IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 5, October 2025, pp. 4211~4225
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i5.pp4211-4225

Exploring social media sentiment patterns for improved cyberbullying detection

Wael M. S. Yafooz[1], Abdulsamad Ebrahim Yahya[2], Abdullah Alsaeedi[1], Reyadh Alluhaibi[3], Faisal Jamil[4], Mahmoud Salaheldin Elsayed[2]
[1]Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia
[2]Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia
[3]Department of Artificial Intelligence and Data Science, College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia
[4]Department of Computer Science, School of Computing, Engineering and Intelligent Systems, University of Ulster, Belfast, United Kingdom

Article Info

Article history:

Received Jul 3, 2024
Revised Jul 5, 2025
Accepted Aug 6, 2025

Keywords:

Cyberbullying
Deep learning
Machine learning
Social media
Transformers

ABSTRACT

Cases of online bullying and aggressive behaviors directed at social media users have surged in recent years. These behaviors have had negative impacts on victims from a wide range of demographic groups. While efforts have been made to address persistent digital harassment, the expected outcome has been limited due to the lack of effective tools to quickly identify cyberbullying behaviors and censor them accordingly on social media platforms. This study presents a scalable and systematic method to detect and analyze offensive behavior and bullying on Twitter (now known as X). Our methodology involves extracting textual, user-related, and network-related attributes to understand the traits of individuals involved in such behaviors. This approach aims to recognize distinctive characteristics that set them apart from regular users. This study proposes a novel model by employing an integrated deep-learning model, combining the bidirectional gated recurrent unit (BiGRU), transformer block, and convolutional neural network (CNN). This model aims to classify X comments into offensive and non-offensive categories. The proposed model's efficacy has been evaluated through several experiments by combining three widely recognized datasets of hate speech. The proposed model achieves an accuracy rate of approximately 98.95%, showing promising results in identifying and categorizing offensive behavior in cyberbullying.

Corresponding Author:

Wael M. S. Yafooz
Department of Computer Science, College of Computer Science and Engineering, Taibah University
Medina 41477, Saudi Arabia
Email: wyafooz@taibahu.edu.sa

1. INTRODUCTION

Cyberbullying is a significant threat to online users. It particularly affects individuals who frequently visit social media platforms [1]–[3]. Unlike traditional bullying, cyberbullying occurs electronically without temporal or spatial limitations [4]–[7]. During the early 20th century, when social media was in its nascent stage, the negative impacts of cyberbullying were not fully acknowledged. However, in the last 15 years, cases of online harassment, cyberstalking, and trolling have been on the rise [8]. In 2017, 41% of US residents experienced some level of online harassment, and 66% experienced some type of hate speech, which was directed at them. Approximately 50% of young people use social media today face cyberbullying on platforms like Twitter (now known as X). Despite attempts to tackle ongoing digital harassment, the results have been constrained by the absence of efficient tools to promptly identify cyberbullying behaviors and take appropriate censorship measures [9], [10]. There has been an increase in interest in detecting cyberbullying in the field of natural language processing (NLP) [11]. A primary goal of cyberbullying detection is to pre-process textual data, such as tweets cleaning, and extract relevant information so that machine learning (ML) algorithms can be used to trained models to understand and classify them. Many methods are used to simplify text representation and translate words into numerical values in text classification/categorization, such as bag of words (BoW), term frequency-inverse document frequency, and Word2Vec, GloV, and fast text. These types of numerical data can be fed into ML classifiers such random forest (RF), decision tree (DT), XGBoost (XGB), k-nearest neighbors (KNN), and support vector machine (SVM), and recurrent neural networks (RNNs) [12]. Although conventional NLP methods have shown considerable efficacy in identifying instances of cyberbullying on social media platforms, several challenges persist. These encompass limitations imposed by character restrictions on social media, differences between offensive and non-offensive comments, inherent ambiguities in natural language, and the extensive use of slang [13].

Over the years, neural network-based models have outperformed conventional ML techniques in various NLP tasks. This superiority owes much to the rich vector representations offered by neural networks, particularly the significant strides made in word embeddings [14]. Unlike conventional ML methods relying heavily on manually crafted, potentially incomplete features that are time-consuming, deep learning (DL) techniques employ hierarchical automatic feature extraction to understand input characteristics. Recent advances in NLP have been driven by the use of neural network models like multilayer perceptrons (MLPs), RNNs, and convolutional neural networks (CNNs). These models have demonstrated promising results across various NLP tasks. Xiao and Cho [15] proposed a novel methodology for text classification that operated at the character level. They incorporated a hybrid model, combining a CNN framework with an RNN architecture. Tai *et al*. [16] proposed another innovative method. They employed long short-term memory (LSTM) for sentence-by-sentence classification to comprehend the semantics of the text and used a CNN to extract local features from the expressions.

Many researchers have explored the detection of cyberbullying in Arabic [6], [17]–[23]. For instance, Haidar *et al*. [17] initially explored cyberbullying detection in Arabic, while Salawu *et al*. [18] classified detection approaches into categories like supervised learning, lexicon-based, rule-based, and mixed-initiative techniques. Rosa *et al*. [6] identified misinterpretations and lack of standardized evaluation methodologies in previous research. In contrast, Al-Garadi *et al.* [19] emphasized crucial factors in detecting aggressive behavior, and Elsafoury *et al.* [20] highlighted the importance of slang-based word embedding techniques. Kim *et al.* [21] stressed human involvement in algorithm development, while Al-Harigy *et al.* [23] focused on prompt and effective cyberbullying detection. Table 1 presents a comprehensive comparative overview of the aforementioned studies, elucidating notable characteristics and methodological strategies.

Table 1. Critical analysis of existing cyberbullying approaches with proposed model

References	DL approaches		Data representation		Data availability
	Cyberbullying application	Limitation and strength	Text	Images	
[21]	✓	✓	✓	×	✓
[22]	×	×	✓	✓	✓
[25]	✓	×	✓	✓	×
[24]	×	×	✓	×	✓
[20]	×	×	✓	×	×
[23]	×	×	✓	✓	✓
[24]	✓	×	✓	✓	×
Proposed	✓	✓	✓	✓	✓

The current surveys on the detection of cyberbullying using DL techniques have certain restraints, as there is no comprehensive survey specifically focused on this field. The lack of comprehensive analysis regarding the strengths and weaknesses of DL models in accurately classifying cyberbullying instances is attributed to the inefficiency of most survey papers. The current surveys do not encompass the taxonomy of DL-based cyberbullying classification, which is crucial for organizing and expanding complex concepts. This study examines image-based data representation techniques, acknowledging their importance in cyberbullying classifying. The careful choice of a suitable framework is essential for the effective implementation of a robust model for classification. The availability of easily obtainable datasets is crucial for researchers to assess the plausibility of their hypotheses. The paper also explores additional factors such as cultural diversity, data representation, multimedia and multilingual content, and the implications for mental well-being. Engaging in these discussions is essential to develop a comprehensive understanding of the challenges and future trends within this field.

Therefore, this study aims to investigate the possibilities of categorizing brief textual content. This investigation is driven by the notable achievements witnessed in diverse DL models, specifically in tasks involving extensive textual content. Our investigation centers on a multichannel DL model, which incorporates elements from three sophisticated DL architectures: bidirectional LSTM (BiLSTM), transformer block, and CNN. Employing advanced NLP and ML classifiers on social networks like X plays a crucial role in automating cyberbullying identification. However, the availability of suitable datasets for training ML classifiers limits this possibility. This limitation underscores the importance of obtaining an extensive dataset encompassing various cyberbullying instances. The purpose of this paper is to present a multichannel technique, a fusion of three DL models, to enhance prediction accuracy. To evaluate the proposed approach and assess the accuracy levels, we combine three extensively acknowledged datasets on hate speech [24]–[26].

