Al-Nahrain Journal of Science
ANJS, Vol.26 (2), June, 2023, pp. 47-55

A Comparative Study of Cyberbullying Detection in Social Media for the Last Five Years

Noor Haydar* and Ban N. Dhannoon

Department of Computer Science, College of Science, Al-Nahrain University, Jadriya, Baghdad, Iraq

Article's Information

Received: 09.03.2023
Accepted: 14.05.2023
Published: 30.06.2023

Keywords:
Cyberbullying detection
Machine learning
Deep learning
Feature extraction
Word embedding

Abstract

The number of users of social media sites has increased nowadays, and while these sites have many benefits, they also have many damages that have grown with the increasing number of users. Among these damages that have spread in social media sites in our time is the phenomenon of cyberbullying. It has become necessary to find solutions to detect it to prevent and hold bullies accountable to reduce the phenomenon of cyberbullying, which has great health and mental effects on the victim in society. There have been many attempts to build models to detect and classify cyberbullying by using machine learning and deep learning algorithms with different sets of data that were collected from social media sites such as Twitter, YouTube, Facebook, Instagram, and others. In this work, we show a group of previous studies that used machine learning and deep learning algorithms in good attempts to detect and classify the phenomenon of cyberbullying.

DOI: 10.22401/ANJS.26.2.08

*Corresponding author: noor.haidar21@ced.nahrainuniv.edu.iq

1. Introduction

In recent years, the number of users of social media sites such as Facebook, Twitter, Instagram, and others has increased, where any user can share their opinions on any topic on social media sites [1,2]. This frequent use of social media sites can cause many problems, especially for teenagers and children aged 13 to 22 years [3]. Users' most significant problems with increased social media usage are personal attacks, abusive language, cyberbullying, and hateful behaviour [4]. Cyberbullying is one of the most important problems facing social media site users. It is defined as a deliberate act of aggression committed by an individual or group using an electronic device and internet technology with social media sites [5-7]. The phenomenon of cyberbullying has begun to increase in range, from 10% of internet users up to 40% of internet users being victims of cyberbullying [3]. The main reasons for the increase in the phenomenon of cyberbullying are:

1. An Increased number of users of social media sites [8].
2. Social media sites are mostly available to all people without restriction.
3. Because cyberbullying is via electronic devices with the victim and not face-to-face bullying, it is challenging to reach bullies (users who cause cyberbullying), and it isn't easy to hold them accountable. This will lead to an increase in the number of bullies and an increase in the phenomenon of cyberbullying [9,10]
4. Freedom of expression about opinions through social media sites is available to all users [5,11].

There are different forms of cyberbullying appeared, including:

1. Sending a message to the victim. The message may include offensive, harmful, aggressive, and others [4].
2. Sharing photos or videos with the victim. Photos or videos may include content that is embarrassing or personal or others to the victim, which makes other users, when seeing these photos or videos, bully the victim [6,12].
3. Creating a fake profile for the victim. The fake profile may contain personal information or offensive information about the victim, which makes other users think that it is the victim's profile and that the victim is the one who published this information or offensive information. They start bullying the victim [6].

The increase in cyberbullying on social networking sites and the diversity of its forms has resulted in negative effects on the victim. The negative effects that appeared on the victims after being exposed to cyberbullying are many, such as negative effects on physical health and mental health like anxiety, depression, thinking, and low self-esteem, and sometimes it led to suicide [1,3].

With the emergence of negative effects and the increase of bullying on social media sites, it has become necessary to find a solution to reduce and prevent the phenomenon of cyberbullying. Many difficulties have appeared in the stage of detecting and preventing cyberbullying, such as users on social media sites do not use the ideal Arabic language, frequent use of abbreviations for actual words by users, and use of sarcasm words among users makes detecting cyberbullying a difficult task [3,10].

Machine learning and deep learning algorithms are widely utilized in a variety of areas to handle complex issues that are difficult to solve with traditional computer methods [13]. Detecting and preventing sentence bullying from social media were utilized using algorithms like Naive Bayes, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gate Recurrent Unit (GRU).

