and homicide rates are shown in Fig. 3 as a result of cyberbullying.

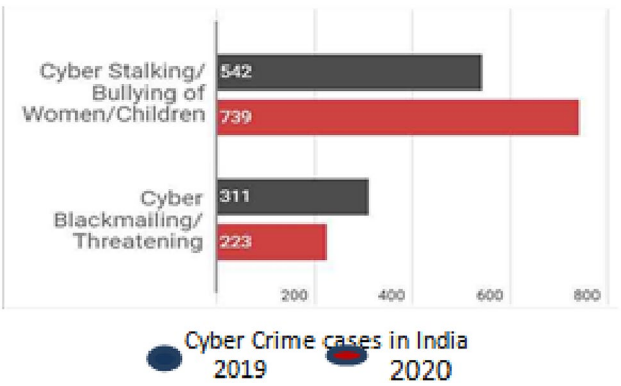

Fig. 1 Cyberbullying incidents in India

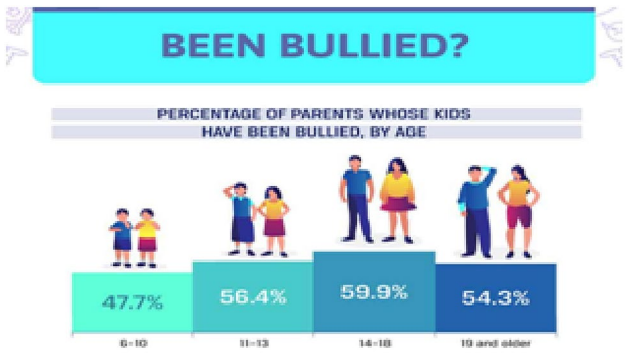

Fig. 2 Cyberbullying percentage by age

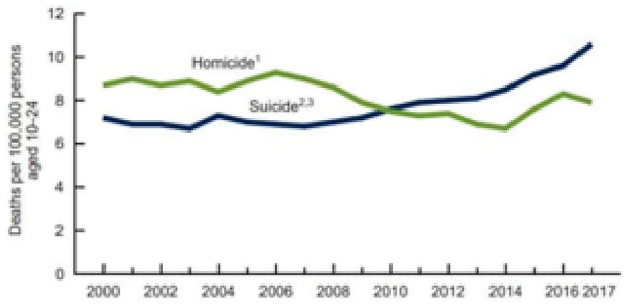

Fig. 3 Suicide and homicide death rates in the USA

Background and Related Work

An effective solution for spotting online abusive and bullying texts is created by combining natural language processing with machine learning techniques. Using distinct properties including phrase frequency, inverse text frequency, and bag-of-words, the accuracy of distinct is tested and used for categorising comments as bullying or non-bullying [1].The study employed Twitter datasets to attempt to suggest an ensemble model by analysing the classification methods used to pinpoint instances of cyberbullying. The suggested ensemble model combines the models of various machine learning classifiers and each of the individual component classifiers on which the datasets are trained to create the predictions in an effort to outperform the individual models. By adding these preliminary forecasts, a final conclusion is reached. Numerous strategies, such as bagging, boosting, alternative stacking, and voting, can be used to make this final forecast. The assessment methods include Adaptive Boosting, Naive Bayes, K nearest neighbours, decision tree, logistic regression, random forest, Bagging classifiers, linear support vector classifier, and stochastic gradient descent [13].

Image-based cyberbullying is recognised using optical character recognition (OCR) technology, and its component effects are assessed on a dummy system. Cyberbullying is automatically detected using natural language processing and machine learning algorithms, and textual data is matched for the corresponding features of the transaction. We developed a semi-supervised method for detecting cyberbullying using the BERT model, which uses the five features that can be utilised to characterise a cyberbullying post or message. The BERT model surpassed traditional machine learning models with an accuracy of 91.90 percent after being trained over two cycles, accounting exclusively for emotive characteristics. The BERT model can generate more accurate findings when a large amount of data is employed [14].

The methods of cyberbullying are investigated using a consolidated deep learning model to automatically detect aggressive behaviour. This method uses three multichannel deep learning models, including convolutional neural networks, bidirectional gated recurrent units, and transformer blocks, to categorise Twitter comments into two groups. Combining data from three well-known datasets on hate speech allowed researchers to assess the effectiveness of the suggested approach. The suggested strategy delivered positive results. The accuracy of the suggested strategy is about 88 percent [5].

