A Survey on Methods for Detecting Cyberbullying in Multilingual Documents

[1]Renetha J B, [2]Deepthi P S[1]

[1]Department of Computer Science and Engineering,
LBS Institute of Technology for Women, Thiruvananthapuram, Kerala, India
[1]APJ Abdul Kalam Technological University, Trivandrum, Kerala, India
Department of Computer Science and Engineering,
[1]Lourdes Matha College of Science and Technology, Trivandrum, Kerala, India
[2] Department of Computer Science and Engineering,
LBS Institute of Technology for Women, Trivandrum, Kerala, India
[2]renetha.jb@lmcst.ac.in, [2] deepthisath@gmail.com

ABSTRACT

Digital technologies are now swallowing the world. People irrespective of the age and gender are influenced by the colourful wings that they provide. Teenagers are the main victims of this digital era. They become addicted to games and the virtual world more quickly than any other age group. Their age is so critical that they are very much sensitive. There is a natural tendency among teenagers to do things so as to catch the attention of others. Sometimes this paves way to bully or harass or embarrass others on the internet or other digital spaces such as social media sites and that causes a negative impact on those who are being targeted then there arises the threat of cyberbullying. Social media users are not merely sticking on English language in their posts or comments but the usage of multilingual code mixing or even code switching is very much prevalent. Surveys have been done among people of different ages in many countries and have demonstrated various consequences of cyberbullying victimisation that lead to change in behaviour and increased anxiety. Researchers identified the necessity of computer-based solutions for determining, preventing, mitigating ow even stopping cyberbullying. This paper is a survey of various computer-based techniques specifically concentrating on machine learning, deep learning and natural language processing that targets to detect cyberbullying in online media.

Keywords: Bullying, Online media, Code-mixed data, Hate speech, Offensive speech, machine learning, deep learning, natural language processing

INTRODUCTION

The adolescent period is a critical state in the lifetime of an individual during which they are influenced more by the society and strangers than those nearby. According to the surveys conducted in various parts of the world, it has been observed that the growth of cyberbullying attacks is alarming. In a survey [1] conducted among tweens belonging to the age group 9 to 12 years in the US it has been observed that about one in five teens are becoming a target, witness, or an aggressor. Like any other crimes cyberbullying has also become a widespread social problem. Accuser, bully, reporter and victim are the four most frequent roles in social media [2]. Cyberbullying is hidden or vaguely visible as aggressive social media posts, usage of abusing words made while gaming, fake accounts for making fun of others, to embarrass known or strangers, threaten, or black-mail, or following some forms of intentional cruelty that can cause deep wounds in young minds.

Nowadays people follow code-mixed languages. This is clearly notable in most of the social media posts. Code switching [4] also can be seen in many social media posts where languages are being alternately used for expressing a situation or a matter. The reason for this is people usually feel that when they want to stress some points or to express their feeling in a well manner they can express well in native language, which in turn lead to the prominent multilingual language usage especially in online media. It is a challenge to detect bullying in such posts. Such cyber-attacks can lead to undesirable situations such as gaining access to people's confidential information, monitor user activities and make the system unavailable to users [30][29]. Bullying involves multiple roles like bully (or bullies), the victims, bystanders, defenders of the victim, assistants to the bully, reinforcers that makes evident the extensive impact of bullying [21]. Proper detection and prevention methodologies need to be formulated for solving this cyber issue.

CYBERBULLYING DETECTION METHODOLOGY AND DOMAIN STUDY

Cyberbullying is being listed as the critical cyber issue in the last decade [33]. In this paper, a narrative literature survey is done to identify the technologies for detecting and preventing cyberbullying. Studies related to cyberbullying and its impact on people of different age groups by analysing various surveys have been done in different countries [1][2][33]. It has been mentioned in different surveys about the vulnerabilities of cyberbullying and the proposal to extend programs dealing with traditional bullying to tackle cyberbullying. The relevance of cyberbullying area in recent research initiatives and also keywords that are linked to it are identified. For this, abstract and briefings about cyberbullying related papers from the year 2012 to 2022 were taken into consideration. Comparative study was done in the selected papers related to cyberbullying detection methodologies.

Cyberbullying works are mostly being done using Machine Learning, Deep Learning and Natural Language Processing. General architecture of cyberbullying detection model is given in Figure 1.