Different datasets like Twitter, YouTube, Instagram, and others were used. Each dataset is preprocessed in several processes before being trained and tested to detect cyberbullying, like removing punctuation, removing numbers, removing several spaces, tokenization, and other processes that preprocess the dataset.

2. Methodology

The methodology to detect cyberbullying using deep learning and machine learning is shown in Figure 1.

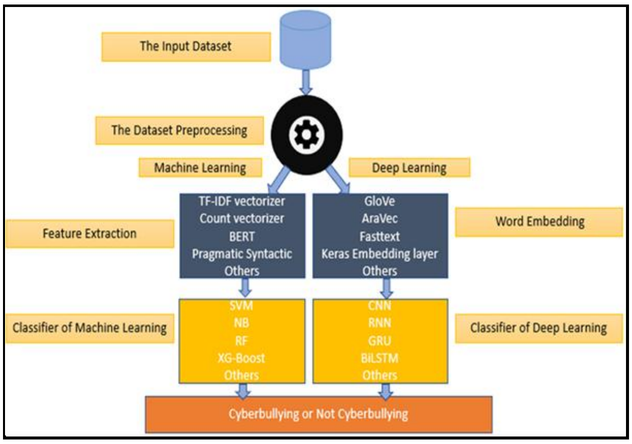

Figure 1. The methodology of deep learning and machine learning to detect the cyberbullying

2.1 The input datasets:

Authors collect their datasets from social media sites such as Twitter, Facebook, YouTube, and Instagram and try to detect and classify the phenomenon of cyberbullying to reduce and prevent it.

2.2 Datasets preprocessing:

The data may contain noise; therefore, there is a need to preprocess to reduce the number of words and sentences by removing unimportant words from the sentences and trying to link or approximate words that have the same meaning or are close to each other and other processes. The data preprocessing stage is carried out by several processes, including tokenization, stemming, stopwords and punctuation removal, and other processes. Tokenization is the process of splitting or fragmenting the text into smaller units, such as dividing the sentence into words, each of which is called a token. Stemming is the process of getting the original word or the word stem in Arabic by using specific algorithms instead of a specific dictionary. Several algorithms follow to apply the stemming process, like light stemmer and root stemmer. These algorithms do different mechanisms to apply the steaming, such as removing any affixes or suffixes attached to the Arabic words. Stemming is a simpler version of lemmatization with faster performance [14]. Stopword removal is the process of removing certain words that are repeated in the texts, called stopwords [15].

2.3 Feature extraction and word embedding:

Computers can only work with digital data. As a result, it is important to understand information on the computer using text representation per the language. The text representation process is significant in natural language processing since it employs methods such as TF-IDF vectorizer, Count vectorizer, Fasttext, Global Vectors (GloVe), and others for word representation with machine learning and deep learning classifiers. Word embeddings are very important in the world of Natural Language Processing. They allow us to capture relationships in language that are difficult to capture otherwise.

Count vectorizer as TF-IDF vectorizer both converts text data into machine-readable formats. The count vectorizer produces an encoded vector with the same length as the entire comment and an integer count of how many times each word appears in the comment.

The internal implementation of Fasttext discards the word order information, depending on the score matrix. It sums up a score when one or more words with a high absolute value significantly influence the final decision [2,16].

The Global vectors (GloVe) word embeddings are based on co-occurrence data and can be used to discover relationships between words by studying these probability ratios. The GloVe is based on matrix factorization techniques applied to the text context matrix. A massive data set was constructed to calculate which "word" and how many times this word appears in a document's "meaning" (columns), and the GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space [8].

2.4 The classifiers:

The researchers use machine learning and deep learning algorithms as a classifier.

2.4.1 Machine learning classifiers:

Naïve Bayes (NB) classifier is a classification technique that is based on Bayes' theorem. Naive Bayes is a family of algorithms that share a common concept: the features used to classify are considered independent of one another. It forecasts the membership probabilities for each class, with the class with the highest likelihood chosen as the most likely [2,17].

