e-ISSN: 2614-8404
p-ISSN:2776-3234

Detection of Hate-Speech Tweets Based on Deep Learning: A Review

Ara Zozan Miran[1*)], Adnan Mohsin Abdulazeez[2],
[1] Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, Akre University for Applied Science- Technical College of Informatics- Akre- Department of Information Technology.
[2] Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
Email: [1]ara.miran@auas.edu.krd, [2] adnan.mohsin@dpu.edu.krd

Abstract* – Cybercrime, cyberbullying, and hate speech have all increased in conjunction with the use of the internet and social media. The scope of hate speech knows no bounds or organizational or individual boundaries. This disorder affects many people in diverse ways. It can be harsh, offensive, or discriminating depending on the target's gender, race, political opinions, religious intolerance, nationality, human color, disability, ethnicity, sexual orientation, or status as an immigrant. Authorities and academics are investigating new methods for identifying hate speech on social media platforms like Facebook and Twitter. This study adds to the ongoing discussion about creating safer digital spaces while balancing limiting hate speech and protecting freedom of speech. Partnerships between researchers, platform developers, and communities are crucial in creating efficient and ethical content moderation systems on Twitter and other social media sites. For this reason, multiple methodologies, models, and algorithms are employed. This study presents a thorough analysis of hate speech in numerous research publications. Each article has been thoroughly examined, including evaluating the algorithms or methodologies used, databases, classification techniques, and the findings achieved. In addition, comprehensive discussions were held on all the examined papers, explicitly focusing on consuming deep learning techniques to detect hate speech.

Keywords – Twitter, Hate Speech, Toxic, Cyberbullying, Deep Learning.

I. INTRODUCTION

Online social media enables the spread of humanities to be associated. However, one disadvantage of these social media is the ability to publish and propagate malicious and harmful content [1]. Hate speech seeks to promote violence and incite hostility against individuals or groups based on characteristics—such as sexual orientation/gender identity, religion, disability, gender, age, veteran status, or age, gender, or disability. The increasing number of social media platforms has caused matters to be excessive [2]. Not all substances are pertinent; some may cause damage to individuals, which is a disgraceful indictment when those who use the media to spread hatred [3]. Hate speech affects everyone regardless of age, which might be too young, an adult, or an older adult [4]. Individuals use a lot of social media platforms daily to express their thoughts, emotions, and progress. Usually, the comments are more hate speech than positive comments, which leads to different kinds of problems; for that reason, it is essential to select a method to detect the words, analyze them, and then show the result of accuracy [5]. Deep learning utilizes networks that have multiple layers [6]. This allows them to understand patterns and representations in data by learning from examples [7]. The training process involves providing these networks with amounts of labeled data, enabling the system to adjust its parameters until it can accurately predict or classify information [8]. This ability makes deep learning especially effective for tasks that involve decision-making and pattern recognition. Deep understanding encompasses such architectures tailored for specific tasks, each designed to address the unique challenges of different data types [9]. Artificial neural networks, inspired by biological neurons, serve as the foundational components of deep learning, replicating the core elements of human intellect [10]. These networks, which have several layers of nodes, have led to many architectures, each specialized for distinct tasks and applications. Convolutional Neural Networks (CNN) excel in analyzing videos and photos, Recurrent Neural Networks (RNN) are adept at addressing issues related to sequential data, and Generative Adversarial Networks (GAN) are at the forefront of generating synthetic data that closely mimics accurate data. [11], [12]. Hate speech detection in social media Twitter is detecting and defining all the cyberbullying an individual can receive on their regular posts [13]. The hate speech might be on racism, sexuality, child abuse, Politicians, and many others. Detecting hate speech is difficult because its nature and context-specific attributes frequently characterize it [14],[15]. Conventional rule-based or keyword-based methods may face problems in accurately identifying the subtleties of hate speech, as hate speech can take on different forms, such as indirect language, changing slang, or statements that depend on the context [16]. Hate speech detection utilizes deep learning, specifically natural language processing algorithms [17]. This paper aims to review other works on detecting hate speech in social media using different types of deep learning methods. The rest of this paper is organized to describe the background theory of deep learning and the algorithms to detect hate speech on Twitter platforms. After that, in section 2, different research on using CNN and GRU, CNN, and LSTM to detect hate speech on social media are described and compared in Table 1 and Table 2. Finally, section 3 describes a discussion session, a conclusion and limitations, and a list of references are shown.

II. RESEARCH METHODOLOGY

Detecting hate speech in social media requires datasets and an algorithm to learn and extract the comments to be analyzed [18]. For that reason, researchers are using deep learning algorithms to detect hate speech. Deep Learning, a subset of machine learning, aims to replicate the brain's function for analyzing and processing data [15]. At its core, Learning utilizes networks of intricate structures made up

e-ISSN: 2614-8404
p-ISSN:2776-3234

of interconnected nodes to model and interpret complex patterns within large datasets. What makes deep learning understandable is its ability to learn representations of data automatically [19], enabling it to identify features and relationships that may pose challenges for conventional machine learning methods [20]. This technology has shown achievements in file recognition, speech processing, natural language understanding, and autonomous systems [21], [22]. The continuous progress in deep learning algorithms alone and the availability of computing resources have paved the way for advancements in artificial intelligence[23]. These advancements shape technology and impact domains like healthcare, finance, and self-driving vehicles. This technology has shown achievements in file recognition, speech processing, natural language understanding, and autonomous systems[24] [20]. The continuous progress in deep learning algorithms alone and the availability of computing resources have paved the way for advancements in artificial intelligence [25],[26]. These advancements shape technology and impact domains like healthcare, finance, and self-driving vehicles. The Advances in learning models, such as (CNNs) for image processing and (RNNs) for sequential data, have resulted in substantial changes in domains such as machine vision, natural language processing, and recognition of words[8]. Deep Learning is executed by neural networks, which consist of multiple layers, a concept that is not novel [27], [28]. Nevertheless, its popularity has surged recently, driven mainly by three factors: Firstly, there has been a significant enhancement in processing capabilities, such as video cards and graphical processors. Secondly, the availability of affordable computer hardware has played a crucial role. Lastly, recent advancements and breakthroughs in Deep Learning research have also contributed to its rise. Deep learning algorithms can be categorized into three groupings based on whether the algorithms are trained to produce achievable results. The subgroups are unsupervised, supervised, and hybrid [27]. The reliability of each deep learning algorithm is demonstrated in Figure. 1.