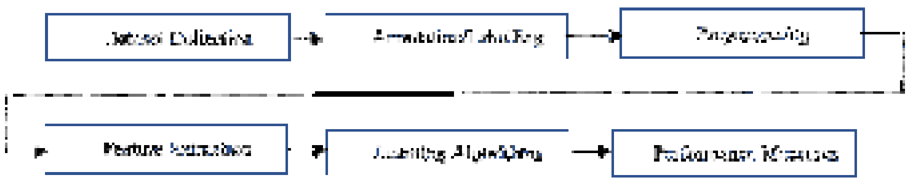

Figure 1. Cyberbullying detection model

A. Machine learning model for detecting cyberbullying

First step in detecting cyberbullying is to collect datasets and perform annotations based on the class of speech they belong to. This is followed by data cleaning and data preprocessing for removing irrelevant, unwanted or erroneous entry. This step helps to reduce the size of the vector dimension where emojis, user information such as email id, mobile number, date and time, URL are being normalized. Duplications are also removed in this step. Data partitioning follows where preprocessed data is partitioned into test and training data.

People often feel comfortable to express their ideas in mixed language, where they use their own native language along with other language, which is termed as code-mixing. Inter sentential switching, also called sentential switching exists when code mixing occurs at the boundaries of textual data. If language swing occurs within the same sentence, then there exists intra sentential switching. Numerous code-mixed data exists such as Hinglish, Singlish where in former Hindi language is written using English alphabetical characters and in latter Sinhala language is written using English alphabetical characters. Code-mixed data can be native language written in English or combination of native language and words written in English or English alphabetic combinations along with native language letters [18].

Pawar et al. [24] for detecting cyberbullying in Hindi and Marathi languages came up with a Multilingual Cyberbullying Detection System. 97% accuracy was obtained for both Hindi and Marathi on experimenting with different datasets. F1 score for both languages were about 96% on many datasets.

Oriola, O. and Kotze, E [5] collected tweets of South African discourses using Twitter API. Before tokenizing and preprocessing the tweets, syntactic and sentiment count indicators were extracted. They applied a multi-tier meta-learning model and performed analysis using Support Vector Machine (SVM), Random Forest, Logistic Regression and Gradient Boost meta-learners. Class imbalance was reduced by applying synthetic minority oversampling technique.

Akhter, M.P, Jiangbin et al. in their study [6] in Urdu and Roman Urdu languages used techniques for analyzing text comments for checking the presence of offensive contents. They conducted experiments using machine learning techniques namely Bayesian model, k nearest neighbours (KNN), random algorithms, regression, SVM, and rule based to identify the usage of abusive language in Urdu and Roman Urdu datasets. Regression models along with character tri-gram features was found to yield best result in their experiments. F-measure of 99.2% for LogitBoost on Roman Urdu dataset and 95.9% for SimpleLogistic on Urdu datasets were obtained.

Sreelakshmi K, Premjith B et al. [3] work based on a machine learning model focused to detect hate speech in Hindi-English code-mixed social media text. Using SVM and Random Forest, classification was done. Using SVM- RBF, they obtained a maximum accuracy of 85.81%.

Oshadhi Liyanage et al. [18] worked in Singlish for detecting abusive or offensive speech. They proposed to select best performing combinations from Logistic Regression, SVM, MNB, KNN, Ada Boost, Decision Tree, Bernoulli Naïve Bayes and Perceptron. Then selected best Ensemble approach using hard voting, soft voting and stacking.

Haidar et al. [27]in their work for detecting cyberbullying used machine learning techniques SVM and Naive Bayes ML learning on an Arabic dataset with 32K tweets in which 1800 tweets were bullying ones. They obtained F1 scores of 92% using SVM and 90% using NB.

Table 1**Error! Reference source not found.** gives a summary of machine learning research for detecting cyberbullying in multilingual documents