Support Vector Machine (SVM) is a supervised machine learning technique that is commonly used to handle classification or regression issues. The SVM is a binary classification technique that classifies data and separates it into two groups by generating an operational separating hyperplane. The support vectors are the data points nearest to the hyperplane, and the hyperplane is a decision space separated into a set of objects of different classes [17,18].

XG-Boost is a supervised machine learning algorithm that is commonly used in classification and regression applications. Its structure is comparable to that of the high-performance gradient decision tree technique that use gradient boosting framework [19]. Its design seeks to make effective use of memory and computational resources. Several qualities are involved in its implementation. It employs sparse awareness, handling missing data values automatically. It employs a block design to enable the building of parallel trees. Training using previously fitted data can be repeated to increase algorithm performance [20]. Random Forest RF is a decision tree ensemble approach in which the features in the original dataset are randomly picked in each sample to create a decision tree model [21]. In machine learning, an ensemble model combines two or more models to improve prediction, accuracy, and durability over individual models [2,22].

2.4.2 Deep learning classifiers:

The convolutional neural network (CNN) model consists of three layers the input layer, the hidden layer, and the output layer. CNN layers, which are used to learn the sequential features of data inputs but have some limitations, have led researchers to use Recurrent Neural Networks (RNN) to achieve better performance [23,24]. RNN stands for a recurrent neural network. RNN is a strong artificial neural network design that can process input sequences of any length. RNN is a classic feedforward neural network that takes only information from the current time and ignores useful information included in the data's temporal sequence and spatial arrangement. As a result, deep learning researchers are beginning to use LSTM, a type of neural network with memory that can save information for an extended period. The LSTM can automatically save information for a long time because it contains a storage memory, and the structure of the LSTM is more complex and intelligent than the RNN. The architecture of LSTM contains an input gate, a forget gate, and an output gate, which solves saving problems. LSTM is appropriate for dealing with and predicting events with significant intervals and delays in time series, which solves the limitations of basic RNN [25]. Bidirectional Long Short-Term Memory (BiLSTM) is an enhanced development of the LSTM architecture that simultaneously processes data in both forward and backward directions. BiLSTM is supposed to collect more information about the input sequence's future and past than standard LSTM. FFNN, also known as multilayer perceptrons and is a type of ANN that processes input in a single direction (forward direction) using many layers of computing units [26]. Most of the time, the GRU network model is used in research publications to deal with the vanishing gradient problem. Because it has three main gates, the GRU is more effective than the LSTM. The material is kept in a hidden format within the GRU for security reasons. The update gate receives both forward and backward information [27].

3. Related Works

This section presents literature studies that relate to detecting and classifying cyberbullying. Studies have begun to use machine learning and deep learning approaches to decrease or detect instances of cyberbullying, and from these studies come the following:

3.1 Machine learning approaches:

The authors in [5] use the SVM classifier to detect cyberbullying words in a Twitter dataset. The dataset size is 17748 tweets, including 14178 cyberbullying tweets and 3570 non-cyberbullying tweets. The Twitter dataset is preprocessed in several processes, including normalization, tokenization, light stemming, and others. TF-IDF word embedding was used to extract the features. The proposed module has done three experiments. The first experiment with the WEKA tool using a light stemmer achieved an accuracy of 85.49% in 352.51 seconds. The second experiment with WEKA using Arbic Stemmer Khoja achieved 85.38% accuracy in 212.12 seconds. The third experiment with Python achieved an accuracy of 84.03 % in 142.68 seconds.

In [6], This study used Arabic YouTube comments as a dataset containing over 15,000 YouTube comments, including 5,817 positive comments. The dataset is preprocessed using normalization, cleaning the text, and stemming. After that, the TF-IDF vectorizer and the count vectorizer were used for feature extraction. The machine learning modules used Multinomial Naïve Bayes (MNB), Complement Naïve Bayes (CNB), and Linear Regression (LR). The best accuracy was 78.6 % with Linear Regression (LR) module and count vectorizer feature extraction.

