IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 3, 2021, accepted January 10, 2022, date of publication January 18, 2022, date of current version January 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3144266

Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model

ERNESTO LEE[1], FURQAN RUSTAM[2], PATRICK BERNARD WASHINGTON[3], FATIMA EL BARAKAZ[4], WAJDI ALJEDAANI[5], AND IMRAN ASHRAF[6]

[1]Department of Computer Science, Broward College, Broward County, FL 33301, USA
[2]Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Punjab 64200, Pakistan
[3]Division of Business Administration and Economics, Morehouse College, Atlanta, GA 30314, USA
[4]Department of Computer Science, Faculty of Sciences, Chouaib Doukkali University, El Jadida 24000, Morocco
[5]Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, USA
[6]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38544, South Korea

Corresponding authors: Imran Ashraf (ashrafimran@live.com) and Furqan Rustam (furqan.rustam1@gmail.com)

This work was supported by the Florida Center for Advanced Analytics and Data Science by Ernesto.Net (under the Algorithms for Good Grant).

ABSTRACT With social media's dominating role in the socio-political landscape, several existing and new forms of racism took place on social media. Racism has emerged on social media in different forms, both hidden and open, hidden with the use of memes and open as the racist remarks using fake identities to incite hatred, violence, and social instability. Although often associated with ethnicity, racism is now thriving based on color, origin, language, cultures, and most importantly religion. Social media opinions and remarks provocating racial differences have been regarded as a serious threat to social, political, and cultural stability and have threatened the peace of different countries. Consequently, social media being the leading source of racist opinions dissemination should be monitored and racism remarks should be detected and blocked timely. This study aims at detecting Tweets that contain racist text by performing the sentiment analysis of Tweets. Owing to the superior performance of deep learning, a stacked ensemble deep learning model is assembled by combining gated recurrent unit (GRU), convolutional neural networks (CNN), and recurrent neural networks RNN, called, Gated Convolutional Recurrent- Neural Networks (GCR-NN). GRU is on the top in the GCR-NN model to extract the suitable and prominent features from raw text, CNN extracts important features for RNN to make accurate predictions. Obviously, several experiments are conducted to investigate and analyze the performance of the proposed GCR-NN within the scope of machine learning and deep learning models indicating the superior performance of GCR-NN with increased 0.98 accuracy. The proposed GCR-NN model can detect 97% of the tweets that contain racist comments.

INDEX TERMS Racism, social media, online abuse, Twitter, deep learning.

I. INTRODUCTION

Social media has become a dominating element in socio-political prospects and controls our minds and actions in different ways. With the wide use of social media platforms over the world and freedom of speech, several vices have emerged over the past few years, racism being one of the leading ones. Social media sites, such as Twitter, represent a new setting in which racism and related stress are apparently prospering [1]. Currently, 22% of United States (US) adults use Twitter [2], while Twitter has 1.3 billion accounts and 336 million active users across the globe, 90% of which has a public profile leading to 500 million tweets per day [3]. Unless tweets are made private, they are publicly available and Twitter users can react to such tweets and engage by sharing them on their profile (retweet), tagging someone's user name, clicking the like button, or responding to the author of the tweet [4]. In Twitter, the expression of feelings, emotions, attitudes, and opinions build the raw data of sentimental analysis [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Hualong Yu.

The growing popularity of social media platforms has led to their wide use for several old and new forms of racist practices [6]. Racism is expressed on such platforms in different surreptitious forms such as memes and openly such as posting Tweets containing racist remarks using fake identities. Although often associated with ethnicity, racism is now thriving based on color, origin, language, cultures, and most importantly religion. Social media opinions and remarks provocating racial differences have been regarded as a serious threat to social, political, and cultural stability and have threatened the peace of different countries. Social media being the leading source of racism opinions dissemination should be monitored and racism remarks should be detected and blocked timely.