Ref	**Methods used**	**Dataset**	**Performance and Remarks**
[18] O. Liyanage et al.	SVM, LR, MNB, AdaBoost	Sinhala-English code-mixed datasets from Youtube and Facebook comments.	Hard voting approach outperformed other baseline algorithms and ensemble approaches with 84% accuracy and f1-score.
[5] O Oriola et al.	LR, Gradient Boosting (GB), SVM, Random Forest (RF)	Twitter-21,350 South African tweets of	Maximum performance was exhibited by optimized GB models of word n-gram, SVM, SVM model of character n-gram, RF and GB multi-tier meta-learning model
[6] Akhter, M.P, Jiangbin et al	NB, Bayes Network (BayesNet), Random Tree and Random Forest, Instance-Based Learning (IBk), J48, Hoeffding Tree, REPTree(Reduced Error Pruning Tree), linear multinomial logistic regression	Roman Urdu and Urdu YouTube videos comments.	LogitBoost and SimpleLogistic achieved F-measure of 99.2% on Roman Urdu and 95.9% on Urdu datasets. Among n-gram feature, character tri-gram was found to be more effective.
[3] Sreelakshmi K, Premjith B et al.	Used SVM, SVM-RBF and Random Forest for classification. Used doc2vec, word2vec and fastText for learning	TwitterAPI shared task HASOC	Accuracy of fastText using SVM- RBF- 85.81% word2vec using SVM-RBF.gave 75.11% accuracy doc2vec gave 64.15% accuracy using Random Forest
[19] Chen et al.	SVM, NB, LSF based framework	Youtube comment boards, dataset includes comments from 2,175,474 distinct users.	LS gave precision of 77.9% and recall of 77.8% using SVM.
[27] Haidar et al.	SVM,NB	32K tweet	F1 score SVM -92% NB – 90%

Table 1. Cyberbullying detection using machine learning

B. Deep learning models for detecting cyberbullying

Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations (BERT) are some of the deep learning methods for cyberbullying detection [15].

P. Deepasree Varma, P. Vinod, et al. [8] performed experiments on tweets and classified tweets as hate or non-hate content using transformer-based models. They worked on languages Malayalam, Code-mixed Malayalam and English. BERT algorithm was used for languages other than English. They used a variety of BERT algorithm namely DistilBERT, a pretrained model. The BERT embeddings are used as training set for different machine learning methods such as LR, SVM and Multinomial Naive bayes. The sentence was analysed using DistilBERT which transfers some of the information it gains to the next model. The logistic regression model along with BERT model produced relatively acceptable results in their study. Compared to other baseline monolingual models, their models obtained equivalent or greater on an aggregated dataset in English, Malayalam, and Code-mixed Malayalam. With DistilBert along with LR maximum accuracy score of 90% was obtained for the Code-mix dataset.

A. Kalaivani and D. Thenmozhi [9] used the pre-trained MBERT (Multi-lingual BERT) models for Tamil, Malayalam, and Kannada languages. Using pre-trained models MBERT with the ktrain library for those three languages, the system was built and predicted the comments for all these languages. Accuracy of MBERT model achieved for Tamil was 0.61,0.71 for Malayalam and 0.61 for Kannada.

Puneet et al. performed classification of abusive Hinglish tweet. A set of 3000 data, labelled as non-offensive, hate-inducing and abusive were taken [10]. They created the dataset HOT (Hinglish Offensive Tweet), using the Twitter Streaming API. Three annotators were used for labelling. For detecting abusive content in Hinglish tweets, they came up with a model called Multi-Input Multi-Channel Transfer Learning based model abbreviated as MIMCT. LSTM and CNN models were trained on English tweets with offensive content. After that models were retrained using the Hindi-English tweets. MIMCT model was found to perform better than SVM with TF-IDF features by 0.166 F1 points and the LSTM transfer learning model by 0.165 F1 points.

Y. Prakash Babu, R. Eswari [13] used the Paraphrase XLM-RoBERTa model and trained the model on code-Mixed language datasets of Malayalam, Tamil and Kannada, and obtained 71.1, 75.3, and 62.5 F1 scores respectively. Alhawarat et al. [7] classified Arabic news documents using Multi-Kernel CNN model and attained an accuracy ranging from 97.58% to 99.90%.

Table 2 gives a summary of few deep learning methods for detecting cyberbullying in textual data.