The authors in [7] use SVM and neural networks (NN) as machine learning modules with TF-IDF feature extraction on a cyberbullying dataset from Kaggle. The dataset includes 12773 belonging to the class cyberbullying, while 11735 belongs to the other class. The dataset was preprocessed in several processes, such as tokenization, text lowering, stopwords cleaning, etc. The higher accuracy was achieved at 92.8 % with a NN, and the higher F1 score was 91.9 % with a NN.

The authors in [28] use different machine learning modules like NB, SVM, Random Forest (RF), Artificial Neural Network (ANN), XGB, SDL, and Consensus-Based Ensemble Model to achieve better performance. They use TF-IDF vectorizer as feature extraction on a dataset of size 23462 samples. The dataset was collected from different resources: Twitter (13471 samples), WhatsApp (1281 samples), Vine (1332 samples), Instagram (6097 samples), and Packet (1281 samples). The best accuracy is 88.54 % with the Consensus-Based Ensemble Model and 88.82 % F1 score with the Consensus-Based Ensemble Model.

The Naïve Bayes (NB) module has been used in [29] on the Arabic dataset collected from YouTube and Twitter data. The accuracy was 95.9 %, and the F1 score was 92.78% with the Naïve Bayes (NB) module.

In [12], the researchers used four datasets. The first dataset from Formspring contains 13,110 posts labelled as bullying and non-bullying. The second dataset from Twitter consists of 13,420 tweets labelled as "offensive" and "not offensive". The third dataset from Twitter include 8817 tweets labelled as either positive (bullying) or negative (non-bullying). The fourth dataset from Twitter is hate-offensive and contains 24,784 tweets labelled as offensive, hate, or none. In this paper, the researchers use five machine learning classifiers like RF, NB, SVM, Logistic Regression (LR), and Ensemble. The better accuracy was achieved at 79.3 % with SVM.

The researchers in [30] use machine learning models with five feature extractions like TFIDF, Sentiment, Semantic, Pragmatic-Syntactic, and Count Vectorizer with Twitter datasets, which are divided into two groups. The first group is the dataset with 9057 tweets, which includes non-bullying 4,852 and 4,204 bullyings, and the second group is the dataset with 22,890 tweets, which includes non-bullying 12,179 and 10,711 bullying. The limitation of this research is that the information included in the dataset was insufficient to achieve better results, so there is a need for a dataset in another social network platforms such as Facebook, YouTube. The machine learning modules used support vector machines (SVM), random forests (RF), LG, K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Multilayer Perceptron (MLP). The better accuracy was achieved at 98 % with RF.

The authors in [31] used two publicly available datasets. Both datasets are manually labelled and free to use with the Bidirectional Encoder Representations from Transformers (BERT)-base model to classify and detect cyberbullying. The first Formspring dataset question and answer (a Q&A forum) contains 12,773 question-answer pair comments manually annotated by three workers and contains 776 posts that are marked bullied by at least two workers. The second Wikipedia talk page (collaborative knowledge repository) contains 11,5864 discussion comments that are manually annotated by ten persons and contain 13,590 comments that are labelled as a personal attack (bully). The authors achieve good accuracy of 98 % with the first dataset and 96 % with the second dataset.

3.2 Deep learning approaches:

The authors in [9] use the comments on the Arabic news channel Aljazeera as a dataset. The dataset contained 32K comments, which included obscene, offensive, and clean comments. The dataset preprocesses several processes, like removing diacritics, numbers, HTML codes, punctuations, and other processes. The dataset was split into three versions: AJComments Original version (Aljazeera comments original version) consists of two classes, first-class merging obscene. Offensive comments in one class, namely cyberbullying class, and the second class include clean comment which, namely non-cyberbullying, AJComments-Balanced version, which includes two classes (cyberbullying and non-cyberbullying class), have the same number of samples, AJComments-Unbalanced version which include two classes first cyberbullying class content obscene comments and the second non-cyberbullying which clean content comments. This paper classifies and detects cyberbullying original, balanced and unbalanced versions of datasets. Each dataset version is done in a different experiment and gets a different result. Best models achieved in balanced version 84% F1-score. The deep learning models used were Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) with AraVec and Fasttext word embeddings, and machine learning modules with TF-IDF vectorizer.