Racist comments and tweets on social media have been regarded as the source of several kinds of mental and body illness leading to adverse health outcomes [7]–[12]. With respect to its use on social media, racism can be categorized into three groups: institutionalized, personally mediated, and internalized [13]. Personally mediated racism can be experienced through racial discrimination or differential racial treatment, or through awareness of discrimination against family and friends. Consequently, the racist behavior of the society adversely affects individuals and ignites several kinds of psycho-social stress often leading to the risk of chronic diseases [14]–[16]. Additionally, racist groups and individuals perpetuate cyber-racism by employing higher skill levels and intricacy through various channels and strategies [5].

Special considerations have been given to the field of sentiment analysis to analyze the text from social media platforms for a large variety of tasks including hatred speech detection, market prediction based on sentiments, and racism detection, etc.

The wide use of social media is a potential source of data generation containing important information regarding people's attitudes, responses, emotions, and opinions regarding specific events, objects, personalities, and entities. Sentiment analysis provides powerful tools to mine such data to analyze emotions. The huge part of Twitter feeds become less characterized by coherent rational discussion, but more by floods of emotion and affect, and can be used to divide the narratives into polarities of good and evil [17], [18]. Research shows that issues may become less obvious than a shared sense of outrage and a compelling sense of shared agreement and Twitter feeds can be quite insular and nodal [19].

Keeping in view its wide use, social media has become an attractive source to apprehend attitudes and analyze interactions over sensitive topics such as racism. In the USA, the discussions about race and ethnicity on Twitter have been considered as indicators of the current state of relations based on race. Additionally, the variation in the types of discussions about racism indicates the geographic variability in racial attitudes and sentiment [20]. So, analyzing the details of how people, events, and circumstances are represented reveals the dynamics of how users communicate, and many problems related to racism can be exposed on this platform. Owing to the extreme and atypical racist attitude an individual faces related to personal traits and attitudes, one can easily become relativized, contextualized, and therefore depoliticized. It leads to distracting attention from the actual and specific structural inequalities in society experienced by certain ethnic groups [21].

Machine and deep learning approaches has proven their strength and superiority over traditional methods in several domains such as image processing [22], [23], text classification [24], [25] and sentiment analysis is no exception. Several recent studies show that machine learning techniques perform better for sentiment analysis tasks [26], [27]. Therefore, this study leverage machine learning and deep learning models to perform sentiment analysis on tweets related to racism and makes the following contributions

- An ensemble model is proposed that makes use of recurrent neural networks. For this purpose, gated recurrent unit (GRU), convolution neural network, and recurrent neural network are stacked to make the GCR-NN model to perform sentiment analysis.
- A large dataset of tweets containing racist comments/text is crawled from Twitter which can be used by the research community. The dataset is annotated using the TextBlob based on the polarity score into positive, negative,and neutral sentiments.
- For performance comparison, several well-known machine learning models are implemented using the optimized parameters such as decision tree (DT), random forest (RF), logistic regression (LR), k nearest neighbor (KNN), and support vector machines (SVM). Term frequency-inverse document frequency (TF-IDF) and bag of words (BoW) are studied as feature extraction techniques.
- For a fair comparison with the proposed approach, GRU, long short term memory (LSTM), CNN, and RNN are implemented as standalone models. Similarly, the performance of several state-of-the-art models is compared with the proposed GCR-NN in terms of accuracy, precision, recall, and F1 score.

The rest of the paper is organized as follows. Section II describes several important research papers related to the current study. The proposed approach, dataset, and description of machine learning algorithms are given in Section III. Section IV provides the analysis and discussion of results. In the end, the conclusion is drawn in Section V.

II. RELATED WORK

The overwhelming effects of hate crimes are increasing to a great extent because of the extensive use of social media [37] and the anonymity enjoyed by online users [38]. Abusive content and intricate stuffing on social media is a problematic phenomenon with more than a few overlapping and coinciding modes and aims [31]. The contents related to harassment and maltreatment arouses negative feelings in online users so they express their feelings in a discourteous way. Cyberbullying and hate speeches are two examples of

TABLE 1. A summary of the discussed research works.