The findings indicate that the proposed method can achieve notable accuracy. The contributions of this research are as follows: this study introduces a novel hybrid model which consists of: transformer block, BiGRU, and CNN all within the multichannel technique, to significantly enhance the prediction accuracy in detection offensive and non-offensive tweets. This study demonstrates the achievement of notable levels of prediction accuracy by integrating three widely recognized hate speech datasets. This study enhances online platforms' detection capabilities, empowering them to strengthen cybersecurity measures and protect users from cyberbullying.

The reminder of the paper is organized as follows: section 3 introduces the integrated DL approach. Section 4 outlines the dataset used for evaluation purposes. Section 5 presents, analyzes and compares the results with closely related approaches. Section 6 encapsulates the conclusion of the research paper, followed by a discussion of potential avenues for future research in section 7.

2. METHODOLOGY

This section presents the proposed intelligent cyberbullying detection platform. The platform is organized into two main components: data analysis and predictive analytics. The platform functions as a robustartificial intelligence and text-mining solution specifically developed for classifying of cyberbullying into offensive and non-offensive tweets. Figure 1 depicts the model's structure for categorizing cyberbullying tweets and the user detection system which was developed to forecast offensive and non-offensive tweets. Furthermore, the system tackles the problem of anonymity in cyberbullying through the implementation of user identification techniques. This solution includes the examination of user behavior patterns, linguistic style, and additional contextual cues to establish distinct user profiles. By implementing this approach, the system can effectively monitor and track potential or existing cyberbullies, deter the occurrence of abusive conduct, and foster the development of an online community characterized by enhanced levels of respect.

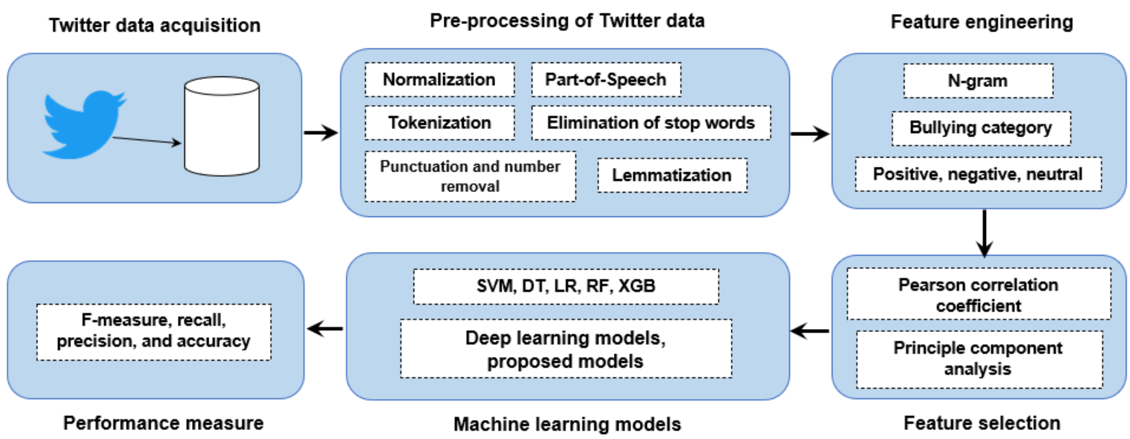


Figure 1. Architecture of proposed intelligent detection of cyberbullying model

The platform employs a hybrid approach, leveraging semantic embeddings, contextual similarity scores, and predictive models to enhance accuracy in classifying tweets. The data analysis module consists of data acquisition, preprocessing, feature extraction, and feature selection steps. It incorporates embedding-based semantic similarity measures, including contextual embedding overlap similarity (CEOS), which enhances the platform's ability to discern subtle nuances in tweets. The predictive analytics module, on the other hand, includes training and testing ML and DL models, optimized for high precision in detection tasks. The methodology integrates DL classification techniques to categorize the tweets and users into two offensive and non-offensive categories. There are five steps in the data analysis module: tweet acquisition from the Twitter API (tweetInvi), now known as X API, tweet pre-processing, feature extraction, feature normalization, and feature selection. Predictive analytics involves training and testing various DL models, it also involves optimizing hyperparameters.

2.1. X data acquisition

The data from X was obtained using the tweetInvi API to detect online cyberbullying. The datasets were comprised of tweets and user networks. These were extracted from textual data, which only are tweets, hastages, comments, and photos description. A dataset of 20,000 tweets as shown in Figure 2, along with userIDs, hashtags, dateTime, and location of origin, was collected in JSON format. Two separate labels for tweets were created to distinguish between offensive and non-offensive tweets. Table 2 provides a detailed breakdown of the dataset's key attributes.

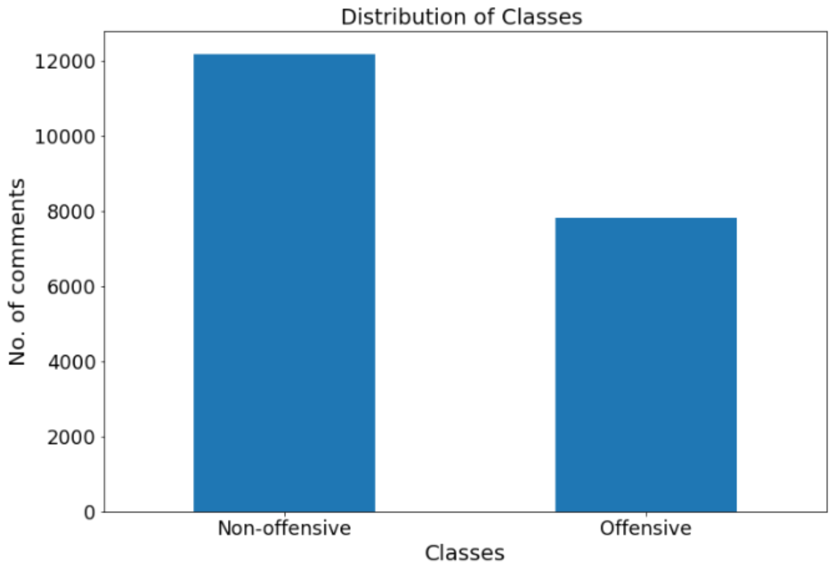


Figure 2. Cuberbullying dataset includes two classes non-offensive and offensive

Table 2. Dataset description that includes features and data count

Sr no	Features	Application programming interface	Data count	Data description
1	Tweet	/tweetinvi/twitter-api/tweets	5,000	The information which users share together
2	HashTags	/tweetinvi/twitter-api/Hashtags	1,000	Shared hashtags projects opening on.
3	Comments	/tweetinvi/twitter-api/Comments	5,000	Comments for shared topic
4	Photos	/tweetinvi/twitter-api/Photos	9,000	Shared photos by various users

2.2. Pre-processing of X data

Data pre-processing plays a vital role in guaranteeing the accuracy and dependability of unprocessed data. Accuracy is enhanced by transforming unprocessed JSON files into a standardized format. Afterward, the textual data is subjected to various pre-processing techniques in NLP. The X data undergoes the following steps.