Each bully tweet on Twitter has a value of 1, indicating that they have all been positively recognised, according to the results of the machine learning techniques used to detect cyberbullying. Different machine learning models receive an equal amount of tweets from the Twitter dataset that are bullies and non-bullies. With a precision of 91%, recall of 94%, and F1-score of 93%, the logistic regression classifier correctly distinguishes between bullying and non-bullying in tweets. Cyberbullying will not hurt users, thanks to users' ability to prevent it [6].

Different machine learning algorithms for cyberbullying detection are covered, as well as the many cyberbullying categories, data sources, and sources of cyberbullying data for research. The dearth of publicly accessible statistics and the absence of multimedia content-based detection were cited as barriers to cyberbullying detection [8].

Very few people have tested methods that do not entail supervision, thus supervised learning techniques were used to detect cyberbullying. This gives researchers more space to work and gives harassers more room to control their targets. Unattended methods are getting more attention, but supervised methods have historically dominated the detection of cyberbullying. It is also feasible to claim that supervised methods fail to address class disparity, whereas unsupervised ones can. The paper focuses on recent investigations of non-supervised text-based cyberbullying detection and makes recommendations for further research. It calls for more research on unsupervised

approaches and emphasises the seriousness and intensity of harassment in the online environment [15].

The use of machine learning techniques has been made to lessen or stop cyberbullying. These attempts, however, work because they rely on the interactions between the victims. Therefore, it's essential to identify cyberbullying without the victims' participation. In this work, we attempted to analyse this problem using a global dataset of 37,373 unique tweets from Twitter. The seven machine learning classifiers used machine learning techniques such, logistic regression, support vector machine, Ada boost, light gradient boosting machine, stochastic gradient descent, random forest, and naive Bayes. performance of all the aforementioned algorithms was assessed based on various metrics, including F1score, precision, accuracy, and recall, to estimate the global dataset's recognition rates. With an accuracy of about 90.57 percent, the experiment's findings show that Logistic Regression is preferable. The greatest F1 score for logistic was 0.928, the highest recall was 1.0, while the highest precision for stochastic gradient descent was 0.968 [16].

The sentiment analysis of the Twitter data is carried out using ordinal regression. Twitter samples from the Corpus of NLTK, including both good and negative tweets, are used in our research. This improvement was achieved by using Text Blob to determine the polarity of the tweets and lemmatization in place of stemming. To summarise our work, we performed sentiment analysis on Twitter data. The sensation is one of the five categories. Positive and negative tweets from the NLTK Corpus Twitter samples were mined for information. Only the pertinent information was preprocessed for the data extraction using lemmatization. We identified the polarity of the tweets using the Text Blob library. There were five groups created for them. On top of that, three machine learning (ML) techniques were used on the information. SVM classifier outperformed the Multinomial Logistic Regression classifier and Random Forest classifier in Twitter Sentiment Analysis with an accuracy of 86.60%. The accuracy of our Support Vector Classifier was enhanced by using GridSearchCV to choose the appropriate hyperparameters while applying SVM. We discovered that at "C":1000 and "gamma," 0.001 was the optimal value [17].

Utilising both user and textual features, the Twitter data was utilised to study abusive posting behaviour from a number of angles. The number of highly interrelated abusive behaviour standards was identified using a deep learning architecture; the suggested work makes use of easily available metadata to combine it with hidden patterns that are automatically extracted from tweeted text. With no need for model architecture modification for each activity, we can detect different types of abusive behaviour using this unified architecture in a seamless and transparent manner. We assess the suggested approach using various datasets that include a range of abusive Twitter behaviours. The findings indicate that it greatly outperforms the other techniques based on the dataset (an improvement in AUC of between 21 and 45%). Training a multi-input network is challenging. A technique that uses two input routes and alternates training between them was developed to further enhance performance across all evaluated datasets. We showed that the suggested training paradigm can perform noticeably better than a variety of other possible training techniques, such as ensemble, feature transfer, and concurrent training [18].

Using a congruent attention fusion capsule network, cyberbullying on social media was discovered. Due to the homogeneity of bullying types, dynamic routing between capsules can more adaptively identify the fine-grained categories in the cyberbullying text to further improve detection performance. An extensible congruent attention mechanism that strikes a balance between the fusions of detailed correlations between various feature subspaces was developed using a unified spatial representation of the composition. To automatically improve bullying characteristics and optimise the word embedding matrix, a similarity weighting approach based on word2vec that assesses the similarity between context words and external bullying vocabulary has also been investigated [19].