Ref.	Model	Dataset	Accuracy
[28]	Varients of BERT and Resnet	https://github.com/kperi/ MultimodalHate-SpeechDetection	0.97
[29]	resnet18 + nlpaueb/greek-bert, Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) with TF-IDF, profile-related and emotion-related features.	3696 tweets, Self-made	0.913
[30]	Random Forest (RF) with TF-IDF and profile related features, Naïve Bayes, Logistic Regression, XGBoost and TF-IDF features	de Gibert, O. a. (2018). Hate Speech Dataset from a White Supremacy Forum. Association for Computational Linguistics, Github	XGBoost with TF-IDF, Recall:0.83, Precision:0.82
[31]	BERT,CNN,GRU and the ensemble of CNN and GRU (CNN+GRU)	selfmade	F1 score: 0.79 CNN
[32]	Distributed Bag of words (DBoW), Distributed Memory Mean (DMM), and Word2Vec CNN	1st dataset: university of Maryland, 2nd dataset: self-made 25000 tweets	1st dataset=96.67%, 2nd dataset=97.5%, Neural Network with 3 hidden layers with Doc2Vec
[33]	Naïve Bayes,Multilayer Preceptron,AdaBoost classifier,Support Vector Machine	Self-made tweeter dataset 4002 tweets	83.4%, MLP with SMOTE 71.2%, AB, MNB, BNB
[34]	Multinomial Naïve Bayes,Linear SVM, Random Forest and RNN	Self-made, Youtube	0.9464 for the first experiment and 0.857 for the second experiment
[35]	NB, RF,LR,DT, SVM and deep learning models	Self made : tweeter	SVM 74.6%
[36]	XGBoost,SVM,LR,NB,and FFNN	YouTube dataset (ICWSM 18 SALMINEN), Reddit dataset (ALMEREKHI 19), Wikipedia dataset (KAGGLE 18), Twitter dataset (DAVIDSON 17 ICWSM)	F1 score =0.92, XGBoost

abusive languages that have vexed the interest of researchers in recent times owing to their harmful effects on society. Decontamination of these contents is very necessary. For this purpose, several studies have been conducted to automatically detect the annoying hate speeches and massages among other contents on social media. Automatic hate speech detection using machine learning algorithms is still new and requires extensive research efforts from both industry and academia [39]. Few recent and related papers have been discussed here [40], [41]. Machine learning algorithms have contributed enormously to hate speech detection and content analysis [37].

The authors present a multimodal hate speech detection model specifically for Greek social media in [28]. The study focuses on Twitter messages, especially racist speech and xenophobia, in Greek aimed at migrants and refugees. The ensemble model, the transfer learning, and fine-tuning of the bidirectional encoder representations from transformers (BERT) and Resnet are used on the collected dataset. Different variants of the BERT and Resnet are used and the highest accuracy of 0.944 is reported using nlpaueb/greek-bert for the text modality and 0.97 with resnet18+ nlpaueb/greek-bert using text+image modality. Similarly, [29] proposes a state-of-the-art machine learning-based system for the automatic detection of hate speech in Arabic social media networks. Several types of emotions are captured and a different set of features are used for analysis. The study uses four different machine learning algorithms such as Naïve Bayes (NB), DT, SVM, and RF with TF-IDF, profile-related, and emotion-related features. RF with TF-IDF and profile-related features achieved the highest accuracy 0.913.

Along the same lines, [30] classifies the fake news and hate speech propaganda using the extracted features from the content containing fake and real news. The study uses NB, LR, and XGBoost with TF-IDF features. XGBoost demonstrates a recall value of 0.83 which indicates that 17% of data contains hatred content and is misclassified by the model. Also, XGBoost achieves the precision value of 0.82 which shows that 18% of data is hateful and the model misclassified it. Authors investigate the hate speech problem in the Saudi Twitter sphere in [31] using different deep learning approaches. A series of experiments are conducted on two datasets using BERT, CNN, GRU, and the ensemble of CNN and GRU (CNN+GRU). Results indicate that the model achieves an F1 score of 0.79 and the area under receiver operating curve (AUROC) of 0.89 using the CNN model.