- Normalization: in this step, texts were converted into a standardized format to improve data quality and enable efficient system processing. Normalization in our system was accomplished through several techniques. First, the duplicate white spaces were eliminated. Next, the text was converted to lowercase. After that, contractions were expanded, and word numerals were converted to their corresponding numerical values. The normalized was then inputted into the tokenization module.

– Tokenization: in this step, paragraphs and sentences were divided into individual token texts. This division facilitates the assignment of semantic meaning to the tokens, thereby enhancing the overall understanding and analysis of the text. Keras, an open-source deep-learning library, was employed for the purpose of text pre-processing, specifically for tokenization. The list of tokens obtained was subsequently crucial for the next pre-processing step.
– Punctuation and number removal: punctuations and numbers were removed in this step. Subsequently, the curated list was transferred to the subsequent phase.
– Lemmatization: in this step, words were transformed into their base forms by considering morphological analysis. For instance, the term ‘computers’ was transformed into the singular form ‘computer’. As part of our model, we used the WordNet lemmatizer provided by the natural language toolkit (NLTK).
– Part-of-speech (POS): this step involved assigning POS tags words to enhance semantic similarity. In our proposed model, we employed POS tagging to improve the precision of semantic similarity.
– Elimination of stop words: in this step, frequently used structural words were removed to optimize text mining. In the ISE model, the removal of stop words was implemented to prioritize relevant information, optimize text mining procedures, and mitigate intricacy.

2.3. Feature engineering

Within this section, the utilization of the feature engineering technique has been implemented to produce additional data from the given dataset. The latent features that have been identified and derived within the proposed platform. These derived features are outlined in Table 3.

Table 3. Derived features and their description

Sr no	Type	Feature name	Description	Reference
1	Embedding feature vector	Work2Vec	The tweet-level feature representation is derived by employing pretrained Word2Vec embeddings.	[27]
2	Sentiment feature vector	SentiStrength	This feature is utilized to evaluate each tweet’s positive and negative sentiment scores.	[28]
3		MPQA subjectivity lexicon	Derives lexicon features at the phrase level to determine the positive and negative contextual polarity of sentiment expression in tweets. Sentiment140 is a collection of tweets labeled positive, negative, or neutral for sentiment analysis.	[29]
4	Lexicon feature vector	Sentiment-140		[30]
5		BingLiu	Using feature mining, customer reviews are categorized as favorable or negative without sentence selection.	[29]
6		AFINN	English words for emotion polarity: positive, negative, or neutral. It scores words numerically for sentiment analysis and text mining.	[31]
7		Expanded NRC-10	Deployed to classify tweets’ emotions or doing sentiment analyses.	[32]
8		NRC hashtag sentiment lexicon	NRC hashtag sentiment lexicon for hashtag-focused social media sentiment research. It identifies frequent social media terms as favorable, bad, or neutral.	[33]
9		SentiWordnet	SentiWordNet ranks English words’ meanings as positive, negative, or neutral. It simplifies NLP sentiment analysis with WordNet synsets and sentiment scores.	[34]
10		NRC-10	The NRC emotion lexicon classifies words into 10 fundamental emotions for text data sentiment analysis and classification.	[35]
11		NRC hashtag emotion association lexicon	The NRC hashtag emotion association Lexicon helps social media data analysts analyze emotions by linking hashtags to emotions.	[36]
12	Knowledge based similarity	CEOS	Computes similarity between tweets embeddings and reference embeddings for offensive and non-offensive content.	Proposed

The proposed method evaluates semantic similarity by leveraging contextual embeddings to measure the overlap between a tweet and predefined reference vectors, representing offensive and non-offensive language. This approach eliminates reliance on frequency-based mechanisms and instead focuses on the semantic relationships in a high-dimensional embedding space. To calculate the similarity, embedding vectors are generated for the input tweet. Let $\vec{E}_{Tweet}$ represent the embedding vector for the tweet, aggregated from word-level embeddings as in (1), where $\vec{E}(w_i)$ is the embedding of the it-th words and n_w words in the tweets. Two reference vectors are precomputed: offensive $\vec{E}_{offensive}$ for offensive language, as shown in (2), and non-offensive $\vec{E}_{non-offensive}$ for non-offensive language as presented in (3). The similarity between the tweet embedding and each reference vector is calculated using cosine similarity. Finaly, the CEOS for a tweet is then calculated as in (4). A positive CEOS_{Tweet} suggests a higher likelihood of offensive content, while a negative score indicates non-offensive tendencies. This method focuses on semantic relationships, offering a robust and context-aware mechanism for distinguishing offensive and non-offensive language.

$$\vec{E}_{Tweets} = \frac{1}{n_w} \sum_{i=1}^{n_w} \vec{E}(w_i) \tag{1}$$

$$Similarity_{offensive} = \frac{\vec{E}_{Tweet} . \vec{E}_{offensive}}{\|\vec{E}_{Tweet}\| \|\vec{E}_{offensive}\|} \tag{2}$$

$$Similarity_{non-offensive} = \frac{\vec{E}_{Tweet} . \vec{E}_{non-offensive}}{\|\vec{E}_{Tweet}\| \|\vec{E}_{non-offensive}\|} \tag{3}$$

$$\text{CEOS}_{Tweet} = Similarity_{offensive} - Similarity_{non-offensive} \tag{4}$$

2.4. Feature selection

To identify the most relevant features for the classification mechanism, we conducted a thorough analysis of the prepared dataset in this module. The inclusion of this step is of utmost importance in the process of getting rid of a potentially expansive feature space. By removing unnecessary characteristics that do not substantially contribute to the analysis procedure, this process optimizes data processing efficiency. Furthermore, it facilitates the identification of data features with the highest and lowest levels of influence, as determined by their respective weights.

In this study, we utilized two well-established methodologies known as principal component analysis (PCA) and Pearson correlation coefficient (PCC) to enhance the selection of initial features, as described in Table 3. In addition, the newly introduced CEOS score, derived from pre-trained embedding models such as BERT or GloVe, has emerged as a critical feature, offering semantic context to the classification process. PCA and PCC primarily emphasize the selection of linear variables. PCA is a technique that is used to calculate the fundamental components, which aids in reducing the dimensionality of a complex feature space to a more manageable one. This process ensures that crucial feature information is retained, as depicted in Figure 3. The CEOS score demonstrates a high correlation with the target variable in PCC analysis (Figure 4), reinforcing its significance as a feature for distinguishing offensive and non-offensive tweets.