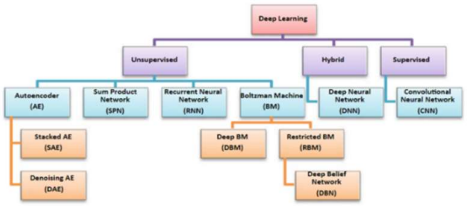

Fig 1. Classification of deep learning techniques, https://www.researchgate.net

A. *Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) for Detecting Hate Speech:*

Due to its potentially devastating effects on society, identifying instances of hate speech in digital communications is a significant difficulty [29]. One viable strategy for this problem is to use cutting-edge technologies like (CNNs). Combining (CNNs) and (GRUs) offers a powerful technique for detecting hate speech, harnessing the benefits of both architectures [14]. While CNNs are great at comprehending short-term dependencies in sequences, (GRUs) excel at understanding long-term dependencies [30]. In this combined method, textual information is tokenized and processed, then supplied into an embedding layer, as shown in Figure 2 [31]. It covers identifying hate speech on social media, particularly Twitter and provides a convolution-GRU-based deep neural network solution [32]. The authors address hate speech identification features such as a bag of words, word and character n-grams, sentiment analysis, linguistic resources, and standard and deep learning methods [33]. The authors also noted a need for comparable evaluation and limited public hate speech datasets.

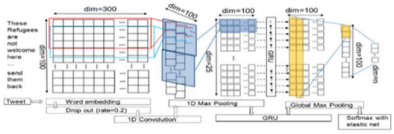

Fig. 2 CNN+GRU Architecture [31]

The CNN's convolutional layers then process the embedded sequences to pull out abstract patterns and features at various levels of granularity [34]. The (GRU) then uses these traits to learn to record sequential and contextual information, drawing connections between previously unrelated parts of the text [22]. The (GRU)'s gated structure allows for the storage and updating of data, which aids in the nuanced understanding of the text's context and the detection of hate speech [35]. Training a mixed model requires iteratively adjusting parameters to achieve a target loss function minimization. To better detect hate speech, the model combines (CNNs) for feature extraction with GRU for sequence modeling [10]. The model's performance and robustness in identifying hate speech across online content can be improved by fine-tuning, optimizing, and studying other architectures and hyperparameters. There are negative impacts on humans' mental and physical health as a direct result of the rapid growth of social media mediums and the constant use by their users. To guard against and avoid any mishaps that might take place behind the scenes on social media platforms, hate speech has become one of the most hazardous subjects to discuss. Researchers turned to deep learning strategies, which will be covered in more detail in the following paragraphs, to identify instances of hate speech on social media sites like Twitter [36], [37] proposed method utilizes a deep learning ensemble approach that combines multiple sub-models to improve classification performance. The paper also mentions the use of a publicly available embedding model. As for the number of datasets used, the article does not provide a specific number. However, the authors mention using the SemEval dataset to compare findings and a split set with an 85/15 ratio for testing purposes. Regarding the results, the paper reports a 5-point improvement in F-measure compared to previous work. The ensemble approach also showed a reduction in variance compared to individual models. [37]discusses a new method for detecting hate speech on Twitter using a Convolution-(GRU)-based deep neural network. The proposed method outperforms previous methods on 6 out of 7 datasets by between 1 and 13% in F1. The comparison was made against several baselines and state-of-the-art on the most extensive publicly available Twitter datasets. The paper also

e-ISSN: 2614-8404
p-ISSN:2776-3234

discusses the implementation, parameter tuning, and evaluation metrics used in the study. [38]systematically analyze hate speech in 9 languages and offer a low-resource detection approach. They test CNN-GRU, BERT, and LASER embedding with logistic regression to detect hate speech in different languages under different scenarios. The authors also present a portfolio of effective models for each language based on data availability. The research sheds light on the difficulties of detecting hate speech in languages other than English and suggests a low-resource detection approach. In [39] they used CNN and GRU algorithms to detect hate speech. CNN is better for local features, and GRU is better for long-range context dependencies. LSTM, another RNN, is slower and more likely to overfit small datasets than GRU. Both models have pros and cons, and hyperparameter adjustment affects performance. [40] suggests utilizing deep neural networks to detect internet hate speech. According to the authors, automatic identification is needed due to digital communication and online hostility. They propose multiclass classification using fastText, BERT embeddings, SVM, and deep neural network classifiers. They find encouraging results using their method on numerous datasets. [41] Combining GRU, ELMo, BERT, and CNN data improves classification performance. The proposed approach has good accuracy and F1 score, making it a feasible option for social media hate speech. However, the approach's success may depend on the training data's quality and representativeness and the classifiers and fusion methods used. [9]discusses a model for detecting hate speech on Twitter using CNN and character-level representation. The authors present their approach to building and cleaning datasets, as well as the architecture of their model. They also discuss related work in sentiment analysis and hate speech detection. The results of their experiments were presented and discussed, with the model achieving satisfactory accuracy compared to the state-of-the-art. [42] highlight various hate speech detection research using these neural network topologies. Zimmerman et al. represented 50 tokens using CNN parameters, while Founta et al. offered a two-layer RNN model using GRU and metadata on people, networks, and content. Some hate speech detection investigations have used hybrid CNN and RNN models, which perform better than either architecture alone. The research thoroughly analyzes CNN and GRU in hate speech detection and shows their potential to improve model accuracy. [43] discusses a study that classified Bengali Facebook comments as Hate Speech, Communal Attacks, Incitement, Religious Hatred, Political Comments, and Religious Comments using (GRU). The study presented an annotated Bengali corpus of six-class comments. The experiment indicates that a GRU-based classification model can achieve 70.10% accuracy. The article found that the BiLSTM performed best with a weighted classification F1 score of 91%. In[36], to identify harmful remarks in virtual communities and social media platforms. The algorithm employs pre-trained word embeddings and many channels to forecast multilabel toxicity indicators precisely. The report additionally examines pertinent literature, outlines the structure of the suggested model, and presents the findings of its evaluation. Moreover, in [37],the research paper discussed the use of deep neural network (DNN) architectures, including (CNN), (GRU), and Universal Language Model Fine-tuning (ULMFiT), for the classification of hate speech in tweets. It highlights the advantages of using DNN models for hate speech detection and emphasizes the superior performance of the ULMFiT model over other architectures. The ULMFiT model, based on the three-layered Average-SGD Weight-Dropped Long Short-Term Memory (AWD-LSTM) architecture, demonstrates substantial improvements in accuracy and F1 score. After that, in [46]. It outlines the method used in the study, which involves transforming speech signals from the time domain to the frequency domain, extracting features using mel-frequency cepstral coefficients, and using a CNN-GRU model to detect and classify dysarthria patients and healthy people. The article also mentions the prevalence of dysarthria in patients with neurological diseases and the potential applications of this technology in healthcare. [47] Investigates the frequency of hate speech on Twitter amidst the COVID-19 outbreak. The study uses machine learning algorithms to discern and examine the data to ascertain patterns and trends within hate speech. The results indicate that there was a significant presence of hate speech about COVID-19 on Twitter during the pandemic, with specific demographics being disproportionately targeted. The authors propose that their discoveries can be employed to counteract hate speech and foster inclusiveness on social media sites. The researcher in [29]Worked on A comprehensive analysis of scholarly research on the identification of hate speech in social media, particularly on Twitter, using the utilization of neural networks. The review comprises a subset of 20 studies from a pool of 565 distinct works, chosen according to specific inclusion and exclusion criteria and the study's primary and secondary inquiries. The review examines the models, algorithms, and methodologies employed in these investigations and the datasets, languages, and nations implicated. The review reveals that the predominant models used are based on Convolutional Neural Networks (CNNs). While these models offer certain advantages over recent models in detecting hate speech, they also exhibit limitations and deficiencies. These include challenges in automated hate speech detection, compatibility with specific languages only, and ambiguity in speech classification when applied to datasets in different languages. The review presents a thorough overview of the experimental particulars and findings of the incorporated investigations, encompassing F1 score, recall, precision, and accuracy. In [48], The authors address the issue of vocabulary mismatch in tweets by using feature expansion to build a corpus of similarity. The study uses four deep learning methods: Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and a combination of the CNN-GRU and GRU-CNN classifications with Boolean representation as feature extraction. The experiment was carried out to find the best performance by comparing the value of the accuracy. The accuracy results were obtained from the average of the five tests in each scenario.