Study [32] investigates the automatic detection of cyberbullying. To review the deep learning and machine learning approaches, the authors use two different datasets. Different word embedding techniques such as distributed BoW (DBoW), distributed memory mean (DMM) and Word2Vec CNN are used to classify online racism. An accuracy of 96.67% for one dataset while 97.5% for the second dataset is achieved using a neural network with 3 hidden layers using Doc2Vec features. In the same way, study [33] explores the automatic detection of Indonesian tweets that contain hate speech or racism. The authors use machine learning models such as multinomial NB (MNB), Multilayer Perceptron (MLP), AdaBoost (AB) classifier, and SVM. Synthetic minority oversampling technique (SMOTE) is used as an upsampling technique and experiments are performed on both SMOTE and non SMOTE features. Results show that MLP with SMOTE features has an accuracy of 83.4% and AB, and MNB has 71.2% accuracy for non-SMOTE features.

Ching She *et al.* work on hate speech detection from social media in [34]. For experiments, the audio data is extracted from videos and converted to text using a speech-to-text converter. MNB, Linear SVM, RF, and RNN are used for experiments. Two different sets of experiments are carried

out where the first experiment involves classifying the video into normal and hateful videos while the second experiment aims at classifying the video into normal, racist, and sexiest classes. Results show that RF shows superior performance in terms of accuracy and achieves an accuracy of 0.9464 for the first set of experiments and 0.857 for the second set of experiments.

Another similar work is [35] which investigates hate speech related to Islam on social media. The study constructs an automated tool that can distinguish between non-islamophobic, weak islamophobic, and strong islamophobic content. Different machine learning algorithms such as NB, RF, LR, DT, SVM, and deep learning models are used. Results suggest that SVM obtains the testing accuracy of 72.17%. The performance of SVM is also evaluated using 10 fold cross-validation which shows a 74.6% accuracy and balanced accuracy of 80.7%. Study [36] proposes a novel system to detect hate speech across multiple social media plate forms like Reddit, YouTube, Twitter, and Wikipedia. A large dataset is built from these social media platforms with 80% labeled as non-hateful and 20% labeled as hateful. Several machine learning algorithms such as XGBoost, SVM, LR, NB, and feed-forward neural networks and tested with BoW, TF-IDF, Word2Vec, BERT, and their combinations. XGBoost outperforms all models with a 0.92 F1 score with all features. Feature importance analysis shows that BERT features have a great effect on predictions.

Taking into account the reported results from deep learning models, this study leverages the deep learning ensemble model to detect racism comments from Twitter. The study aims at obtaining high classification accuracy by stacking recurrent neural networks. Racism detection is performed using sentiment analysis where the ratio of tweets containing negative sentiments indicates the racist tweets.

III. MATERIALS AND METHODS

A. PROPOSED METHODOLOGY

This study proposes an approach for racism detection on social media platforms using machine learning and deep learning technique. Figure 1 shows the flow of the steps carried out in the proposed approach. As the first step is crawled from Twitter, followed by data cleaning and preprocessing, and finally the data annotation. In the end, the proposed stacked ensemble model is trained and tested on the datasets and its performance is compared with several other deep learning and machine learning models.

B. DATASET DESCRIPTION

The racism tweets dataset is collected from Twitter. Twitter has been the first choice of the majority of researchers for text and sentiment analysis due to its being the most common platform widely used by a large number of people to express their feelings, views, comments, and opinions. In particular, this study intends to study the racism trends

based on Twitter posts.

For data collection, tweets related to racist comments have been collected. For this purpose, several keywords are used such as, '#racism', '#racial', and '#racist', etc. for data collection for the period of 29 July 2021 to 6 August 2021. A total of 169,999 tweets have been collected that match the criteria. The data are collected using the 'Twint library' and important attributes such as 'username', 'date', 'location', and 'content' are extracted. A specimen of collected tweets is provided in Table 2.