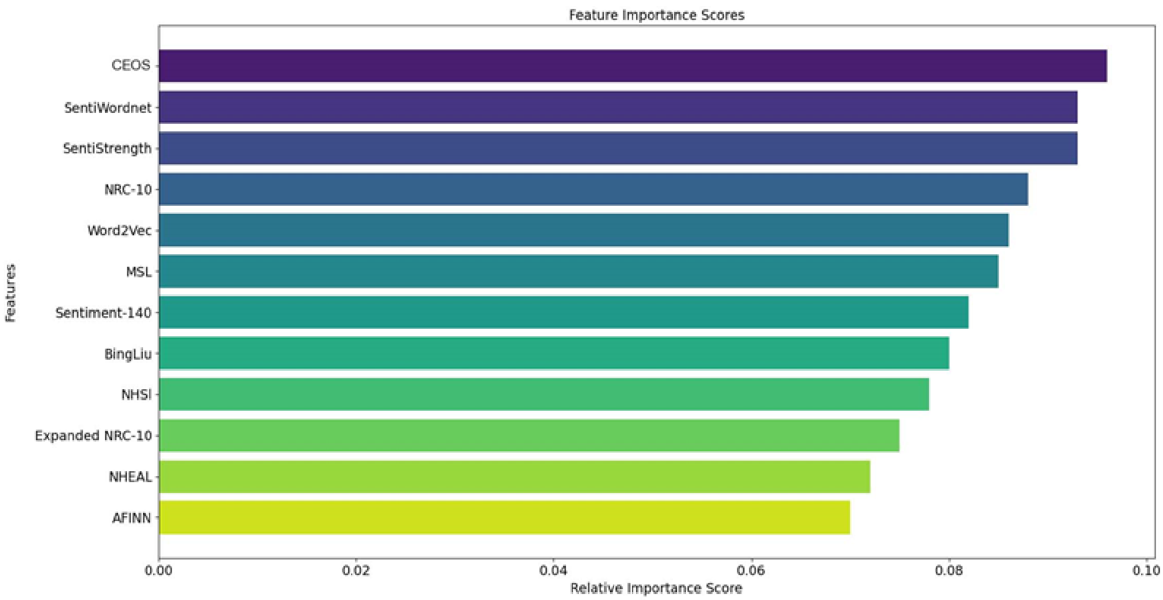


Figure 3. Feature importance analysis based on PCA

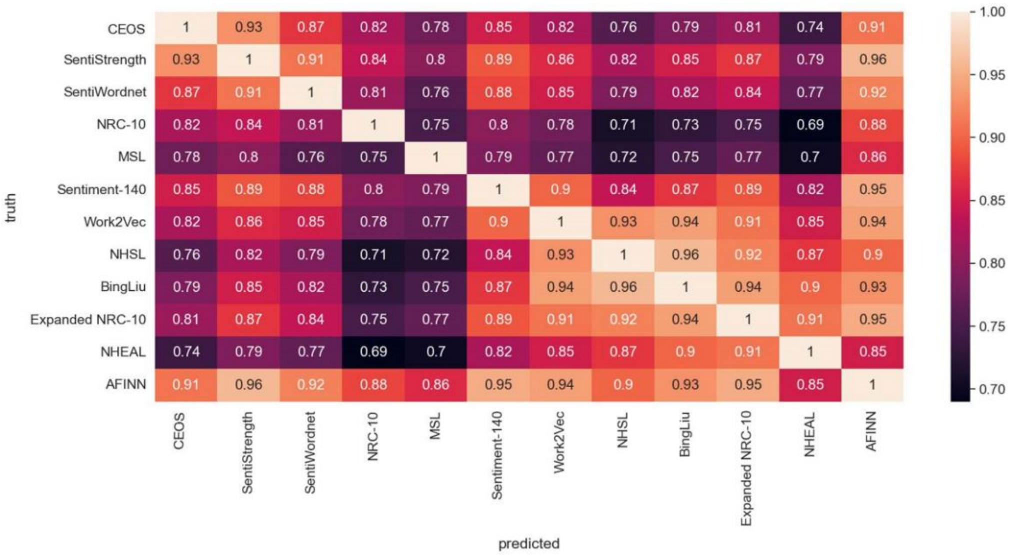


Figure 4. Pearson correlation analysis of input features

2.5. Predictive analysis of X data for cyberbully classification based on ML algorithms

This section focuses on our proposed cyberbullying detection system's ML classifiers. The purpose of this study is to perform a thorough comparative analysis of the efficacy of various ML classifiers in differentiating between offensive and non-offensive tweets. The proposed platform incorporates a range of classification models for supervised data classification, as detailed in Table 4.

In order to thoroughly assess the performance of the classifier, we implemented a k-fold cross-validation methodology, utilizing a value of k equal to 10. This methodology systematically evaluates and compares the outcomes of the applied classifiers. The dataset was divided into k equal segments, where each segment was utilized for training a classifier.

Table 4. Applied classifiers on the label dataset

SR no #	ML classifiers
1	XGB
2	SVM
3	DT
4	Logistic regression (LR)
5	RF
6	KNN

Each iteration used a different segment for testing, while the remaining k-1 segments were allocated for training. The deliberate partitioning of the dataset into distinct training and testing sets is implemented to address the problem of overfitting and improve the accuracy of classification. Each classifier was evaluated by measuring the proportion of correctly classified instances out of the total number of classifications performed by the classifier.

In this study, we employed various ML classifiers, such as RF, DT, XGB, KNN, and SVM [37], [38] with quadratic and linear kernels, to forecast the classification of tweets. Every algorithm function based on the principles of supervised learning, which requires a dataset for training and allows for the prediction of class labels for instances that are not known. In conclusion, a range of performance metrics, including precision, f-measure, accuracy, and recall, were utilized to evaluate the efficacy of the implemented ML classifiers.

This study introduces a multichannel DL framework that integrates three advanced models: transformer block, BiGRU, and CNN to enhance the accuracy of cyberbullying detection by leveraging their complementary strengths as shown in Figure 5. Each component plays a distinct role in capturing different aspects of the textual data, contributing to a holistic feature representation. The transformer block is designed to capture long-range dependencies and relationships within textual data using self-attention mechanisms. By processing sequences in parallel, the transformer effectively captures global contextual information, allowing the model to uncover intricate patterns in the input text. This capability is critical for understanding subtle cues in tweets that may indicate offensive behavior. Complementing this, the BiGRU processes sequences bidirectionally, retaining contextual information from both past and future words in a sentence. This bidirectional approach ensures that the model comprehends the broader context around each word. The BiGRU's outputs are further refined through spatial dropout, global average pooling, and global maximum pooling, which enhance feature retention while minimizing noise.

The CNN component focuses on extracting local and spatial features from the input text, such as n-grams (e.g., bigrams and trigrams), which are critical for identifying offensive language. By applying convolutional filters, the CNN identifies patterns that other models might miss, such as localized phrases or word groupings. Rectified linear unit (ReLU) activation introduces non-linearity to enhance the model's learning capacity, while a 50% dropout rate reduces overfitting and improves generalization.

To achieve a unified representation, the outputs of the transformer block, BiGRU, and CNN are concatenated into a single feature vector, denoted in (5), where T, B, and C represent the individual outputs of the respective models. This concatenated feature vector is then processed through two fully connected dense layers for further feature integration and dimensionality reduction. The transformations applied by the dense layers are given in (6) and (7), where W_1 and W_2 are weight matrices, b_1 and b_1 are biases, and (f) represents the ReLU activation function.

$$V = [T, B, C] \tag{5}$$

$$D_1 = (W_1 . V_1 + b_2) \tag{6}$$

$$D_2 = f(W_2 . D_1 + b_2) \tag{7}$$

The configuration of the two dense layers, with 60 and 30 neurons respectively, was determined through empirical experimentation and hyperparameter tuning. Various configurations were tested to achieve an optimal balance between model complexity and performance. This configuration provided the best results in terms of accuracy and generalization, as reflected in the evaluation metrics. The final classification into offensive or non-offensive categories is performed by a softmax activation layer, defined in (8). To optimize the model's predictions, the binary cross-entropy loss function BCE is employed, which quantifies the classification error. This is formulated in (9).

$$P(y) = Softmax(W_3 . V_3 + b_3) \tag{8}$$

$$BCE = \sum_c y_c \log(S_0(x)_c) \tag{9}$$

The end-to-end workflow begins with preprocessing and tokenization of input tweets to ensure compatibility with the DL models. The preprocessed text is simultaneously fed into the transformer block, BiGRU, and CNN, each extracting unique features that are fused into the concatenated vector v. This vector is refined through the dense layers, and the final classification is achieved in the softmax layer. By integrating global, sequential, and local features, this multichannel framework captures a comprehensive representation of the data, resulting in improved prediction accuracy. The overall architecture, illustrated in Figure 5, highlights the interactions among the components and demonstrates how their integration enhances the model's performance in detecting offensive behavior on social media platforms.