e-ISSN: 2614-8404
p-ISSN:2776-3234

Table 1. Overview of the literature on detecting hate speech Tweets based on CNN and GRU.

e-ISSN: 2614-8404
p-ISSN:2776-3234

Ref	Year	Datasets	Based Model	Accuracy	Advantages	Limitations
[31]	2018	Seven datasets	CNN+GRU	range from 0.72% to 0.94%	Capacity to catch word sequence and order in brief texts, high Twitter hate speech detection accuracy.	The lack of comparison judgments makes it hard to evaluate individual works. The existence of abstract ideas like ' sexism,' 'racism,' or 'hate' is difficult to discern in literary content, thus knowledge of social groupings.
[36]	2018	SemEval dataset	CNN	5%	Shows a significant improvement in F-measure compared to previous work. The authors also provide detailed experimental results and analysis, including confusion matrices, to aid reproducibility and comparison with other methods.	One limitation is using pre-trained word embeddings, which may not capture all relevant information for the specific task—also, there is a lack of explanation for the decisions made by the model, which may be necessary for applications. Additionally, the proposed approach may not be suitable for low-resource languages or domains with limited training data.
[37]	2018	Seven public datasets	CNN+GRU	13%	The proposed method uses only word-based features, which are more straightforward and interpretable than character-based features yet achieve better results.	The proposed method may have limitations and potential issues not discussed in the paper.
[9]	2019	Seven datasets	CNN	0.8893%	It addresses the challenge of multilingual hate speech detection, which is a significant advantage given the diverse nature of social media content. The model was found to be more than satisfactory when compared to the state-of-the-art model, indicating its effectiveness in dealing with the issue of hate speech on the internet.	Potential biases in the dataset, constraints in the model's generalizability, or challenges in real-time implementation.
[39]	2019	Million tweets	CNN+GRU	0.678%	Gives a summary of the essential works in social media hate speech automatic detection for scholars and practitioners.	Focuses on hate speech detection's technological components, such as machine learning techniques and feature engineering, rather than its ethical and societal ramifications.
[40]	2019	24,883 tweets	CNN, SVM, GRU, BERT	11.6%	Examine multiple models and find that BERT fine-tuning performed best, showing that deep neural networks can detect internet hate speech. Their approach is adaptable to additional languages and domains outside Twitter.	It is focused on English-language tweets and says the confusion between hate speech and offensive speech needs further study.

e-ISSN: 2614-8404
p-ISSN:2776-3234

[41]	2020	Two datasets	CNN, BERT, GRU	0.764 to 0.787	It uses machine learning algorithms like ELMo, BERT, and CNN to detect hate speech on social media. Combining classifier results improves classification performance. The proposed approach has good accuracy and F1 score, making it a feasible option for social media hate speech.	Performance may depend on training data quality and representativeness and classifier and fusion algorithms used.
[38]	2020	16 datasets	CNN-GRU, BERT and LASER	0.93%	Employs pre-trained Google News Corpus word embeddings to capture semantic links between words and improve model performance. The model also uses convolutional and recurrent layers to collect local and global text information.	Requirements for massive training data, trouble capturing complicated syntactic patterns, and a potentially confusing architecture.
[44]	2021	223,549	CNN + BiGRU	75%	Achieves state-of-the-art performance on several evaluation metrics, including ROC_AUC score, precision, recall, and F1-score, compared to other existing models. Additionally, the proposed MCBiGRU model is designed to handle multilabel classification, which is more challenging than binary classification.	The dataset used for training and evaluation is limited; there is a lack of study regarding interpretability and extensive analysis of the computational resources needed for training and deployment.
[42]	2022	Numerous hate speech datasets	CNN + GRU	0.88%	Relevance of using current datasets and text context to improve hate speech detection models. The authors also emphasize the necessity to develop characteristics that may be applied to varied datasets and topics of interest to create more generalized and high-performance models.	Hate speech detection research is limited by a lack of unanimity on hate speech, bias in hate speech datasets, poor quality datasets, and the need for more generic model development.
[43]	2022	5,126 comments	BiLSTM, CNN, CNN-LSTM, GRU	70.10%	It gives unique insights into hate speech detection, its impact on society, and deep learning models' capacity to remedy this issue.	Language experts selected and labeled the dataset used to train and test deep learning models, which may not reflect popular opinion.
[46]	2022	One large dataset	CNN + GRU	78.57%	The ability to provide high accuracy.	The models worked on a small sample size, and further research is needed to validate the effectiveness of the CNN-GRU model on a larger scale.
[47]	2022	1180 tweets	CNN	0.615 %	Simple to use, not making assumptions about the data, and achieving high levels of accuracy. The authors conducted an F1-Score of 0.85 using the BERT algorithm, considered high accuracy.	The study was limited to a specific period and may not represent hate speech on Twitter. The study only focused on hate speech related to Asian individuals during the COVID-19 pandemic and did not consider other forms of hate speech or different periods.