TABLE 2. Sample text from the dataset.

User	Text
@_LeBale racism	@_LeBale racism is good
tonyhasanidea	@manoutdoors4 @AJ_Lady_Liberty @FBIWFO @TheJusticeDept @FBI it is clear to hundreds of millions of people of all walks that this country has a severe problem with systemic racism. your denial is discussing.

C. DATA PREPROCESSING

Several steps are carried out at the preprocessing level to clean the data. It is vital to preprocess and clean the document adequately so a model can be trained appropriately. This study combines natural language processing (NLP) methods using the natural language toolkit (NLTK) of Python to preprocess the reviews.

- Tokenization is the process of splitting natural texts into tokens without any white spaces. It involves breaking sentences down into constituent words set. Although looks simpler and straightforward, deciding which tokens are appropriate is not a trivial task.
- Stemming: The text contains different forms of the same word which can create complexity in machine learning models. Words such as 'go', 'gone', and 'going' are the modified forms of 'go'. Stemming converts each word into its root form such as 'gone', and 'going' will be transformed into 'go'. Stemming is performed using the Stemmer porter algorithm.
- Lemmatization: It is a similar procedure to that of tokenization, however, produces a different output. Tokenization simply removes 's' or 'es' at the end of a word to change it to its root form which often results in wrong words/spelling. Lemmatization retains the root form of a word by considering the context in which a word is used. It also lowers the unique occurrence's count of similar words. This approach is used in the suggested strategy for word preprocessing in their canonical format to limit the unique occurrences count of identical text tokens.
- Stop Words Exclusion: Stop words are words that do not contribute to the training of the machine learning algorithms. Instead, they create complexity by increasing the feature space. So, stop words such as a, am, and an, etc., are removed to increase the learning efficiency of models in this study.
- Case Normalization: Because precise words having various cases must be treated in a similar way, such as "Racism" & " racism," the entire text must be converted to lowercase letters. It is commonly referred to

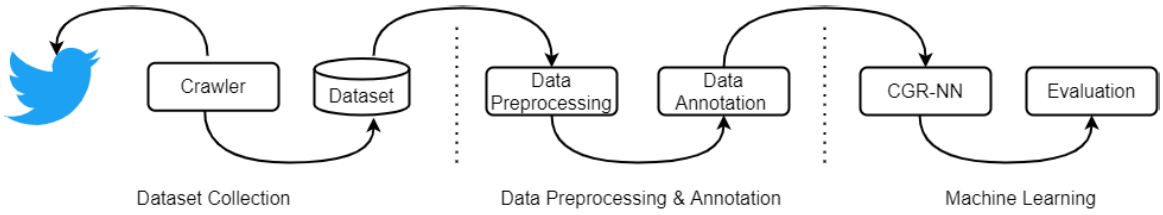

FIGURE 1. Architecture of the proposed methodology.

as data cleansing because it aids in minimizing the repetition of similar features that vary only with regard to case sensitivity.

- Noise Removal: This stage removes any noise that could degrade the performance of the classification. Special characters, numeric data, id, and '#' signs, etc. are examples of noise types deleted in this phase.

The sample tweets are preprocessed using the above-discussed steps and the resulting text is given in Table 3.

TABLE 3. Sample text before and after the preprocessing.

Before preprocessing	After preprocessing
@_LeBale racism is good	racism good
@manoutdoors4 @AJ_Lady_Liberty @FBIWFO @TheJusticeDept @FBI it is clear to hundreds of millions of people of all walks that this country has a severe problem with systemic racism. your denial is discussing. the world is changing , get on board or get left	clear hundr million people walk country sever problem system racism denial

D. DATA ANNOTATION

To annotate the dataset with positive, negative, and neutral sentiments, this study uses the TextBlob library. Textblob finds the polarity score for a given text which is used to assign a sentiment label to the text. Textblob polarity score range varies between −1 to 1. The polarity score range for positive, negative, and neutral sentiments is shown in Table 4.