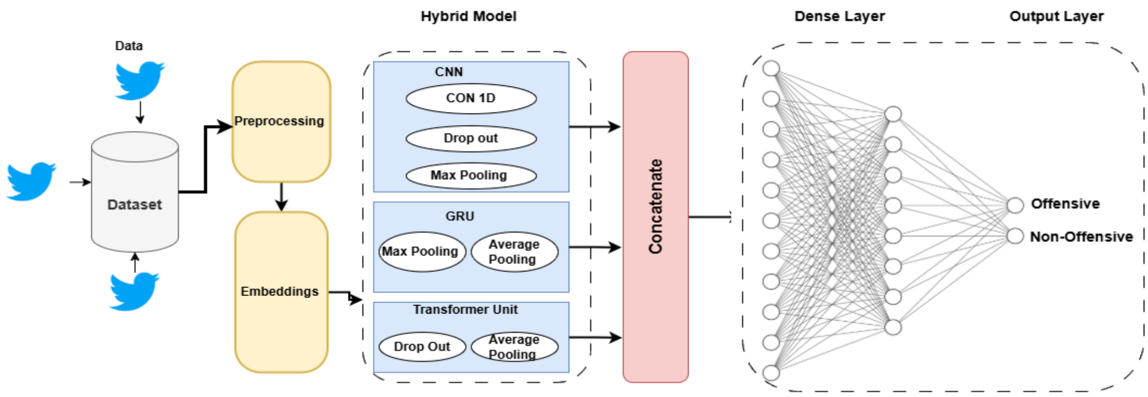


Figure 5. Proposed DL framework employing a combination of BiGRU, transformer unit and CNN within a multichannel structure for intelligent cyberbullying detection

3. IMPLEMENTATION ENVIRONMENT

This section presents a comprehensive overview of the various tools and technologies utilized in creating the proposed cyberbullying platform (Table 5). Python version 3.6 was utilized as the fundamental programming language to develop the platform. The powerful integrated development environment (IDE) known as Visual Studio Code 2020 was employed to enhance the functionality of high-level programming interfaces. This environment provided support for the expansion of fundamental libraries such as Keras, TensorFlow, and Scikit-learn.

The process was initiated by retrieving tweets from the X API using the tweetInvi library. After that, the tweets obtained were subjected to tokenization using the nltk.tokenize library. This process resulted in the generation of a list of tokens. The token lists were subsequently utilized to generate n-grams (specifically, uni-gram, bi-gram, and tri-gram) by employing the nltk.ngram library. After computing these features, the PCA libraries were imported into the designated platform. The highest-performing characteristics were then employed to train and evaluate different ML classifiers and DL models by utilizing the Keras, TensorFlow, and Scikit-learn libraries. Various classification algorithms based on ML have been utilized to assess the efficacy of the cyberbullying platform under consideration. The outcomes of each classification algorithm undergo a rigorous validation process using k-fold cross-validation.

Table 5. Proposed implementation environment of cyberbullying detection

Component	Tools and technologies	Description
Used hardware	Programming language	Python version 3.6
	IDE	VS Code 1.74 V
	RAM	18 GB0
	OS	Windows 10
	CPU	Intel (R) Core (TM) i7-6700 CPU 3.40 GHz
Core libraries	Matplotlib, NLKT, Keras	Data visualization libraries text analysis support toolkit and libraries NN libraries support
	Pandas	Data preparation Pandas libraries
	Sklearn	ML support library for classification
	TweetInvi	X API support library

4. EXPERIMENTAL RESULTS

This section presents the empirical findings of the proposed cyberbullying detection platform. The platform was organized based on four distinct models. The factual decision attributes were used to forecast tweets' categories within the traditional ML classification model. Subsequently, the feature selection methodology was utilized to identify relevant features for forecasting offensive and non-offensive tweets. This identification was accomplished by employing the aforementioned methods. The results of these experiments are presented in Figure 6.

Critical analysis of the graph shows that each classification model utilizing the current set of features achieved a true classification accuracy rate exceeding 77%. The SVM and KNN classifiers were noteworthy due to their remarkable precision in predicting outcomes. The SVM classifier achieved an accuracy of 94%, while the KNN classifier achieved an accuracy of 91.5%. Comparative analysis among different models revealed an accuracy of 89%, 88.7%, 87.2%, and 83.8% for DT, RF, and XGB, respectively as presented in Table 6.

Furthermore, when utilizing a subset of features, the predictive accuracy of all ML classifiers surpassed a threshold of 78.7%. To train and test the ML models, PCA and PCC were used to select the relevant subset of decision features. Model complexity and overfitting were effectively mitigated by the feature reduction approach. The graph shows the positive impact of implementing this strategy on the overall accuracy of the ML models. Th model's performance was superior to other classifiers, achieving an impressive accuracy rate of 98%. Similarly, the KNN, DT, RF, XGB, and LR models had an accuracy of 92%, 92%, 90.3%, 88%, and 84.5%, respectively. Multiple assessment metrics were calculated to evaluate the models' performance, such as accuracy, precision, recall, and F1-score [39].

Accuracy is defined as the proportion of correctly predicted occurrences (true positive) out of the total number of instances. The measure of accuracy for the multi-classification problem is shown in (10). Using the number of true positive instances divided by the total number of instances, we can calculate the accuracy of a model. Recall quantifies the proportion of expected positive outcomes that are accurately identified as positive. In a multi-classification issue, the *i*-th label of the class is determined by summing the values in a column of the confusion matrix. The calculation of recall is determined in (11).

$$\text{Accuracy} = \frac{True\ Positive}{Total\ Instance} \tag{10}$$

$$Recall = \frac{True\ Positive}{Total\ Actual\ Positive} \tag{11}$$

Divide the number of true positives by the total number of actual positives to calculate the recall. Precision refers to the proportion of accurately anticipated positive cases out of all the actual positive cases. Summing the values in a column of the confusion matrix determines the j-th label of the class in a multi-classification issue. Accuracy is defined as the ratio of the number of true positive predictions to the number of true negative predictions. The ($TruePositive_j$) predicted positive instances as a proportion of the total ($TotalPredictedPositive_j$) as calculated in (12). Let m be a matrix, where i represents the rows (predicted label) and j represents the columns (actual label).

$$Precision = \frac{True\ Positive_j}{Total\ Predicated\ Positive_j} \tag{12}$$

To evaluate the robustness and generalization capability of the proposed model, we employed k-fold cross-validation k=10. In this method, the dataset was divided into 10 equal subsets, or folds. For each iteration, one-fold was used as the testing set, while the remaining k-1 folds were used for training. This process was repeated 10 times, ensuring that every data point was used for both training and testing exactly once. The performance metrics, including accuracy, precision, recall, and F1-score, were averaged across all folds to provide a comprehensive evaluation of the model's performance. This approach minimizes the risk of overfitting and provides a reliable estimate of the model's effectiveness across different data splits.

In Table 6, both the whole and simplified feature sets are illustrated for the experimental models. In terms of accuracy, precision, and recall, the SVM consistently outperformed other classification models in all classes. On average, the SVM consistently shows the highest level of accuracy. In contrast, the KNN exhibited improved prediction outcomes, specifically in relation to memory. However, LR had the lowest performance compared to the other classification models. It showed poor performance across all assessment measures, irrespective of employing full or reduced decision feature sets.