Ref	Methods used	Dataset	Performance and Remarks
[8] P. Deepasree Varma et al.	BERT, DistilBERT. LR, SVM, Multinomial Naive bayes MNB	Data from Google searches. Data from Github repositories, Dynamically Generated Hate Speech Dataset	Logistic Regression with DistilBert-90% (highest accuracy)
[9] Kalaivani et al.	pre-trained Multi-lingual BERT, MBERT	Dravidian-CodeMix-FIRE2021 dataset, ktrain library	Accuracy of MBERT model Tamil -0.61, Malayalam 0.71 and Kannada 0.61
[10] Puneet et al.	MIMCT, LSTM and CNN	HOT dataset	maximum performance-MIMCT model
[13] Prakash Babu et al.	Paraphrase XLM-RoBERTa model	YouTube comments	Fi scores Tamil 71.1, Malayalam 75.3, and Kannada 62.5 for code-Mixed language datasets
[7] Alhawarat et al.	Multi-Kernel CNN, SATCDM	NADA, SANAD	accuracy range from 97.58% to 99.90%,

Table 2. Deep learning methods for cyberbullying detection

C. Natural Language Processing model for detecting cyber bullying

NLP is a technique for performing text mining that does linguistic analysis which helps a machine to understand textual data or speech. Syntax and semantic analysis are the core jobs of NLP [20].

Feature extraction which is a crucial step for classifying the text data for determining whether bullying exists need to extract simple surface features, word generalization, sentimental analysis, lexical resources, linguistic features, knowledge-based features, meta information and multimodal information.

Term Frequency-Inverse Document Frequency (TF-IDF) uses a combination of TF and IDF, and Word2Vec is a two-layer neural net that “vectorizes” words to process text. These two are feature extraction techniques.

Work by Puneet et al. [10] used Twitter word2vec (Tw) and FastText (Ft). Tw and Ft exhibited superior performance. They proposed Multi-Input Multi-Channel Transfer Learning based model (MIMCT) for detecting abusive speech and retrained the model on the Hinglish Offensive Tweet (HOT) dataset using transfer learning coupled with multiple feature inputs. And used Glove, fastText and Twitter word2vec for the word embedding. They concluded that in terms of F1-score and precision score, SVM model supplemented with TF-IDF features achieved highest performance.
In work by Neeraj Vashistha et al. [17], logical regression model along with TFIDF and POS features improves the performance. Alhawarat et al. [7] used n-gram word

embedding called Superior Arabic Text Categorization Deep Model (SATCDM). They achieved an accuracy ranging from 97.58% to 99.90%, for Arabic text document.

CHALLENGES AND LIMITATIONS

Chris Emmery, Ben Verhoeven et al. [14] classified cyberbullying detection techniques into three categories: binary, fine-grained, metadata approaches. Binary approach focussed on binary classification where it checks if a message contains bullying or not. Fine-grained approaches aim to check the role of actors in a bullying scenario or the content type. Text-based features are the focus in these two approaches. Metadata approaches take other details such as user profile, network details in addition to message. Usually cyberbullying occurs behind the curtains in social media where tracing such contents will be a great challenge. Data scarcity is also a hinderance for accurately determining the cyberbullying occurrences.
Ambiguity sometimes stands as a hindrance in the existing cyberbullying detection systems for distinguishing hate from non-hate [16]. One such situation is when some words may be treated offensive in some country but may be treated a normal in another country. Change in gesture while using some normal words may make them offensive. This is also not possible to detect through analysing the textual data. Another more general issue in cyberbullying prevention research is the scale and approach of program evaluation [24].

DATASETS

Modha et al. [23] focused coarse-grained binary classification that classified tweets into Hate or Offensive (HOF) vs Non-Hate and Non-Offensive (NOT) classes. They also focused on fine-grained classification task to classify tweets into HATE (Hate speech), Offensive(OFN), Profane or unacceptable words(PRFN).
Bharathi Raja Chakravarthi1, Navya Jose et al. [11] presented a corpus for option mining of annotated code-mixed Malayalam-English text. Their experiments concluded that transformer-based models performed better than other traditional machine learning models. Better performance was obtained from XLM-RoBERTa[12] and transformer-based models. The different multilingual datasets available in literature are listed in Table 3.

Reference	**Language**	**Classes of speech**	**Source**
[22]Bohra et al.	Hinglish	tags namely, ‘eng’, ‘hin’ and ‘other’, Hate Speech , Normal Speech	Twitter
[23]Modha et al.	Hinglish	HOF, NOT ,HATE, OFFN, PRFN	Twitter, HASOC Hindi data set
[25] Chakravarthi et al.	Code-mixed Tamil-English (Tanglish)	sentiment-annotated	YouTube comments. Used YouTube Comment Scraper tool

[26] Remmiya Devi et al.	Tanglish and Hinglish	Entities	Twitter
[11] Chakravarthi et al.	Code-Mixed Malayalam-English	sentiment-annotated- Positive state, negative state, mixed feelings, neutral, not in intended language	YouTube comments. Used YouTube Comment Scraper tool
[28] Adeep Hande et al.	Kannada Code Mixed Dataset	Positive state, negative state, mixed feelings, neutral, not in intended language	YouTube comments.