e-ISSN: 2614-8404
p-ISSN:2776-3234

[29]	2023	2400 tweets	CNN	94%	Advantages of some of the models in detecting abusive language in non-generalizable (unseen) problems and in terms of their speed and ability to load all the data quickly.	Handling different topical focuses and targets and the ability to load data quickly.
[48]	2023	183,472	CNN, CNN, and GRU	CNN= 87.94% CNN+GRU= 0.8733%	Used deep learning models and feature extraction techniques to achieve high accuracy in hate speech detection. The study also provides a detailed description of the methodology used, which can be helpful for researchers working on similar problems.	The study did not compare the performance of their proposed method with other state-of-the-art methods for hate speech detection. It did not analyze the computational resources required to train and test the models.

B. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) for Detecting Hate Speech:

CNNs allow for fast processing of text data, extracting linguistic features and patterns that can be used to identify hate speech from other types of expression [49], [50]. Embeddings, convolution, pooling, and dense connections are just some of the layers that can train CNN models to recognize the linguistic and contextual cues indicative of hate speech. [51]addresses the difficulties of identifying hate speech on social media and the strategies that have proven successful in doing so. When evaluating the precision of various machine learning and deep learning approaches, the authors find that BiRNN (Bidirectional Recurrent Neural Network) excels. [52]Each time step, the LSTM neural network processes one word embedding from a multilayer perceptron (MLP) with two hidden layers, keeping word order. LSTM neural network output is processed via hyperbolic tangent activation. The number of terms must be specified before training the recurrent neural network. HaterNet detected at most 33 terms after preprocessing the tagged corpus tweets. All tweets under 33 terms have padding rows of 0s. In. After that, [53]benefits from identifying risky comments using LSTM neural networks. The authors cleaned data with NLTK and built their own stop words. The model had great precision, recall, and accuracy. The paper tries to filter hate speech and make social media safer. [54] Presents a comprehensive approach to detecting hate speech on Twitter using traditional machine and deep learning techniques. The authors explore shallow learning algorithms, including logistic regression, random forest, decision tree, naïve Bayes, K-NN, SVM, and deep learning methods such as LSTM, BiLSTM, and CNN. Further examines the difficulties linked to detecting hate speech on social media platforms. It provides insights into potential applications of the proposed approach for mitigating the impact of hate speech on individuals and communities. On the road to realizing their full potential, these models are trained extensively using labeled datasets and then fine-tuned [55]. Aside from helping with automatic identification, using this model for hate speech detection allows for prompt interventions and mitigation methods, making the internet safer for everyone. CNNs and (LSTM) networks are crucial in identifying hate speech in textual data[56]. CNNs detect specific patterns and characteristics in the text using convolutional and pooling layers to extract important linguistic clues about hate speech. Meanwhile, (LSTM) models excel in comprehending the sequential structure of language, effectively capturing contextual interdependencies and enduring associations among words or phrases. By leveraging the capabilities of (CNNs) for extracting features and (LSTMs)[57] ,[58]. This collaboration allows for identifying hate speech content that is nuanced and embedded within its context. In [49],the study focuses on detecting hate speech diffusion on Twitter using graph-based methods. It finds that classification based on the sharing graph yields strong F1 scores for hate speech detection and highlights the vulnerability of existing textual hate speech detection methods to adversarial attacks. The study also considers the effects of automated bots in sharing hate speech content. The implications of the findings are relevant for addressing hate speech and online safety on social media platforms.[7]The DeepHate model utilizes multi-faceted text representations, including semantic, sentiment, and topical information, to improve hate speech detection. By combining pre-trained word embeddings, sentiment analysis, and Latent Dirichlet Allocation (LDA), the model achieves better performance compared to other configurations. Additionally, empirical studies provide insights into the salient features that aid in hate speech detection. [7]discusses the challenges of detecting hate speech in social media, particularly on platforms like Twitter. It explores deep learning approaches, such as various embeddings, to improve the detection of different types of hate speech. The experiments on publicly available datasets showed significant improvements in accuracy and F1-score, offering promising solutions to combat hate speech online. The authors [16] discussed cyberbullying detection using machine learning and deep learning approaches. Cyberbullying's potential dangers are highlighted, as are the benefits of effective detection methods. Methods for categorizing cyberbullying are