TABLE 4. Data annotation using polarity score.

Sentiment	Polarity score
Neutral	=0
Positive	>0
Negative	<0

After the annotation, the distribution of tweets are shown in Figure 2. It shows the ratio of positive, negative, and neutral sentiments in the dataset. The number of records for the three classes is almost similar, with neutral sentiments making the major part of the dataset.

E. FEATURES EXTRACTION

BoW and TF-IDF are used for features extraction to train the machine learning models. Each feature extraction technique gives 125,461 features for models' training.

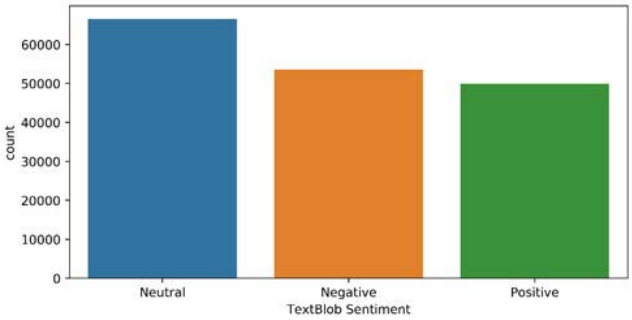

FIGURE 2. Ratio of sentiment in dataset.

1) TERM FREQUENCY - INVERSE DOCUMENT FREQUENCY

TF-IDF is among the most commonly employed scoring metrics for summarization and information retrieval. It is utilized to measure the significance of the term within a given text [42]. The TF-IDF extraction function takes two inputs: IDF and TF. TF-IDF provides tokens that seem to be uncommon within a dataset. When uncommon words appear in multiple documents, their relevance grows.

$$TF - IDF_{t,d,D} = TF_{t,d} * IDF_{t,D} \quad (1)$$

where t denotes terms, d denotes each document, and D is the documents set. The parameter n-gram range is used in conjunction with TF-IDF. TF-IDF is used to compute word weights, which offer corpus weights for a given word. The weighted word matrix is the output. The TF approach is frequently used for extracting features and therefore is widely utilized for text categorization. During classifier training, the incidence frequency of terms' is used as a parameter. TF function does not consider the importance of rare words, in contrast to the TF-IDF, which gives less weight to more frequent terms. TF-IDF results on the sample preprocessed data are shown in Table 5.

2) BAG OF WORDS

The BoW is another commonly used feature extraction used in NLP tasks.

It is the most convenient and adaptable approach to get a document's features [43].

The Word's histogram within the text is examined in BoW. The frequency of the words is employed as a function for the training of the set. The BoW approach is implemented in this study by utilizing the Count Vectorizer from the Scikit-learn

TABLE 5. Results of TF-IDF feature extraction on sample tweets.

clear	country	denial	good	hundr	million	people	problem	racism	sever	system	walk
0.000000	0.000000	0.000000	0.814802	0.000000	0.000000	0.000000	0.000000	0.579739	0.000000	0.000000	0.000000
0.308515	0.308515	0.308515	0.000000	0.308515	0.308515	0.308515	0.308515	0.219511	0.308515	0.308515	0.308515

TABLE 6. Results of BoW feature extraction on sample tweets.

clear	country	denial	good	hundr	million	people	problem	racism	sever	system	walk
0	0	0	1	0	0	0	0	1	0	0	0
1	1	1	0	1	1	1	1	1	1	1	1

library of Python. The technique of obtaining numerical vectors by transforming a textual data set is termed vectorization. The frequency of words is counted indicating that tokens have been counted and making the token vectors. The BoW assigns a value to every attribute based on the frequency of those features. BoW results on sample preprocessed data are shown in Table 6.

F. MACHINE LEARNING MODELS

For racism detection from tweets, machine learning models have been adopted due to their superior performance over traditional models. Some of the renowned models such as RF, LR, DT, SVM, and KNN are discussed briefly in this paper for completeness. The performance of these models is optimized by fine-tuning several hyperparameters. A complete list of parameters used in this study is provided in table 7 along with the range used for optimization, as well as, the used values for experiments.