The LSTM as a deep neural network has been used in [3] on three datasets with Sentiment specific word embedding (SSWE) layer. The datasets from Wikipedia, Twitter, and Formspring each have 3000 examples, bringing the total number of examples to 9000. Datasets are preprocessed in several processes, such as by removing punctuation marks, blank spaces, symbols, numbers, and others. The accuracy was 75.5 % with the Wikipedia dataset, 72 % with the Formspring dataset, and 79.1 % with the Twitter dataset. This paper dosen’t performing well due to insufficient datasize.

The authors in [1] used two datasets to detect cyberbullying. The first dataset (Binary aggressive cyberbullying dataset) includes 115,661 post samples, distributed as 101,082 aggressive posts and 14,782 non-aggressive posts, which were collected from the Wikipedia Talk website. The second dataset (Multiclass Cyberbullying Dataset) is openly accessible and collected from social media sites like Twitter, which includes 39,869 samples of tweets divided into five classes. The datasets are preprocessed in several steps, including removing the punctuation, tokenization, removing unnecessary words and emojis, and others. After the preprocessing step, the deep learning modules used BiLSTM and CNN-BiLSTM. The dataset in this research are limited to an English language and there is an overfitting of the proposed models, particularly when using the binary class dataset. The accuracy was 94.1% with the BiLSTM and binary datasets and 99% with the BiLSTM and multiclass datasets.

In [32], the authors employ Convolution Neural Network (CNN) as a deep learning module on Twitter comments as a dataset in the proposed module. The accuracy was 81.60% in the CNN module with GloVe as the feature extraction. The dataset in this study has a small size, so there is a need to use a big dataset to achieve better performance.

In [10], the authors use deep learning modules like LSTM, BiLSTM, GRU, and RNN with the Kaggle dataset from Wikipedia articles. The dataset includes 100,000 comments on Wikipedia articles, and the dataset was preprocessed in several processes, such as text cleaning, tokenization, stemming, lemmatization, and stopword removal. The accuracy was 80.86 % with LSTM, 82.18 % with BiLSTM, 81.46 % with GRU, and 81.01 % with RNN.

The authors in [33] used CNN modules with GloVe as feature extraction on a Twitter dataset. The dataset consists of 69874 tweets. The proposed module achieved an accuracy of 93.97 %.

In [34], The authors employ Feed Forward Neural Networks (FFNN) as deep learning modules with one hot encoding as word embedding. The dataset contains 34,890 records, including 3,015 as bullying and 31,875 as non-bullying. It was preprocessed in several processes to reduce the dataset size to achieve the best results. The accuracy was 93.33% with the FFNN module. Table 1 shows a comparison between some of the related works discussed earlier, where most of their limitations is the insufficient datasize.

Table 1. Comparison between some of the related works.