Table 3. Multilingual datasets

FUTURE SCOPE

The main challenge in cyberbullying detection, especially the usage of emojis, local language words, shortened words in social media communication demands the need of dynamic datasets [31]. Usage of small, heterogeneous datasets, without a thorough evaluation of applicability is a limitation to effectively detect cyberbullying. Most of the cyberattacks take place in private conversations, which also create a hinderance. Handling obfuscated words that can by-pass the automated screening software for toxic language is another challenge in cyberbullying detection. Collaborative bullying, where more than one user targets a victim, also creates hinderance. So more sophisticated social media analysis is highly desirable.

There are various future directions in the cyberbullying research area. Real time chats need to be analysed and if found bullying, action should be taken automatically. Fine-grained cyberbullying categories such as threats, curses and expressions of racism and hate also paves way to further research. Another research scope is to develop algorithms capable of detecting cyberbullying from several multimedia sources [32].

CONCLUSIONS

Research on automatic detection of cyberbullying is going on in different parts of world because of its social implication. Nowadays it is a trend among people to share videos, photos and animations of their precious moments and they are made publicly visible which make them spread across the network in a minute. Accusers' private data may get manipulated and can lead to personal harassment. Importance of building a real-time cyberbully detection platform that can instantly detect and prevent the cyberbullying is highly desirable for protecting especially teens against the bullying attack. There are many challenges in cyberbullying research area, one among them is the increased usage of emoji along with textual messages. Research is now carried out in different areas like machine leaning, deep leaning and natural language processing for detecting, preventing and mitigating cyberbullying. Recall, precision, area under the curve score (AUC), accuracy (ACC) score are the commonly used evaluation parameters. Generic features, such as BOW or embeddings gives reasonable classification performance. Among deep learning techniques, pre-trained BERT model exhibiter better performance than other models.

REFERENCES

[1] Patchin, J.W. and Hinduja, S., 2022. Cyberbullying among tweens in the United States: prevalence, impact, and helping behaviors. *The Journal of Early Adolescence*, 42(3), pp.414-430.

[2] Hinduja, S. and Patchin, J.W., 2015. Bullying beyond the schoolyard: preventing and responding to cyberbullying. Thousand Oak.

[3] Sreelakshmi, K., Premjith, B. et al., 2020. Detection of hate speech text in Hindi-English code-mixed data. Procedia Computer Science, 171, pp.737-744.

[4] Beatty-Martínez, A.L. et al., 2020. Codeswitching: A bilingual toolkit for opportunistic speech planning. Frontiers in Psychology, p.1699.Oriola, O. and Kotze, E (2020) Evaluating machine learning techniques for detecting offensive and hate speech in South African tweets *IEEE Access 8*, 21496-21509

[5] Oriola, O. and Kotzé, E., 2020. Evaluating machine learning techniques for detecting offensive and hate speech in South African tweets. *IEEE Access*, 8, pp.21496-21509.

[6] Akhter, M.P., Jiangbin, Z., et al., 2020. Automatic detection of offensive language for urdu and roman urdu. IEEE Access, 8, pp.91213-91226.

[7] Alhawarat, M. and Aseeri, A.O., 2020. A superior Arabic text categorization deep model (SATCDM). *IEEE Access*, 8, pp.24653-24661.

[8] Varma, P.D., Vinod, P et al., 2022, June. Hate Speech detection in English and Malayalam Code-Mixed Text using BERT embedding. *In 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)* (pp. 1-6). IEEE.

[9] Kalaivani, A. and Thenmozhi, D., 2020. Multilingual Sentiment Analysis in Tamil, Malayalam, and Kannada code-mixed social media posts using MBERT. FIRE (Working Notes).

[10] Mathur, P., Sawhney, R. et al. 2018, October. Did you offend me? classification of offensive tweets in hinglish language. *In Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)* (pp. 138-148).

[11] Chakravarthi, B.R., Jose, N et al., 2020. A sentiment analysis dataset for code-mixed Malayalam-English. arXiv preprint arXiv:2006.00210.

[12] Chakravarthi, B.R et al., 2021. Findings of the sentiment analysis of dravidian languages in code-mixed text. arXiv preprint arXiv:2111.09811.