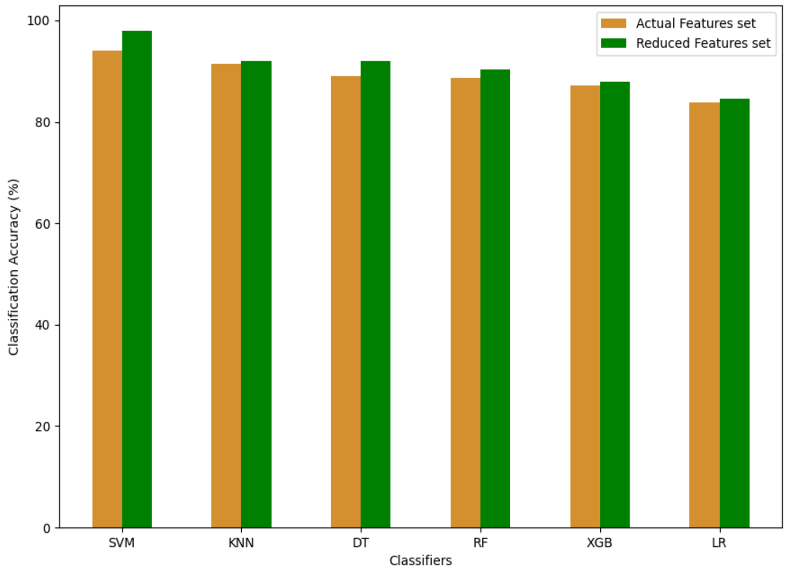


Figure 6. Tweets classification (offensive and non-offensive) results using actual and reduced features set

Table 6. Evaluating classification models based on both full and reduced features

Classifiers\metrics	Actual features set performance measures				Reduced features set performance measures			
	F1	Accuracy	Recall	Precision	F1	Accuracy	Recall	Precision
SVM	0.94	94	0.935	0.921	0.98	98	0.961	0.981
KNN	0.93	91.5	0.903	0.895	0.92	92	0.905	0.932
DT	0.90	89	0.901	0.892	0.92	92	0.901	0.881
RF	0.91	88.7	0.89	0.86	0.91	90.3	0.890	0.891
XGB	0.89	87.2	0.881	0.8752	0.89	88	0.853	0.865
LR	0.84	83.8	0.842	0.821	0.86	84.5	0.824	0.823

Furthermore, a comparative analysis was performed with well-established ML algorithms to evaluate the precision of the proposed approach. The algorithms are known for their exceptional accuracy in various NLP applications, such as the CNN, BiLSTM, and transformer block. Tokenization was utilized as a preliminary step for all of these methods. Table 7 compares the results of the technique and other algorithms. As shown in Table 7, the proposed model achieves superior performance compared to the other algorithms when the data is divided into 75% for training and 25% for testing.

Our solution has a remarkable accuracy of 98.95%. This accuracy level surpasses the next most accurate algorithm (BiGRU) by approximately one percentage point and outperforms the third- and fourth-ranked algorithms by about two percent. Furthermore, we performed a thorough assessment by considering precision, recall, F1-score, and a confusion matrix, as well as accuracy. In non-offensive cases, our suggested technique showed impressive performance with an accuracy of 96%, recall of 98%, and an F-score of 98%. Our technique demonstrated exceptional performance in identifying offensive situations, with an accuracy rate of 96%, a recall rate of 98%, and an F1-score of 87%. It accurately categorized around 98% of non-offensive cases (15,700 out of 20,000 samples) and 97% of offensive cases (14,322 out of 15,000 samples).

Table 7. Evaluating the performance of DL models using both the full set of features and a reduced feature set

Classifiers\metrics	Actual features set performance measures				Reduced features set performance measures			
	F1	Accuracy	Recall	Precision	F1	Accuracy	Recall	Precision
Transformer block	0.95	95.75	0.94	0.93	0.96	96.25	0.95	0.93
CNN	0.94	95.8	0.95	0.93	0.96	97.01	0.96	0.95
Bi-GRU	0.95	96.7	0.96	0.94	0.97	97.85	0.972	0.94
Proposed method	0.97	98.6	0.98	0.96	0.98	98.95	0.98	0.96

To further evaluate the performance of the DL models, we computed the receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values. The ROC curve illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR) across varying thresholds. A higher AUC value indicates better discriminatory ability. Figure 7 displays the ROC curves for the proposed method and baseline models. The proposed method achieved the highest AUC value of 0.980, demonstrating superior performance in distinguishing offensive and non-offensive tweets. The Bi-GRU and CNN models also performed well, with AUC values of 0.965 and 0.955, respectively. The transformer block achieved an AUC of 0.945. These results highlight the robustness and reliability of the proposed multichannel DL approach.

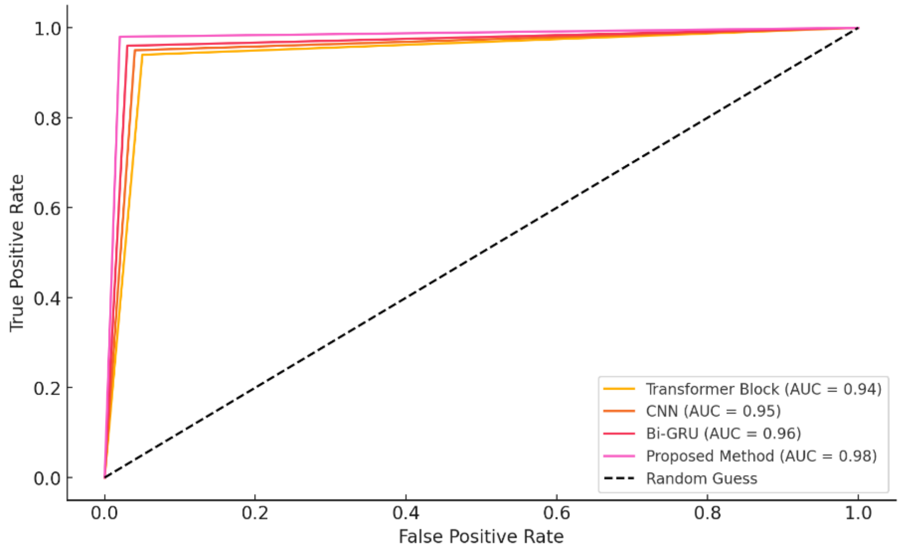


Figure 7. ROC curve and AUC analysis for the proposed multichannel method and baseline DL models

5. CONCLUSIONS

In summary, the prevalence of online social networks has revolutionized communication but also introduced challenges like cyberbullying. Using a combination of three DL architectures: BiGRU, transformer blocks, and CNN models, this work proposes a novel method to detect cyberbullying. The method effectively categorizes concise communications, achieving around 98.95% accuracy using 75% training and 25% testing data split across three cyberbullying datasets. Future work includes exploring multimodal integration, fine-tuning architectural parameters, developing real-time detection, assessing generalization across platforms, and improving model explainability for safer online environments. Possible future work may include investigating multimodal integration, optimizing architectural parameters through fine-tuning, developing real-time detection capabilities, evaluating generalization across different platforms, and improving model explainability and interpretability. The objective of these directives, by promoting the creation of safer and more secure online environments, is to advance the area of cyberbullying detection. Further study will concentrate on enhancing the adaptability of the model to changing online behaviors and upcoming platforms, guaranteeing a proactive strategy for preventing and mitigating cyberbullying. Additionally, the large language models (LLMs) can be utilized using the proposed similarity measure. They can also perform fine-tuning for LLMs through exploring the semantic similarity using sentence-BERT (SBERT), XLNet, and RoBERTa.