The goal of the study was to identify cyberbullying actors using text analysis and user credibility to educate users about the harm that cyberbullying causes. Information was taken from Twitter. Because the data were unlabeled, we created a web-based labelling tool to separate tweets that contained cyberbullying from tweets that did not. The technology provided us with 2053 derogatory words, 129 tweets containing swear words, 301 tweets about cyberbullying, 399 tweets with non-cyberbullying. Then, using SVM and KNN, we discovered and identified cyberbullying texts. The results show that SVM generates the greatest F1-score, which is 67% [20]. People occasionally use sarcasm to convey their feelings and thoughts, indicating the complete opposite of what they say. Sarcastic content can be shared by users in a number of different ways, including audio, photos, podcasts, text, and videos. The study uses text data taken from Twitter to examine COVID-19 ironic material with negative attitudes. We extracted the data using specific terms such hash tag-related, irony, sarcastic information, and sarcasm, and then we did an aggressive and offensive character study of people. Decision Tree, Naive Bayes, and the linear support vector classifier were used to conduct this research. The decision tree demonstrated the highest level of accuracy compared to libSVM and Naive Bayes with 90% accuracy (Table 1).

Cyberbullying Detection Algorithms

The two main parts of the cyberbullying detection techniques are discussed.

Table 1 Comparative summary of different approaches in cyberbullying detection

References	Year of Publication	Features	ML Techniques	Outcomes	Research gaps
[1]	2021	BoW and TF-IDF	SVM	Combining ML with natural language processing	The textual categorization method can be used to identify cyberbullying
[5]	2021	Multilevel automatic feature	Deep learning model and Multichannel deep learning framework	Blocks with bidirectional gated recurrent unit transformers and convolutional neural networks	The effectiveness of the results can be increased by using a larger dataset
[13]	2021	Feature vector, dependent feature	Ensemble model	Logistic regression, linear SVC for ensemble and multinomial NB	The tweet content will be translated into different languages and deep learning technology will be used for improved outcomes in the future
[14]	May-2022	Sentimental features	Naïve Bayes and SVM	BERT model, image-based cyberbullying can be recognised using optical character recognition (OCR)	With a huge data collection, detection using an image dataset and a distinct modality can produce more accurate findings
[6]	Dec-20	Latent semantic features, bag of words features	Logistic Regression classifier	Support vector classifier, logistic regression, naïve Bayes, random forest classifier	The textual categorization method can be used to identify cyberbullying
[15]	2020	Bag of n-grams model, Bag-of-words, embedding and RNN	Lexical syntactic analysis	Unsupervised methods	Unsupervised techniques can be the subject of further study
[16]	2020	TF-IDF and Word2Vec feature extraction have been used	LGBM, LR and SGD,	Evaluation of classifier performance metrics	Several features can be used to increase the SGD and LR classifiers' detection rates
[17]	2021	Text blob	MLR,RF,SVM,	Using ordinal regression, analyse Twitter sentiment	Using SVM, sentiment analysis was carried out
[18]	2018	Word vectors, metadata features	Multi-input classifier with integrated classification	TF-IDF weights for naive Bayes model	You can spot toxic behaviour
[19]	2020	Multiple subspace features, Word-2vec,	Naive Bayes	Network of fusion capsules paying consistent attention	In-depth interaction of complementary traits is necessary to reduce bullying's homogeneity
[20]	2018	Rule based feature extraction	SVM, KNN	Using a web-based labelling tool, user credibility analysis was carried out	Utilising the POS Tagger tool will improve the results of feature extraction
[21]	2021	Textual features	BERT	Decision tree classifier, machine (SVM) with a linear Kernel, support vector and naive Bayes classifier	BERT model had a 91.90% accuracy rate
[22]	2021	Term frequency-inverse document frequency (TF-IDF)	Naïve bayes(nb), linear support vector classifier (libsvm), and decision tree	With Naive Bayes, sarcasm may be recognised with great accuracy	With a huge dataset, sarcasm content on online social media can be predicted more accurately

Part 1: using natural language processing to detect cyberbullying.
Part 2: machine learning.

Cyberbullying Detection in Natural Language Processing

One area of research in this area is the detection of offensive information using natural language processing (NLP). The most effective method for describing how Natural Language Processing works is "LANGUAGE LEVELS" [23]. These levels are used by people to interpret spoken or written languages. Processing applications make use of linguistic abilities [23, 24].

Machine Learning in Cyberbullying Detection

The datasets are used with any of the following machine learning algorithms to find bullying-related messages on social media.