Paper	Year	Dataset	Size	Feature extraction/ word embedding	Classifier	Accuracy	F1-Score
[32]	2018	Twitter dataset	20,000 random tweets.	GloVe	CNN	81.60%	
[34]	2018	Datasets from Facebook and Twitter	4,913 records	one hot encoding	FFNN	93.33%	
			34,890 records				
[33]	2019	Twitter dataset	69874 tweets.	GloVe	CNN	93.97%	
[29]	2019	Arabic dataset (YouTube and Twitter)	25,000 comments and tweets		Naïve Bayes (NB)	95.9%	92.7%
[7]	2019	Kaggle	24508 records	TF-IDF vectorizer	SVM	90.3%	89.8%
					NN	92.8%	91.9%
[31]	2020	Formspring dataset	12,773 (a Q&A forum)	BERT	BERT-base model	98%	
		Wikipedia talk pages dataset (collaborative knowledge repository)	115,864 discussion comments			96%	
[12]	2020	Formspring	13,110 posts		RF	76.7 %	
		Twitter datasets	13,420 tweets		NB	75.7%	
			8817 tweets		SVM	79.3%	
			24,784 tweets		LR	78.9%	
					Ensemble	79 %	
[10]	2020	Kaggle dataset on Wikipedia articles.	100,000 comments		LSTM	80.86%	
					BiLSTM	82.18%	
					GRU	81.46%	
					RNN	81.01%	

[9]	2020	AJComments-Original	32K	AraVec	CNN-BiLSTM-AVG		44%
				Fasttext	CNN-BGRU-ATT		44%
				TF-IDF vectorizer	SVM		38%
		Balanced AJComments		AraVec	CNN-AVGPOOL	84%	
				Fasttext	BiLSTM-multiCNN-ATT	84%	
					CNN-BiLSTM-MAX	84%	
				TF-IDF vectorizer	RF	85%	
					XG-Boost	85%	
					SVM	85%	
		Unbalanced AJComments		AraVec	MultiCNN-ATT		67%
				Fasttext	CNN-ATTENTION		66%
				TF-IDF vectorizer	Linear SVC		71%
[6]	2021	Arabic YouTube comments dataset	15,000 comments	Count vectorizer	MNB		78.4%.
					CNB		76.6%
					LR		78.6%
				TF-IDF vectorizer	MNB		77.0%
					CNB		78.5%
					LR		76.8%
[3]	2021	Wikipedia	9000 examples	SSWE layer	LSTM	75.5%	
		Twitter				79.1%	
		Formspring				72%	
[28]	2021	Twitter	23462 samples including	TF-IDF vectorizer	ANN	86.32%	86.32%
		WhatsApp			XGB	85.51%	85.87%
		Instagram			NB	73.23%	77.60%
					SVM	87.17%	87.52%
		Vine			RF	87.24%	87.52%
		packet			SDL	83.03%	82.03%
					Consensus-Based Ensemble Model	88.54%	88.82%
[5]	2021	Arabic comments Twitter dataset	17748 comments tweet	TF-IDF vectorizer	SVM	85.49%	
[1]	2022	Kaggle platform: (Binary Aggressive Cyberbullying)	115,864 samples	Keras Embedding layer	BiLSTM	94.1 %	74%
					CNN-BiLSTM	93%	72.3%
		Kaggle platform: (Multiclass Cyberbullying)	39,869 samples		BiLSTM	99%	
					CNN-BiLSTM	95%	

[30]	2022	Twitter datasets	31,947 tweets	TFIDF vectorizer	SVM	73%	76%
					RF	85%	98%
					LG	73%	73%
					KNN	55%	82%
					NB	44%	54%
					MLP	79%	87%
				Semantic	SVM	74%	
					RF	82%	
					LG	74%	
					KNN	74%	
					NB	37%	
					MLP	74%	
				Sentiment	SVM	77%	71%
					RF	93%	97%
					LG	77%	69%
					KNN	83%	82%
					NB	74%	44%
					MLP	77%	82%
				Count Vectorizer	SVM	77%	
					RF	79%	
					LG	75%	
					KNN	51%	
					NB	43%	
					MLP	80%	
				Pragmatic Syntactic	SVM	75%	
					RF	75%	
					LG	75%	
					KNN	75%	
					NB	74%	

4. Discussion

In most papers, BiLSTM shows high efficiency and good accuracy by using deep learning architecture as a classifier compared to other deep learning algorithms. Also, using a BERT-based model as a sentence embedding achieves very good results and high efficiency with a huge dataset. Most papers show the feature extraction or word embedding phase to represent the words (word came from input data and after input in the preprocessing phase) and use modern methods such as GloVe and TF-IDF to represent the words and improve the accuracy. Several papers depend on traditional models to train machine learning and deep learning algorithms. Nevertheless, the model built on a BiLSTM deep learning module shows the highest accuracy with the BERT embedding method at 99% and 98%. In my opinion, the use of sentence embedding like BERT, SentenceBERT, and InferSent with BiLSTM or BiGRU classifiers with large dataset could improve the accuracy.