e-ISSN: 2614-8404
p-ISSN:2776-3234

investigated in the research. The research examines machine learning and deep learning strategies for cyberbullying detection using two datasets: the Wikipedia Talk Corpus and the Twitter Hate Speech Corpus. The paper also addresses the role that technology plays in making cyberbullying more severe than in-person harassment. Overall, the research gives valuable insights into the detection of cyberbullying and the use of machine learning and deep learning technologies to solve this issue.[60] examines the challenges of detecting hate speech on social networking sites and the research in natural language processing (NLP) and machine learning (ML) to address this issue. The authors used an up-sampling method to balance the data. They implemented deep learning models like Long Short-Term Memory (LSTM) and Bi-directional Long Short-Term Memory (Bi-LSTM) for improved accuracy in detecting hate speech. LSTM was found to have better accuracy, precision, and F1 score, while Bi-LSTM had a higher recall. [61]Explores the problem of harmful language on social media platforms and suggests a technique for identifying it through the utilization of conventional machine learning models, as well as BERT and fastText embedding with deep neural networks. The authors merged the ALONE and HASOC'20 datasets to create a consolidated dataset for their research. The researchers pre-processed data and applied various machine learning and ensemble techniques, including TF-IDF, POS tagging, and trigrams. Among these approaches, LR and XGBoost yielded the most favorable results. In the second case, word embedding techniques such as fastText and BERT were employed to generate embeddings, which were subsequently utilized as inputs for DNN classifiers. The researchers used multiple deep neural network classifiers and found optimal performance was achieved by combining BERT embeddings with a convolutional neural network (CNN). [62] focused on hate speech detection methods using deep and shallow learning techniques. The study provides insights into the detection accuracy, computational efficiency, and practical implications of using pre-trained models and domain generalization. The paper presents a large-scale empirical evaluation of 14 shallow/deep classification-based hate speech detectors, evaluated on three large and publicly available hate speech detection benchmarks. [63] a study on detecting hate speech on Twitter using deep convolutional neural networks. The study uses machine learning-based classifiers such as Logistic Regression, Random Forest, Naive Bayes, Support Vector Machines, Decision Trees, and K-Nearest Neighbors to identify hate speech-related tweets on Twitter. The study also uses word embedding methods such as LSTM and Bi-LSTM models. The results show that the Deep Convolutional Neural Network architecture performs well in detecting hate speech on Twitter. Furthermore, in [54],the authors discussed detecting hate speech in social media using the LSTM algorithm. Results show that the capability to detect hate speech in online text automatically, high levels of accuracy, recall, and F1 score, and the potential to counteract hate speech on social media platforms are all features of this technology. In [55],BERT and Hate Speech Word Embedding with Deep Model detect hate speech in text data. The paper describes feature and classifier approaches, reviews recent investigations, and includes datasets, embedding models, and experiment results. It successfully detects hate speech in English text data. [66]examines fuzzy categorization CNN-LSTM and Random Forest deep learning models. The research discusses Twitter hate speech detection issues and model efficacy. The discovery could be used in pattern recognition, machine learning, and AI beyond Twitter. [67]the research proposes an attentional multi-channel convolutional stacking Bidirectional LSTM network that uses word representation approaches to capture semantic relations at multiple windows. The model is compared to five state-of-the-art and five baseline models on three Twitter benchmark datasets. The presented model outperforms the others in most circumstances, and the absence of channels and attention mechanisms has the most significant influence, as proven in Figure 3. An online social network hate speech detection methodology can be utilized for data-driven cyber security.

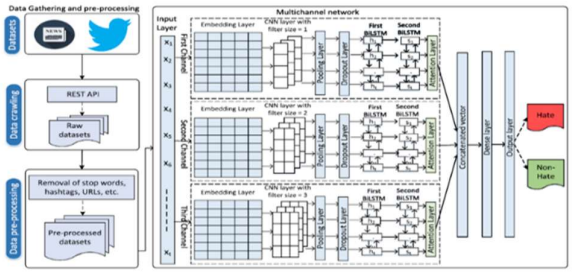

Fig. 1 CNN+LSTM proposed model [67]

e-ISSN: 2614-8404
p-ISSN:2776-3234

Table 2. Overview of the literature on detecting hate speech Tweets based on CNN and LSTM.

e-ISSN: 2614-8404
p-ISSN:2776-3234

Ref	Year	Datasets	Based Model	Accuracy	Advantages	Limitations
[51]	2019	Two datasets	LSTM	0.90%	Compares numerous Twitter hate speech detection techniques utilizing varied data ratios for a more complete evaluation. The article also evaluates machine learning and deep learning models for Twitter hate speech detection.	Only detects hate speech in tweets, not other social media or online platforms.
[52]	2019	16k tweets dataset	LSTM+MLP	0.895% and 0.892%	An integrated hate speech detection method using deep learning, feature engineering, and ensemble learning.	Due to the deep learning model's computational cost, the approach may not scale to massive datasets. The authors note that their method may not detect implicit or coded hate speech.
[53]	2020	Two datasets	LSTM	94.94%	The model classifies a statement as toxic or non-toxic and gives its proportion. High accuracy, precision, and recall scores on test data show that the model can distinguish harmful and non-toxic phrases.	It focuses on English comments exclusively; its performance in other languages is unknown. This may limit the model's use in multilingual online platforms.
[54]	2020	2 Wikipedia Talk Corpus and the Twitter Hate Speech Corpus	CNN+ DBOW, CNN+ DMM	CNN+ DBOW=94% CNN+ DMM=98.20 %	CNN models are computationally efficient and can capture local features, while the DBOW and DMM models are more straightforward and faster to train.	DBOW and DMM models may not capture the order of words in a sentence, which can be important for hate speech classification.
[59]	2020	99,799	LSTM	F1 scores of 0.80% and recall scores of 0.93%	The study does highlight that graph convolutional networks yield robust F1 scores for the imbalanced classification task and high hate speech detection precision. Additionally, the study reproduces existing results showing how adversarial attacks can weaken text detection models for hate speech detection.	Adversarial attacks can easily fool current text-based detection methods, and simple manipulations of messages can impair the detection system's abilities.
[7]	2020	3 Public datasets	RNN+LSTM+Hybrid CNN	Not mentioned	Utilizes multi-faceted textual representations for automatic hate speech detection in social media. Empirical analyses of the Deep Hate model provide insights into the salient features that helped detect hate speech in social media.	Not mentioned any limitations.

e-ISSN: 2614-8404
p-ISSN:2776-3234

[16]	2020	16K	CNN, LSTM, BiLSTM	CNN=90.47 % LSTM=90.55 %	The ability to detect various types of hate speech, even on platforms like Twitter, where contextual information is limited. These approaches significantly improve accuracy and F1 score, providing promising solutions to combat hate speech online.	The quality and representativeness of the training data, the generalization to new and diverse types of hate speech, and the potential biases in the training data.
[60], [61]	2021	51962 tweets	LSTM and BiLSTM	LSTM= 0.9785% BiLSTM= 0.9781%	LSTM excelled in accuracy, precision, and F1 score, whereas Bi-LSTM excelled in recall. The up-sampling technique adopted by the authors to achieve data balance ultimately led to better model results.	Without a large, labeled dataset, the models built using pure NLP concepts can be slower than those made using machine learning or deep learning models.
[61]	2021	25,000 tweets	CNN+BERT LSTM+GRU	13%	Combining two datasets increases the data's diversity and improves the results' generalizability. It also comprehensively analyzes various machine learning and deep learning techniques for hate speech detection on social media platforms.	This research is used only to detect English tweets and hate speech without categorizing them into different types of cyberbullying.
[62]	2022	25K	CNN+ BiLSTM	F1 score of 0.61% and a weighted average F1 score of 0.73 %. Weighted average model= 0.96-0.97%	Offers valuable insights into these methodologies' precision, effectiveness, and applicability. The authors also provide open access to their programs on GitHub, which might be helpful for scholars and practitioners seeking to reproduce or expand upon the work.	The study exclusively detects English-language hate speech; findings may not apply to other languages. The work only tests the models on three datasets; therefore, the results may not represent all hate speech scenarios. The results may be difficult to interpret because the publication needs to provide overall accuracy.
[63]	2023	3 Different datasets	LSTM, BiLSTM, and CNN	CNN=0.892 % LSTM=0.901 % BiLSTM=0.902%	We achieved higher accuracy than traditional machine learning algorithms, with BiLSTM outperforming other deep learning models.	A limited dataset, lack of generalizability, computational complexity, interpretability, and limitations of feature extraction techniques.
[64]	2023	Fixed partitioned datasets were used.	Bi-LSTM	Bi-LSTM=0.72 %	Ability to automatically detect hate speech in online text, high accuracy, recall, and F1 score, and its potential to combat hate speech on social media platforms.	It only addresses hate speech issues with text data and does not include images or videos. It also only uses tweets written in English, which limits its generalizability to other languages.
[65]	2023	Three datasets	CuDNNLSTM+BERT	0.96%	Pre-trained models can be fine-tuned for specific tasks, improving performance. Pre-trained models can also learn from enormous volumes of data to detect language nuances.	Code words or implicit language may prevent these models from detecting hate speech. Hate speech detection is subjective, and people may have various definitions.