1) RANDOM FOREST

RF is a tree-based classifier that builds trees based on a random vector taken from the input vector [44]. Initially, RF builds a forest by producing multiple decision trees using random features. Later, voting is performed by aggregating the decision from all decision trees to make the final prediction. Votes from a decision tree with a low error rate are given a higher weight and vice versa. By using decision trees with low error rates, reduces the chances of wrong prediction [1]. RF can be defined by the equations:

$$p = mode\{T_1(y), T_2(y), \ldots, T_m(y)\} \tag{2}$$

$$p = mode\{\sum_{m=1}^{m} T_m(y)\} \tag{3}$$

2) LOGISTIC REGRESSION

LR is a statistical-based classifier that is mostly used for the analysis of binary data in which one or more variables are used to find the results. It is also used for probability evaluation of class association [45]. LR is especially recommended for categorical data due to its superior performance. It finds the affiliation between the dependent and one or more independent variables of the categorical data using approximation. For probability approximation, LR makes use of a logistic function. A logistic function or logistic curve is a common "S" sloped or sigmoid curve defined as

$$f(x) = \frac{L}{1 + e^{-m(v - v_0)}} \tag{4}$$

3) SUPPORT VECTOR MACHINE

SVM is a well-known machine learning algorithm that is widely used for the classification of linear, as well as, non-linear data. For binary classification problems, it is the first choice of many researchers and it is available in various kernel functions [25]. The main purpose of the SVM classifier is to estimate the hyperplane based on feature set to classify data points [44]. The dimensions of the hyperplane vary with respect to the number of features. As multiple possibilities exist for hyperplanes in n-dimensional space, the task is to derive hyperplanes that maximize the margins between samples of classes. The cost function used to determine the hyperplanes is given by

$$J(\theta) = \frac{1}{2}\sum_{j=1}^{n} \theta_j^2, \tag{5}$$

such that

$$\theta^T x^{(i)} \geq 1, \quad y^{(i)} = 1, \tag{6}$$

$$\theta^T x^{(i)} \leq -1, \quad y^{(i)} = 0, \tag{7}$$

4) K NEAREST NEIGHBOR

KNN is a simple and widely used machine learning algorithm for both classification and regression problems. KNN assumes that similar data can be found in close proximity, so it uses the concepts of 'neighbors'. It estimates the distance of the new data points to its neighbors by using distance calculation metrics such as Euclidean distance, Manhattan distance, and Minkowski distance, etc. In KNN, the value of K determines the number of neighbors to be considered for prediction. Well-known distance calculation metrics are given here [46]:

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}, \tag{8}$$

$$Manhattan\ Distance = \sum_{i=1}^{k}|x_i - y_i|, \tag{9}$$

TABLE 7. Hyperparameters and their range used for fine tuning machine learning models.

Model	Hyperparameter	Hyperparameter tuning range
DT	Max_depth=300	Max_depth=100 to 500
RF	Max_depth=300, N_estimators=300	Max_depth=100 to 500, N_estimators=100 to 500
LR	Solver='saga', multi_class='ovr', C=3.0	Solver=saga,sag, multi_class='ovr', C=1.0 to 5.0
KNN	N_neighbour=5	N_neighbour=2 to 15
SVM	kernel=linear, C=3.0	kernel=linear, poly, C=1.0 to 5.0

$$Minkowski\ Distance = \left(\sum_{i=1}^{k} |x_i - y_i|^q\right)^{\frac{1}{q}}, \qquad (10)$$

5) DECISION TREE

DT is a ruled-based supervised machine learning algorithm. DT is a renowned and powerful predictive model which can handle regression and classification problems efficiently. Attribute selection is the major problem in DT [47] and information gain and Gini index are the most used methods for attribute selection. Information gain is the rate of increase or decrease in the entropy of attributes where entropy shows how homogeneous a dataset is [43].