Decision Tree

Decision tree classifiers can be used for both classification and regression [1]. It can help in decision-making and representation. It is a framework that resembles a tree, with each leaf node standing in for a choice and each inside node for a condition. A classification tree returns the class of the target. A regression tree will show the anticipated value for an addressed input.

Naive Bayes

Naive Bayes is a powerful machine learning method that is based on the Bayes theorem [25]. The probability of an object is used by the algorithm to create predictions. Multiclass classification and binary problems can be swiftly fixed with this technique.

Support Vector Machine

Similar to a decision tree, the support vector machine (SVM) is a supervised machine learning technique that may be applied to both classification and regression. It can distinguish the classes in n-dimensional space in a special way. In practice, SVM creates a series of hyperplanes in an infinite-dimensional space using a kernel that turns an input data space into the desired form. For instance, Linear Kernel employs instances dot products in the following manner: $K(x, xi) = \text{sum}\ (x, xi)$.

Evaluation Metrics

Researchers assess how well a suggested model separates cyberbullying from non-cyberbullying using a range of evaluation metrics. Examining typical evaluation criteria employed by academics is crucial to understanding the efficacy of models. The most popular measures for evaluating cyberbullying classifiers on social media websites are as follows:

Accuracy

Enhanced accuracy in identifying cyberbullying content and determining the victim's emotional state following a cyberbullying incident via real-time streaming data

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

The prediction models for cyberbullying could be assessed using.

Precision: determines the percentage of successfully or accurately identified samples

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall: provides the percentage of real positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1 Measure: provides the choral group with memory and accuracy

$$F1 - \text{Measure} = \frac{-2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative, respectively.

Data for the performance measure evaluation metrics for social media websites of cyberbullying classifiers are shown in Table 2. The data demonstrates unequivocally that the stochastic gradient descent (SGD) classifier has the highest accuracy of 90.60% [16].

Conclusions and Future Directions

Social networks have mostly taken over our everyday routines because using them makes it so simple to engage with others. However, the development of antisocial behaviour like trolling, hate speech, and cyberbullying on social networks like Twitter and the negative effects that social media users experience make this a crucial topic to research

Table 2 Algorithms performance summary based on evaluation metrics

Sl.no.	Machine learning algorithm	Accuracy rate	Precision value	Recall value	F1 scores	Prediction time
i	Logistic regression (LR)	90.57%	0.9518	0.9053	0.928	0.0015
ii	Light gradient boosting machine (LGBM)	90.55%	0.9614	0.8951	0.9271	0.0515
iii	Stochastic gradient descent (SGD)	90.60%	0.9683	0.889	0.927	0.0016
iv	Random forest (RF)	89.84%	0.9338	0.9134	0.9235	2.5287
v	AdaBoost (ADB)	89.30%	0.9616	0.8756	0.9166	0.1497
vi	Naive Bayes (NB)	81.39%	0.7952	0.9736	0.8754	0.0034
vii	Support vector machine (SVM)	67.13%	0.6713	1	0.8033	39.9592

[5]. An effective strategy for addressing the cyberbullying issue is the use of deep learning techniques to identify the social media content that encourages it. The examination of research conducted revealed that there has been very little research on the identification of cyberbullying, and that detection using popular multimedia content, such as videos, music, etc., has been overdone using text-based research. The use of ML approaches is restricted by the availability of data because a testing and training dataset is necessary. In addition, the majority of the available datasets only contain text data, therefore the researchers had to create their own data. The concept of implementing a real-time cyberbullying detection system, which will be useful for identifying and stopping cyberbullying instantly, can be another area of research, and working with various different languages can open up research avenues [26]. Deep Learning techniques can be used to classify cyberbullying in a more precise manner, and they will perform better than the machine learning algorithms currently in use.

Funding No funding was received for this research work.

Declarations

Conflict of Interest The authors declare no conflict of Interest.