e-ISSN: 2614-8404
p-ISSN:2776-3234

[66]	2023	One dataset is divided into four categories	CNN-LSTM	75.35% to 88.69%	More complex and flexible categorization to catch hate speech and other damaging language. Decision boundaries can be more flexible, improving data acquisition, and classifications can account for uncertainty by enabling different confidence levels.	It is computationally intensive and may demand more resources than binary classification.
[67]	2023	Two datasets	BiLSTM	0.93%	On imbalanced datasets, the suggested model outperforms comparative models by 10%. Ablation analysis is used to evaluate model neural network components.	The dataset was restricted to English-language hate speech, ignored external influences, and did not analyze model interpretability.

III. RESULTS AND DISCUSSION

Hate speech detection on social media is a big concern in machine learning [68], with numerous researchers actively addressing hate speech originating from various sources [29]. CNN is considered one of the most effective algorithms for detecting hate speech, as it utilizes several methodologies [69], algorithms, and techniques [70]. Table 1 of this research presents an overview of 15 studies focusing on detecting hate speech using (CNN) and (GRU) and Table 2 gives an overview of 15 studies focusing on detecting hate speech using (CNN) and (LSTM). Moreover, the reviewed papers have been updated and published within the last five years. The CNN-based models achieve an accuracy ranging from 90% and higher in eighteen of the thirty works of literature—each of the previous research papers utilized separate datasets comprising several Twitter comments. The findings indicate that the quantity and nature of datasets have minimal impact on the precision of CNN-based models. In the end, the review articles prove that machine learning models are also used for detecting hate speech texts. Still, the accuracy and reliability of deep learning models such as CNN, LSTM, and GRU are higher among other algorithms. Half the researchers found that models designed for a single dataset may make mistakes. Another gap is that the models cannot detect other languages because they are trained on English tweets only.

IV. CONCLUSION

The application of techniques for "deep learning" to identify hate speech on Twitter is an exciting strategy for reducing the negative impacts of cyberbullying. Since language on social media constantly changes, rule-based approaches could be more effective; however, deep learning models are ideally adapted to this challenge. These models employ natural language processing and deep neural networks to analyze hate speech, and they show promise in picking up on nuanced contextual differences and shifting patterns of intolerance. However, there are still obstacles to overcome, such as the requirement for extensive and varied labeled datasets, the possibility of biased training data, and the persistent development of language and online communication. Fine-tuning algorithms to reduce false positives is also essential because it is difficult to compromise between suppressing hate speech and protecting free speech. Incorporating deep learning models into content moderation systems on platforms like Twitter can make the internet safer as technology and research in this area evolve. Researchers, platform developers, and communities must continue to work together to improve models, combat biases, and cultivate a more welcoming digital space. This research reviews fifteen studies on detecting hate speech on Twitter using deep learning methods. Different models and other datasets were used, and the results show us that CNN is the most accurate, reliable, and easier to detect comments and then classify as offensive, sexual, racist…etc.

REFERENCES

[1] W. Alorainy, P. Burnap, H. Liu, and M. Williams, "The Enemy Among Us: Detecting Hate Speech with Threats Based 'Othering' Language Embeddings," 2018, [Online]. Available: http://arxiv.org/abs/1801.07495

[2] N. A. Kako and A. M. Abdulazeez, "Peripapillary Atrophy Segmentation and Classification Methodologies for Glaucoma Image Detection: A Review," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 18, no. 11, pp. 1140–1159, Mar. 2022, doi: 10.2174/1573405618666220308112732.

[3] L. Ketsbaia, B. Issac, and X. Chen, "Detection of hate tweets using machine learning and deep learning," *Proceedings - 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2020*, pp. 751–758, 2020, doi: 10.1109/TrustCom50675.2020.00103.

[4] G. (Computer scientist) Wang, IEEE Computer Society, and Institute of Electrical and Electronics Engineers., *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications : proceedings :*

e-ISSN: 2614-8404
p-ISSN:2776-3234

29 December 2020-1 January 2021, Guangzhou, China.

[5] O. S. Kareem, A. M. Abdulazee, and D. Q. Zeebaree, "Skin Lesions Classification Using Deep Learning Techniques: Review," *Asian Journal of Research in Computer Science*, pp. 1–22, May 2021, doi: 10.9734/ajrcos/2021/v9i130210.

[6] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "2D Facial Images Attractiveness Assessment Based on Transfer Learning of Deep Convolutional Neural Networks," in *ICOASE 2022 - 4th International Conference on Advanced Science and Engineering*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 13–18. doi: 10.1109/ICOASE56293.2022.10075585.

[7] R. Cao, R. K. W. Lee, and T. A. Hoang, "DeepHate: Hate Speech Detection via Multi-Faceted Text Representations," in *WebSci 2020 - Proceedings of the 12th ACM Conference on Web Science*, Association for Computing Machinery, Inc, Jul. 2020, pp. 11–20. doi: 10.1145/3394231.3397890.

[8] K. Ismael Taher and A. Mohsin Abdulazeez, "Deep Learning Convolutional Neural Network for Speech Recognition: A Review," 2021, doi: 10.5281/zenodo.4475361.

[9] A. Elouali, Z. Elberrichi, and N. Elouali, "Hate speech detection on multilingual twitter using convolutional neural networks," *Revue d'Intelligence Artificielle*, vol. 34, no. 1, pp. 81–88, 2020, doi: 10.18280/ria.340111.