5. Conclusion

In this paper, we presented a literature review of the scientific papers in the field of detecting the phenomenon of cyberbullying to help researchers to come out with the algorithms and the parameters used in classification and detecting the phenomenon of cyberbullying and to help researchers choose the best feature extractor and classifier to work within their future research. We have concluded from our observations of the results of related works to the trend to use deep learning algorithms in particular trend to use of the BiLSTM classifier and BERT to achieve better results in future research. In the case of using machine learning algorithms, we tend to use SVM and NB as classifiers to achieve better results in future research.

References

[1] Theyazn, H. H.; Mosleh, H. A.; Saleh, N. A. "Cyberbullying Identification System Based Deep Learning Algorithms". Electronics (Switzerland); 11(20), October, 2022.

[2] Mohammad, W. H.; Zainab, N. S. "Twitter Sentiment Analysis Using Different Machine Learning and Feature Extraction Techniques". Al-Nahrain Journal of Science; 2(3): 50-54, September, 2021.

[3] Maheep, M. "Detecting Cyberbullying across Multiple Social Media Platforms Using Deep Learning". International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE: 299-301, March, 2021.

[4] Mukul, A.; Dr, R. E. "Classification of abusive comments in social media using deep learning". 3rd

international conference on computing methodologies and communication (ICCMC), IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, 2019.
[5] Samar, A.; Mohamed, A. F. "Arabic Cyber Bullying Detection Using Arabic Sentiment Analysis". Egyptian Journal of Language Engineering; 8(1), 2021
[6] Tahani, A.; Danyah A.; "Comparison of Machine Learning Techniques for Cyberbullying Detection on YouTube Arabic Comments". IJCSNS International Journal of Computer Science and Network Security; 21(1), 2021.
[7] John, H.; et al.; " Social Media Cyberbullying Detection using Machine Learning". International Journal of Advanced Computer Science and Applications; 10(5): 703-707, 2019.
[8] Noor, A. H.; Ban, N. D.; "The Detection of Sexual Harassment and Chat Predators Using Artificial Neural Network". Karbala International Journal of Modern Science; 7: 301-312, 2021.
[9] Benaissa, A. R.; Harbaoui, A.; Hajjami, H. B.; "Classification of Cyberbullying Text in Arabic". International Joint Conference on Neural Networks (IJCNN), IEEE, 2020.
[10] Celestine, I.; et al.; "Cyberbullying detection solutions based on deep learning architectures". in Multimedia Systems, 2020.
[11] Vishu, T.; Ashwini, K.; Sanjoy, D.; "Sentiment Analysis on Twitter Data Using Deep Learning approach". in Proceedings-IEEE, 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN: 187-190, December, 2020.
[12] Aaminah, A.; Adeel, M. S.; "Cyberbullying Detection Using Machine Learning". Volume: SI, Number: 01: 45- 50, 2020.
[13] Seno, M. E.; Mohammad, O. K.; Ban, N. D.; "CLR: Cloud Linear Regression Environment as a More Effective Resource-Task Scheduling Environment (State-of-the-Art)". International Journal of Interactive Mobile Technologies (IJIM); 16(22): 157-175, 2022.
[14] Mohamed, A. E.; et al.; " Recent Advances in NLP: The Case of Arabic Language". Springer studies in Computational Intelligence; 874, 2019.
[15] Aditya, J.; Gandhar, K.; Vraj, S.; "Natural Language Processing". International Journal of Computer Sciences and Engineering; 6(1), January, 2018.