References

1. Islam MM, Uddin MA, Islam L. Cyberbullying Detection on social networks using machine learning approaches. In: 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)|978-1-6654-1974-1/20/$31.00 ©2020 IEEE. https://doi.org/10.1109/CSDE50874.2020.9411601.
2. Akram W, Kumar R. A study on positive and negative effects of social media on society. Int J Comput Sci Eng. 2017;5(10):351–4.
3. Al-Garadi MA, Hussain MR, Khan N, Murtaza G, Nweke HF, Ali I, Mujtaba G, Chiroma H, Khattak HA, Gani A. Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges. IEEE Access. 2019;7:70701–18.
4. Alsaed Z, Eleyan D. Approaches to cyberbullying detection on social networks: a survey. J Theor Appl Inform Technol 2021;99(13).
5. Alotaibi M, Alotaibi B, Razaque A. A multichannel deep learning framework for cyberbullying detection on social media. Electronics. 2021;10:2664. https://doi.org/10.3390/electronics10212664.
6. Shah R, Aparajit S, Chopdekar R, Patil R. Machine learning based approach for detection of cyberbullying tweets. Int J Comput Appl. 2020. https://doi.org/10.5120/ijca2020920946.
7. Salawu S, He Y, Lumsden J. Approaches to automated detection of cyberbullying: a survey. IEEE Trans Affect Comput. 2017. https://doi.org/10.1109/TAFFC.2017.2761757.
8. Mahlangu T, Tu C, Owolawi P. A review of automated detection methods for cyberbullying. 2018 IEEE.
9. Nadali S, Murad M, Sharef N, Mustapha A, Shojaee S. A review of cyberbullying detection. An overview. In: 13th International Conference on Intelligent Systems Design and Applications (ISDA), 2013.
10. Haidar B, Chamoun M, Yamout F. Cyberbullying detection a survey on multingual techniques. European Modelling Symposium, 2016.
11. Zainudin N, Zainal K, Hasbullah N, Wahab N, Ramli S. A review on cyberbullying in Malaysia from digital forensic perspective. In: 2016 International Conference on Information and Communication Technology (ICICTM), 16th–17th May 2016, Kuala Lumpur, Malaysia, 2016.
12. Romsaiyud W, Nakornphanom K, Prasertsilp P, Nurarak P, Konglerd P. Automated cyberbullying detection using clustering appearance patterns, 2017.
13. Azeez NA, Idiakose SO, Onyema CJ, Van Der Vyver C. Cyberbullying detection in social networks: artificial intelligence approach. J Cyber Secur Mob. 2021;10(4):745–74. https://doi.org/10.13052/jcsm2245-1439.1046.
14. Kargutkar SM, Chitre V. A study of cyberbullying detection using machine learning techniques. In: ICCMC, 2020; pp. 734–739, https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000137.
15. El-Seoud SA, Farag N, McKee G. A review on non-supervised approaches for cyberbullying detection. Int J Eng Pedagog. 2020. https://doi.org/10.3991/ijep.v10i4.14219.
16. Muneer A, Fati SM. A comparative analysis of machine learning techniques for cyberbullying detection on Twitter. Future Internet. 2020;12:187. https://doi.org/10.3390/fi12110187.
17. Ahmed M, Goel M, Kumar R, Bhat A. Sentiment analysis on Twitter using ordinal regression. In: 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) Pune, India, Oct 29–30, 2021.
18. Founta AM, Chatzakou D, Kourtellis N, Blackburn J, Vakali A, Leontiadis I. A unified deep learning architecture for abuse detection. In: 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
19. Wu F, Gao B, Pan X, Su Z, Ji Y, Liu S, Liu Z. FACapsnet: a fusion capsule network with congruent attention for cyberbullying detection. Neurocomputing. 2023;542:126253.

20. Nurrahmi H, Nurjanah D. Indonesian Twitter cyberbullying detection using text classification and user credibility. In: 2018 International Conference on Information and Communications Technology (ICOIACT).
21. Desai A, Kalaskar S, Kumbhar O, Dhumal R. Cyber bullying detection on social media using machine learning. In: ITM Web of Conferences 40, 03038 (2021) ICACC-2021 https://doi.org/10.1051/itmconf/20214003038.
22. Kumar R, Bhat A. An analysis on sarcasm detection over Twitter during COVID-19. In: 2021 2nd International Conference for Emerging Technology (INCET) Belgaum, India. May 21–23, 2021.
23. Louppe G. Understanding random forests: from theory to practice. arXiv 2014, https://arXiv.org/1407.7502.
24. Novalita N, Herdiani A, Lukmana I, Puspandari D. Cyberbullying identification on Twitter using random forest classifier. J Phys Conf Ser. 2019;1192:012029.
25. Al-Hassan A, Al-Dossari H. Detection of hate speech in social networks: a survey on multilingual corpus.
26. Eshan S, Hasan M. An application of machine learning to detect abusive Bengali text. In: 2017 20th International Conference of Computer and Information Technology (ICCIT), 22–24 December, 2017, 2017.