[10] M. Jameel Barwary and A. Mohsin Abdulazeez, "Impact of Deep Learning on Transfer Learning : A Review IJSB Literature Review," 2021, doi: 10.5281/zenodo.4559668.

[11] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[12] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "FIAC-Net: Facial Image Attractiveness Classification Based on Light Deep Convolutional Neural Network," in *2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCSEA54677.2022.9936582.

[13] A. Chaudhari, A. Parseja, and A. Patyal, "CNN based hate-o-meter: A hate speech detecting tool," in *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 940–944. doi: 10.1109/ICSSIT48917.2020.9214247.

[14] R. J. Hassan and A. Mohsin Abdulazeez, "Deep Learning Convolutional Neural Network for Face Recognition: A Review Literature Review," 2021, doi: 10.5281/zenodo.4471013.

[15] J. N. Saeed and A. M. Abdulazeez, "Facial Beauty Prediction and Analysis based on Deep Convolutional Neural Network: A Review," *Journal of Soft Computing and Data Mining*, vol. 02, no. 01, Apr. 2021, doi: 10.30880/jscdm.2021.02.01.001.

[16] P. Kapil', A. Ekbal', and D. Das, "Investigating Deep Learning Approaches for Hate Speech Detection in Social Media."

[17] H. T. Sadeeq and A. M. Abdulazeez, "Metaheuristics: A Review of Algorithms," *International journal of online and biomedical engineering*, vol. 19, no. 9. International Association of Online Engineering, pp. 142–164, 2023. doi: 10.3991/ijoe.v19i09.39683.

[18] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles," *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/app13179728.

[19] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[20] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[21] H. T. Sadeeq and A. M. Abdulazeez, "Giant Trevally Optimizer (GTO): A Novel Metaheuristic Algorithm for Global Optimization and Challenging Engineering Problems," *IEEE Access*, vol. 10, pp. 121615–121640, 2022, doi: 10.1109/ACCESS.2022.3223388.

[22] C. H. Salh and A. M. Ali, "Breast cancer recognition based on performance evaluation of machine learning algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 2, pp. 980–989, Aug. 2022, doi: 10.11591/ijeecs.v27.i2.pp980-989.

[23] H. Saeed Yahia and A. Mohsin Abdulazeez, "Medical Text Classification Based on Convolutional Neural Network: A Review," 2021, doi: 10.5281/zenodo.4483635.

[24] K. Xia, J. Huang, and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020, doi: 10.1109/ACCESS.2020.2982225.

[25] Z. A. Aziz and A. M. Abdulazeez, "Application of Machine Learning Approaches in Intrusion Detection System," *Journal of Soft Computing and Data Mining*, vol. 2, no. 2, Oct. 2021, doi: 10.30880/jscdm.2021.02.02.001.

[26] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Trans Cybern*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020, doi: 10.1109/TCYB.2020.2983860.

[27] E. Benavides, W. Fuertes, S. Sanchez, and M. Sanchez, "Classification of Phishing Attack

e-ISSN: 2614-8404
p-ISSN:2776-3234

Solutions by Employing Deep Learning Techniques: A Systematic Literature Review," in *Smart Innovation, Systems and Technologies*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 51–64. doi: 10.1007/978-981-13-9155-2_5.

[28] H. Chen *et al.*, "A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources," *Agric Water Manag*, vol. 240, Oct. 2020, doi: 10.1016/j.agwat.2020.106303.

[29] A. Z. Miran and H. S. Yahia, "Hate Speech Detection in Social Media (Twitter) Using Neural Network," *Journal of Mobile Multimedia*, vol. 19, no. 3, pp. 765–798, 2023, doi: 10.13052/jmm1550-4646.1936.

[30] C. H. Salh and A. M. Ali, "Comprehensive Study for Breast Cancer Using Deep Learning and Traditional Machine Learning", doi: 10.21271/zjpas.

[31] Z. Zhang, J. Tepper, and D. Robinson, "Detecting hate speech on Twitter using a convolution-GRU based deep neural network," 2018. [Online]. Available: https://www.researchgate.net/publication/323723283

[32] C. H. Salh and A. M. Ali, "Unveiling Breast Tumor Characteristics: A ResNet152V2 and Mask R-CNN Based Approach for Type and Size Recognition in Mammograms," *Traitement du Signal*, vol. 40, no. 5, pp. 1821–1832, Oct. 2023, doi: 10.18280/ts.400504.

[33] S. Modha, P. Majumder, T. Mandl, and C. Mandalia, "Detecting and visualizing hate speech in social media: A cyber Watchdog for surveillance," *Expert Syst Appl*, vol. 161, Dec. 2020, doi: 10.1016/j.eswa.2020.113725.

[34] D. Arya *et al.*, "Transfer Learning-based Road Damage Detection for Multiple Countries," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.13101

[35] H. T. Sadeeq and A. M. Abdulazeez, "Car side impact design optimization problem using giant trevally optimizer," *Structures*, vol. 55, pp. 39–45, Sep. 2023, doi: 10.1016/j.istruc.2023.06.016.

[36] S. Zimmerman, C. Fox, and U. Kruschwitz, "Improving Hate Speech Detection with Deep Learning Ensembles." [Online]. Available: https://www.economist.com/news/europe/21734410-

[37] Z. Zhang, D. Robinson, and J. Tepper, "Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2018, pp. 745–760. doi: 10.1007/978-3-319-93417-4_48.

[38] S. S. Aluru, B. Mathew, P. Saha, and A. Mukherjee, "Deep Learning Models for Multilingual Hate Speech Detection," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.06465

[39] A. Al-Hassan and H. Al-Dossari, "DETECTION OF HATE SPEECH IN SOCIAL NETWORKS: A SURVEY ON MULTILINGUAL CORPUS," Academy and Industry Research Collaboration Center (AIRCC), Feb. 2019, pp. 83–100. doi: 10.5121/csit.2019.90208.

[40] AshwinGeetd'Sa, IrinaIllina, and DominiqueFohr, "Ashwin," *Hal Open Science*, pp. 1–12, Jan. 2021.

[41] Y. Zhou, Y. Yang, H. Liu, X. Liu, and N. Savage, "Deep Learning Based Fusion Approach for Hate Speech Detection," *IEEE Access*, vol. 8, pp. 128923–128929, 2020, doi: 10.1109/ACCESS.2020.3009244.