ACKNOWLEDGMENT

The authors would like to thank the Scientific Research Dean's Office of Northern Border University, Arar, Saudi Arabia.

FUNDING INFORMATION

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2025-1437-01.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Wael M. S. Yafooz	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Abdulsamad Ebrahim Yahya	✓			✓		✓	✓			✓				✓
Abdullah Alsaeedi	✓				✓				✓	✓				
Reyadh Alluhaibi		✓				✓			✓		✓			
Faisal Jamil		✓		✓	✓				✓					
Mahmoud Salaheldin Elsayed		✓				✓				✓			✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [WMSY], upon reasonable request.

REFERENCES

[1] M. U. S. Khan, A. Abbas, A. Rehman, and R. Nawaz, "HateClassify: a service framework for hate speech identification on social media," *IEEE Internet Computing*, vol. 25, no. 1, pp. 40–49, Jan. 2021, doi: 10.1109/MIC.2020.3037034.

[2] W. M. S. Yafooz, A. Al-Dhaqm, and A. Alsaeedi, "Detecting kids cyberbullying using transfer learning approach: transformer fine-tuning models," in *Kids Cybersecurity Using Computational Intelligence Techniques*, Cham: Springer, 2023, pp. 255–267, doi: 10.1007/978-3-031-21199-7_18.

[3] T. Wullach, A. Adler, and E. Minkov, "Towards hate speech detection at large via deep generative modeling," *IEEE Internet Computing*, vol. 25, no. 2, pp. 48–57, Mar. 2021, doi: 10.1109/MIC.2020.3033161.

[4] A. Razaque, F. Amsaad, D. Halder, M. Baza, A. Aboshgifa, and S. Bhatia, "Analysis of sentimental behaviour over social data using machine learning algorithms," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12798 LNAI, 2021, pp. 396–412, doi: 10.1007/978-3-030-79457-6_34.

[5] N. S. Ansary, "Cyberbullying: concepts, theories, and correlates informing evidence-based best practices for prevention," *Aggression and Violent Behavior*, vol. 50, Jan. 2020, doi: 10.1016/j.avb.2019.101343.

[6] H. Rosa *et al.*, "Automatic cyberbullying detection: a systematic review," *Computers in Human Behavior*, vol. 93, pp. 333–345, Apr. 2019, doi: 10.1016/j.chb.2018.12.021.

[7] L. Cheng, Y. N. Silva, D. Hall, and H. Liu, "Session-based cyberbullying detection: problems and challenges," *IEEE Internet Computing*, vol. 25, no. 2, pp. 66–72, Mar. 2021, doi: 10.1109/MIC.2020.3032930.

[8] A. Razaque *et al.*, "Avoidance of cybersecurity threats with the deployment of a web-based blockchain-enabled cybersecurity awareness system," *Applied Sciences*, vol. 11, no. 17, p. 7880, Aug. 2021, doi: 10.3390/app11177880.

[9] D. Chatzakou *et al.*, "Detecting cyberbullying and cyberaggression in social media," *ACM Transactions on the Web*, vol. 13, no. 3, pp. 1–51, Aug. 2019, doi: 10.1145/3343484.

[10] M. Arif, "A systematic review of machine learning algorithms in cyberbullying detection: future directions and challenges," *Journal of Information Security and Cybercrimes Research*, vol. 4, no. 1, pp. 01–26, Jun. 2021, doi: 10.26735/gbtv9013.

[11] S. S. Jacob and R. Vijayakumar, "Sentimental analysis over twitter data using clustering based machine learning algorithm," *Journal of Ambient Intelligence and Humanized Computing*, Jan. 2021, doi: 10.1007/s12652-020-02771-9.

[12] S. Sadiq, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Aggression detection through deep neural model on Twitter," *Future Generation Computer Systems*, vol. 114, pp. 120–129, Jan. 2021, doi: 10.1016/j.future.2020.07.050.

[13] O. Gencoglu, "Cyberbullying detection with fairness constraints," *IEEE Internet Computing*, vol. 25, no. 1, pp. 20–29, Jan. 2021, doi: 10.1109/MIC.2020.3032461.

[14] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [Review Article]," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, Aug. 2018, doi: 10.1109/MCI.2018.2840738.

[15] Y. Xiao and K. Cho, "Efficient character-level document classification by combining convolution and recurrent layers," *arXiv*, 2016.

[16] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*, vol. 1, pp. 1556–1566, 2015, doi: 10.3115/v1/p15-1150.

[17] B. Haidar, M. Chamoun, and F. Yamout, "Cyberbullying detection: a survey on multilingual techniques," in *Proceedings - UKSim-AMSS 2016: 10th European Modelling Symposium on Computer Modelling and Simulation*, Nov. 2017, pp. 165–171, doi: 10.1109/EMS.2016.037.

[18] S. Salawu, Y. He, and J. Lumsden, "Approaches to automated detection of cyberbullying: a survey," *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 3–24, Jan. 2020, doi: 10.1109/TAFFC.2017.2761757.

[19] M. A. Al-Garadi *et al.*, "Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges," *IEEE Access*, vol. 7, pp. 70701–70718, 2019, doi: 10.1109/ACCESS.2019.2918354.

[20] F. Elsafoury, S. Katsigiannis, Z. Pervez, and N. Ramzan, "When the timeline meets the pipeline: a survey on automated cyberbullying detection," *IEEE Access*, vol. 9, pp. 103541–103563, 2021, doi: 10.1109/ACCESS.2021.3098979.

[21] S. Kim, A. Razi, G. Stringhini, P. J. Wisniewski, and M. De Choudhury, "A human-centered systematic literature review of cyberbullying detection algorithms," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW2, pp. 1–34, Oct. 2021, doi: 10.1145/3476066.

[22] R. Alhejaili, A. Alsaeedi, and W. M. S. Yafooz, "Detecting hate speech in arabic tweets during COVID-19 using machine learning approaches," in *Lecture Notes in Networks and Systems*, vol. 479, 2023, pp. 467–475, doi: 10.1007/978-981-19-3148-2_39.

[23] L. M. Al-Harigy, H. A. Al-Nuaim, N. Moradpoor, and Z. Tan, "Building towards automated cyberbullying detection: a comparative analysis," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–20, Jun. 2022, doi: 10.1155/2022/4794227.

[24] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, vol. 11, no. 1, pp. 512–515, May 2017, doi: 10.1609/icwsm.v11i1.14955.

[25] Z. Waseem and D. Hovy, "Hateful symbols or hateful people? predictive features for hate speech detection on twitter," in *HLT-NAACL 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Student Research Workshop*, 2016, pp. 88–93, doi: 10.18653/v1/n16-2013.

[26] DataTurks, "Tweets dataset for detection of cyber-troll," *Kaggle*. 2018. Accessed: Jan. 01, 2025. [Online]. Available: https://www.kaggle.com/dataturks/dataset-for-detection-of-cybertrolls

[27] S. Petrovic, M. Osborne, and V. Lavrenko, "The edinburgh Twitter corpus," *Computational Linguistics*, no. June, pp. 25–26, 2010.

[28] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 1, pp. 163–173, Jan. 2012, doi: 10.1002/asi.21662.

[29] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2005, pp. 347–354, doi: 10.3115/1220575.1220619.