[13] Babu, Y.P. and Eswari, R., 2021. Sentiment Analysis on Dravidian Code-Mixed YouTube Comments using Paraphrase XLM-RoBERTa Model. *Working Notes of FIRE.*

[14] Emmery, C et al., 2021. Current limitations in cyberbullying detection: On evaluation criteria, reproducibility, and data scarcity. *Language Resources and Evaluation*, 55(3), pp.597-633.

[15] Zhou, Y., Yang, Y. et al., 2020. Deep learning-based fusion approach for hate speech detection. *IEEE Access*, 8, pp.128923-128929.

[16] Kumaresan, K. and Vidanage, K., 2019, October. Hatesense: Tackling ambiguity in hate speech detection. *In 2019 National Information Technology Conference (NITC)* (pp. 20-26). IEEE.

[17] Vashistha, N. and Zubiaga, A., 2020. Online multilingual hate speech detection: experimenting with Hindi and English social media. *Information*, 12(1), p.5.

[18] Liyanage, O. and Jayakumar, K., 2021, December. Hate Speech Detection in Sinhala-English Code-Mixed Language. *In 2021 21st International Conference on Advances in ICT for Emerging Regions (ICter)* (pp. 225-230). IEEE.

[19] Chen, Y., Zhou et al., 2012, September. Detecting offensive language in social media to protect adolescent online safety. *In 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing* (pp. 71-80). IEEE.

[20] Al-Makhadmeh, Z. and Tolba, A., 2020. Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach. *Computing*, 102(2), pp.501-522.

[21] Xu, J.M., Jun, K.S. et al., 2012, June. Learning from bullying traces in social media. *In Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 656-666).

[22] Bohra, A., et al., 2018, June. A dataset of Hindi-English code-mixed social media text for hate speech detection. *In Proceedings of the second workshop on computational modeling of people's opinions, personality, and emotions in social media* (pp. 36-41).

[23] Mandl, T., Modha, S. et al., 2021. Overview of the hasoc subtrack at fire 2021: Hate speech and offensive content identification in english and indo-aryan languages. arXiv preprint arXiv:2112.09301.Pieschl, Stephanie, Kourteva, Penka, and Stauf, Leonie. *'Challenges in the Evaluation of Cyberbullying Prevention – Insights from Two Case Studies'.* 1 Jan. 2017 : 45 – 54.

[24] Pawar, R. and Raje, R.R., 2019, May. Multilingual cyberbullying detection system. *In 2019 IEEE international conference on electro information technology (EIT)* (pp. 040-044). IEEE.

[25] Bharathi Raja Chakravarthi, et al. 2020. Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text. *In Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pp. 202–210, Marseille, France. European Language Resources association.

[26] Remmiya Devi, G., Veena, P.V et al., 2016. AMRITA-CEN@ FIRE 2016: Code-mix entity extraction for Hindi-English and Tamil-English tweets. *In CEUR workshop proceedings* (Vol. 1737, pp. 304-308).

[27] Haidar, B., Chamoun, M. et al., 2017, October. Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content. *In 2017 1st cyber security in networking conference (CSNet)* (pp. 1-8). IEEE.

[28] Hande, A., Priyadharshini, R. et al. 2020, December. KanCMD: Kannada CodeMixed dataset for sentiment analysis and offensive language detection. *In Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media* (pp. 54-63).

[29] Kumar, A. and Sachdeva, N., 2019. Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis. *Multimedia Tools and Applications*, 78, pp.23973-24010.

[30] Haque, A.B., Bhushan, B. and Dhiman, G., 2022. Conceptualizing smart city applications: Requirements, architecture, security issues, and emerging trends. *Expert Systems*, 39(5), p.e12753.

[31] Arif, Muhammad. (2021). A Systematic Review of Machine Learning Algorithms in Cyberbullying Detection: Future Directions and Challenges. *Journal of Information Security and Cybercrimes Research.* 4. 01-26. 10.26735/GBTV9013.

[32] Nirmal, N., Sable, P et al., 2021. Automated detection of cyberbullying using machine learning. *Int. Res. J. Eng. Technol.(IRJET)*, pp.2054-2061.

[33] Slonje, R., Smith, P. K., & Frisén, A. (2013). The nature of cyberbullying, and strategies for prevention. *Computers in human behavior*, 29(1), 26-32.