[42] F. Alkomah and X. Ma, "A Literature Review of Textual Hate Speech Detection Methods and Datasets," *Information (Switzerland)*, vol. 13, no. 6. MDPI, Jun. 01, 2022. doi: 10.3390/info13060273.

[43] G. O. Ganfure, "Comparative analysis of deep learning based Afaan Oromo hate speech detection," *J Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00628-w.

[44] A. K. J, A. S, T. E. Trueman, and E. Cambria, "Comment toxicity detection via a multichannel convolutional bidirectional gated recurrent unit," *Neurocomputing*, vol. 441, pp. 272–278, Jun. 2021, doi: 10.1016/j.neucom.2021.02.023.

[45] *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. IEEE.

[46] D. H. Shih, C. H. Liao, T. W. Wu, X. Y. Xu, and M. H. Shih, "Dysarthria Speech Detection Using Convolutional Neural Networks with Gated Recurrent Unit," *Healthcare (Switzerland)*, vol. 10, no. 10, Oct. 2022, doi: 10.3390/healthcare10101956.

[47] W. Zaghouani, J. Alberto Benítez-Andrades, U. de León, S. Mabrouka Besghaier, A. Abdelali, and A. Toliyat, "Asian hate speech detection on Twitter during COVID-19," 2019.

[48] K. U. Wijaya and E. B. Setiawan, "Hate Speech Detection Using Convolutional Neural Network and Gated Recurrent Unit with FastText Feature Expansion on Twitter," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, pp. 619–631, 2023, doi: 10.26555/jiteki.v9i3.26532.

[49] A. Sharma, A. Zozan, and Z. R. Ahmed, "The 3D Facemask Recognition: Minimization for Spreading COVID-19 and Enhance Security."

[50] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B. W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020, doi: 10.1109/ACCESS.2020.3019735.

[51] L. Jiang, K. Japan, and Y. Suzuki, "Detecting hate speech from tweets for sentiment analysis." [Online]. Available: https://www.kaggle.com/pandeyakshive97/hate-speech-dataset.

[52] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, and M. Camacho-Collados, "Detecting

e-ISSN: 2614-8404
p-ISSN:2776-3234

and monitoring hate speech in twitter," *Sensors (Switzerland)*, vol. 19, no. 21, Nov. 2019, doi: 10.3390/s19214654.

[53] K. Dubey, R. Nair, M. U. Khan, and P. S. Shaikh, "Toxic Comment Detection using LSTM," in *Proceedings of 2020 3rd International Conference on Advances in Electronics, Computers and Communications, ICAECC 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/ICAECC50550.2020.9339521.

[54] L. Ketsbaia Northumbria University, B. Issac Northumbria University, and X. Chen Northumbria University, "Detection of Hate Tweets using Machine Learning and Deep Learning", doi: 10.1109/TrustCom50675.2020.00103/20/$31.00.

[55] P. K. Sahoo, S. Mishra, R. Panigrahi, A. K. Bhoi, and P. Barsocchi, "An Improvised Deep-Learning-Based Mask R-CNN Model for Laryngeal Cancer Detection Using CT Images," *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228834.

[56] T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A. G.-T. Nguyen, "Hate Speech Detection on Vietnamese Social Media Text using the Bi-GRU-LSTM-CNN Model," Nov. 2019, [Online]. Available: http://arxiv.org/abs/1911.03644

[57] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, "CNN-LSTM architecture for predictive indoor temperature modeling," *Build Environ*, vol. 206, Dec. 2021, doi: 10.1016/j.buildenv.2021.108327.

[58] S. Khan *et al.*, "BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4335–4344, Jul. 2022, doi: 10.1016/j.jksuci.2022.05.006.

[59] M. Beatty, "Graph-Based Methods to Detect Hate Speech Diffusion on Twitter," in *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 502–506. doi: 10.1109/ASONAM49781.2020.9381473.

[60] C. Paul and P. Bora, "Detecting Hate Speech using Deep Learning Techniques." [Online]. Available: www.ijacsa.thesai.org

[61] P. Malik, A. Aggrawal, and D. K. Vishwakarma, "Toxic Speech Detection using Traditional Machine Learning Models and BERT and fastText Embedding with Deep Neural Networks," in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 1254–1259. doi: 10.1109/ICCMC51019.2021.9418395.

[62] J. S. Malik, G. Pang, and A. van den Hengel, "Deep Learning for Hate Speech Detection: A Comparative Study," Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.09517

[63] A. Toktarova *et al.*, "Hate Speech Detection in Social Networks using Machine Learning and Deep Learning Methods." [Online]. Available: www.ijacsa.thesai.org

[64] S. Shekhar Pandey, I. Chhabra, R. Garg, and S. Sahu, "Hate Speech Detection," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 5, p. 897, 2023, doi: 10.35629/5252-0504897903.

[65] H. Saleh, A. Alhothali, and K. Moria, "Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model," *Applied Artificial Intelligence*, vol. 37, no. 1, 2023, doi: 10.1080/08839514.2023.2166719.

[66] A. Abraham, A. J. Kolanchery, A. A. Kanjookaran, B. T. Jose, and D. PM, "Hate Speech Detection in Twitter Using Different Models," *ITM Web of Conferences*, vol. 56, p. 04007, 2023, doi: 10.1051/itmconf/20235604007.

[67] M. Fazil, S. Khan, B. M. Albahlal, R. M. Alotaibi, T. Siddiqui, and M. A. Shah, "Attentional Multi-Channel Convolution With Bidirectional LSTM Cell Toward Hate Speech Prediction," *IEEE Access*, vol. 11, pp. 16801–16811, 2023, doi: 10.1109/ACCESS.2023.3246388.

[68] D. Marrugo, J. Carlos Martinez Santos, E. Puertas, D. Andres Marrugo-Tobón, and J. Carlos Martinez-Santos, "Natural Language Content Evaluation System For Multiclass Detection of Hate Speech in Tweets Using Transformers," 2023. [Online]. Available: https://github.com/EdwinPuertas

[69] G. M. Zebari, D. A. Zebari, D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and K. Yurtkan, "Efficient CNN Approach for Facial Expression Recognition," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Dec. 2021. doi: 10.1088/1742-6596/2129/1/012083.

[70] M. Jakubec, E. Lieskovská, B. Bučko, and K. Zábovská, "Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation," *Applied Sciences (Switzerland)*, vol. 13, no. 9. MDPI, May 01, 2023. doi: 10.3390/app13095810.