[30] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," *CEUR Workshop Proceedings*, vol. 718, pp. 93–98, 2011.

[31] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2004, pp. 168–177, doi: 10.1145/1014052.1014073.

[32] S. M. Mohammad, S. Kiritchenko, and X. Zhu, “NRC-Canada: building the state-of-the-art in sentiment analysis of tweets,” *SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, vol. 2, pp. 321–327, 2013.

[33] F. B.-Marquez, E. Frank, S. M. Mohammad, and B. Pfahringer, “Determining Word-Emotion Associations from Tweets by multi-label classification,” in *Proceedings - 2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016*, Oct. 2017, pp. 536–539, doi: 10.1109/WI.2016.0091.

[34] S. Kiritchenko, X. Zhu, and S. M. Mohammad, “Sentiment analysis of short informal texts,” *Journal of Artificial Intelligence Research*, vol. 50, pp. 723–762, Aug. 2014, doi: 10.1613/jair.4272.

[35] S. Baccianella, A. Esuli, and F. Sebastiani, “SENTIWORDNET 3.0: an enhanced lexical resource for sentiment analysis and opinion mining,” *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, pp. 2200–2204, 2010.

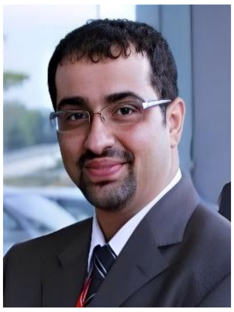
[36] S. M. Mohammad and P. D. Turney, “Crowdsourcing a Word-Emotion Association Lexicon,” *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, Aug. 2013, doi: 10.1111/j.1467-8640.2012.00460.x.

[37] A. E. Yahya, W. M. S. Yafooz, and A. Gharbi, “Mapping graduate skills to market demands: a holistic examination of curriculum development and employment trends,” *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 14793–14800, 2024, doi: 10.48084/etasr.7454.

[38] W. M. S. Yafooz, A. Alsaeedi, R. Alluhaibi, and A. H. Mohamed Emara, “Enhancing multi-class web video categorization model using machine and deep learning approaches,” *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 3176–3191, Jun. 2022, doi: 10.11591/ijece.v12i3.pp3176-3191.

[39] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, “Prediction of diabetes disease using machine learning algorithms,” *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, pp. 284–290, 2022, doi: 10.11591/ijai.v11.i1.pp284-290.

BIOGRAPHIES OF AUTHORS



Wael M. S. Yafooz is a Professor of Artificial Intelligence in the Department of Computer Science, Taibah University, Saudi Arabia. He received his bachelor’s degree in the area of Computer Science from Egypt in 2002, while his Master of Science in Computer Science was received from the University of MARA Technology (UiTM)-Malaysia 2010, additionally, his Ph.D. in Computer Science which was received in 2014 was also obtained from UiTM. He is an IEEE Senior Member and has obtained the Fellow of the Higher Education Academy (FHEA) recognition. He was awarded many Gold and Silver Medals for his contribution to local and international expos of innovation and invention in the area of computer science. Besides, he was awarded the Excellent Research Award from UiTM. He served as a member of various committees in many international conferences. Additionally, he chaired IEEE international conferences in Malaysia and China. Furthermore, He is a volunteer reviewer with different peer-review journals. Moreover, he supervised a number of students at master and Ph.D. levels. Additionally. He was invited as a speaker in many international conferences held in Bangladesh, Thailand, India, China and Russia. His research interests include data mining, machine learning, deep learning, natural language processing, social network analytics, computer vision, and data management. He can be contacted at email: waelmohammed@hotmail.com or wyafooz@taibahu.edu.sa.

Abdulsamad Ebrahim Yahya is an Assistant Professor in Department of Information technology, College of Computing and Information Technology, Northern border University, Rafha, Saudi Arabia. Dr. Yahya received his Ph.D. in Computer Science from Faculty of Information Science and Technology, National University of Malaysia (UKM), Malaysia, 2013 and Higher diploma engineering in Computer Engineering from Higher Institute of Mechanical and Electrical Engineering-Sofia (Republic of Bulgaria), 1994. Dr. Yahya has published several conferences and journal papers. His research area covers artificial intelligent, computer vision, pattern recognition, image processing, image analysis, and biometrics. He can be contacted at email: abdulsamad.qasem@nbu.edu.sa.

Abdullah Alsaeedi received the B.Sc. degree in computer science from the College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia, in 2008, M.Sc. degree in advanced software engineering, The University of Sheffield, Department of Computer Science, Sheffield, UK, in 2011, and the Ph.D. degree in computer science from the University of Sheffield, UK, in 2016. He is currently an Associate Professor at the Department of Computer Science, Taibah University, and Madinah, Saudi Arabia. His research interests include: software engineering, software model inference, grammar inference, and machine learning. He can be contacted at email: aasaeedi@taibahu.edu.sa.

Reyadh Alluhaibi received the B.E degree from Taibah University in 2005. He received M.Sc. degree from Tulsa University in 2009. After working as a lecturer (from 2009 to 2012) in the Deparment of Computer Science, Taibah University. He received the Ph.D. degree from Manchester University in 2017. After working as an Assistant Professor (from 2017) in the Deparment of Computer Science, Taibah University. Currently, he is an Associate Professor in the Department of Artificial Intelligence and Data Science and currently serves as the Dean of Admissions and Registration at Taibah University, Saudi Arabia. He received his B.E. from Taibah University (2005), M.Sc. from the University of Tulsa (2009), and Ph.D. in Artificial Intelligence from the University of Manchester (2017). His research interests include: machine learning, natural language processing, computational semantics, knowledge representation, and temporal logics. He can be contacted at email: rluhaibi@taibahu.edu.sa.

Faisal Jamil currently serves as a lecturer in Computer Science at the University of Ulster, UK. Prior to his position in the UK, he worked as a post-doctoral research fellow at the Department of ICT and Natural Sciences, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Norway, from 2022 to 2023. In recognition of his research excellence, Dr. Faisal received the Presidential Award for Research Excellence in 2021 from President Jeju National University, South Korea. Previously, Dr. Faisal held affiliations with Gachon University as a Postdoctoral Fellow, Jeju National University as a Senior Researcher, and the Capital University of Science and Technology as a Lecturer. He holds a Ph.D. in Information Technology and his research interests encompass applied artificial intelligence (AI), digital health, machine learning, data science, digital twins' application, autonomous navigation, and IoT and blockchain. Dr. Faisal has contributed to the academic field with over 52 publications in peer-reviewed international journals. He can be contacted at email: f.Jamil@hud.ac.uk.

Mahmoud Salaheldin Elsayed received the B.S. degree in Mathematics and Computer Sciences and the Ph.D. degree in Computer Sciences from Menoufia University, Faculty of Science, Menoufia, Egypt. From 2006 to 2008, he was with the Teachers College, Arar, Saudi Arabia. He is Lecturer of Computer Sciences with Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia from 2008 until present. He has a fair number of publications in international journals and conference, which focus on deadlock control for single and multi-unit resource systems. His research interests include: Petri net theory and applications, supervisory control of discrete event systems, and robust supervisory control of flexible manufacturing systems. He received the Excellent Paper Award from the International Institute of Engineers and Researchers (IIER) in 2016, and published a book entitled Deadlock Analysis and Resolution in Distributed Systems in the LAP LAMBERT Academic Publishing, Germany in 2017. He can be contacted at email: mahmoud.al-sayed@nbu.edu.sa.