[16] Jincheng, X.; Qingfeng, D.; "A Deep Investigation into Fast Text ". 2019 IEEE 21st International Conference on High Performance Computing and Communications, IEEE 17th International Conference on Smart City, IEEE 5th International Conference on Data Science and Systems, 2019.
[17] Mohamed, A. E.; Zainab, N. S.; "A Review of Machine Learning Approach for Twitter Sentiment Analysis". Al-Nahrain Journal of Science; 24(4): 52-58, December, 2021.
[18] Innocent, C.; et al.; "Predicting HIV Status among Men Who Have Sex with Men in Bulawayo & Harare, Zimbabwe Using Bio-Behavioural Data, Recurrent Neural Networks, and Machine Learning Techniques". Trop Med Infect Dis; 7(9), September, 2022.
[19] Amer, Noor, and Ban N. Dhannoon; "Machine learning approaches to detect online harassment using bag of words." In AIP Conference Proceedings, 2457(1): 040009. AIP Publishing LLC, 2023.
[20] Theyazn, H. H.; et al.; "Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models". Int J Environ Res Public Health; 19(19), October, 2022.
[21] Tahseen, A. W.; Ban, N. D.; "Improving Prediction of Arabic Fake News Using Fuzzy Logic and Modified Random Forest Model". Karbala International Journal of Modern Science; 8(3): 477-485, 2022.
[22] Paritosh P.; Anju B.; Prashant, S. R.; "Consensus based Ensemble model for Spam detection". International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015.
[23] Rahim, B.; Mohammad, T. A.; Jan, A.; "Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model". Stochastic Environmental Research and Risk Assessment; 34(2): 415-433, February, 2020.
[24] Ahmad, M. S.; Ban, N. D.; "Color Model Based Convolutional Neural Network for Image Spam Classification" Al-Nahrain Journal of Science; 23(4): 44-48, December, 2020.
[25] Ru, N.; Huan, C.; "Sentiment Analysis based on GloVe and LSTM-GRU". Proceedings of the 39th Chinese Control Conference; July, 2020.
[26] Xuan, H. L.; et al.; "Comparison of Deep Learning Techniques for River Streamflow Forecasting". IEEE Access; 9: 71805-71820, 2021.
[27] Xiaomei, W.; et al.; "Intelligent Hybrid Deep Learning Model for Breast Cancer Detection". Electronics (Switzerland); 11(17), September, 2022.
[28] Asma, A. A.; Abdulbasit, A. D.; "Consensus-Based Ensemble Model for Arabic Cyberbullying Detection". Computer Systems Science & Engineering CSSE; 41(1), 2022.
[29] Djedjiga, M.; et al.; "Detection of Arabic Cyberbullying on Social Networks Using Machine Learning". 16th ACS/IEEE International Conference on Computer Systems and Applications AICCSA, Al Ain University & Crowne Plaza, Abu Dhabi, UAE: 03-07, November, 2019.
[30] Hadiya, B.; "Cyber Bullying Detection in Twitter using Machine Learning Algorithms". Journal of Computer Science IJCSIS, 2022.
[31] Jaideep, Y.; Devesh, K.; Dheeraj, C.; "Cyberbullying Detection using Pre-Trained BERT Model".

Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020): 02-04, July, 2020.

[32] Monirah, A. A.; Mourad, Y.; "Optimized Twitter Cyberbullying Detection based on Deep Learning". 21st Saudi Computer Society National Computer Conference: SCS-NCC' 2018: Riyadh, Kingdom of Saudi Arabia: 25-26, April, 2018.

[33] Vijay B.; Jui, T.; Pooja, G.; Pallavi, V.; "Detection of Cyberbullying Using Deep Neural Network". 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019.

[34] Batoul, H.; Maroun, C.; Ahmed, S.; "Arabic Cyberbullying Detection: Using Deep Learning". 7th International Conference on Computer and Communication Engineering (ICCCE 2018): Kuala Lumpur, Malaysia. IEEE: 19-20, September, 2018.