$$E(D) = -P(positive)log2.P(positive) - P(negative)log2.P(negative)$$

The above equation computes the entropy E of a given dataset D which contains the positive and negative decision attributes. Gain of the attribute X is calculated by the formula:

$$Gain\ (attribute\ X) = Entropy(Decision\ Attribute\ Y) - Entropy(X, Y)$$

6) PROPOSED GATED CONVOLUTIONAL RECURRENT NEURAL NETWORKS

The proposed model GCR-NN is a combination of GRU, CNN, and RNN. This study combines these models in a stack as GRU is working on the top, CNN is working in middle followed by the RNN. The selection of these models to make an ensemble is based on their individual performance. GRU takes the input from the embedding layer with a 5000 vocabulary size. This input is processed by the GRU model to extract features for the following layers. GRU architecture is used with 64 units, followed by a CNN layer that uses the output from the GRU model. CNN layer is used with 64 filters and a kernel with 4×4 kernel size. CNN layer is followed by the max-pooling layer with a pooling size of 4. A dropouts layer with a 0.2 dropouts rate is also used to reduce the complexity in GCR-NN because the dropout layer will randomly delete the neurons and reduce the chances of model overfitting. RNN is working at the end of the GCR-NN model with 16 units. The outputs of the GRU and CNN are directed to the RNN model. At the end of RNN, a dense layer is used with 3 neurons and a softmax activation function because of three target classes. The model is compiled with categorical_crossentropy loss function because of multi-class problem and 'adam' optimizer is used for training [48]. The model is fit using 100 epochs and a batch size of 16.

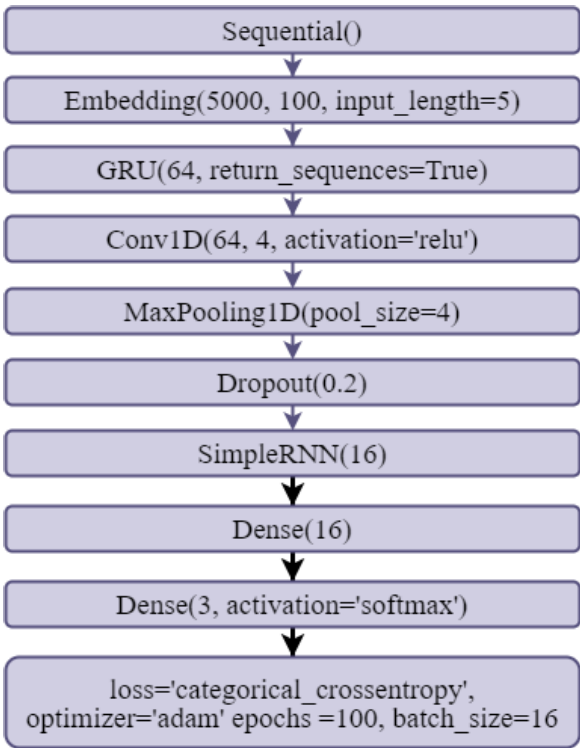

FIGURE 3. Structure of the proposed GCR-NN.

IV. RESULTS AND DISCUSSIONS

Experiments for sentiment analysis on racism tweets have been carried out using an Intel Corei7 11th generation machine operating on Windows 10. Machine learning and deep learning models are implemented on Jupyter in python language using Tensor-flow, Kara's, and Sci-kit learn frameworks. The performance of all models is evaluated in terms of accuracy, precision, recall, F1 score, number of correct predictions, and number of wrong predictions.

A. VISUAL REPRESENTATION OF SENTIMENT DISTRIBUTION

For providing the distribution of the dataset, with respect to country, data is divided into the top four countries with respect to the highest number of tweets. Figure 4a shows that the highest number of tweets are posted from the US, followed by the United Kingdom (UK), Nigeria, and Republic of South Africa (RSA) when racist content is considered.

Tweets sentiments distribution for each of the top four countries is given in Figure 4. It shows that the majority of the tweets belong to the neutral class for the US, UK, and

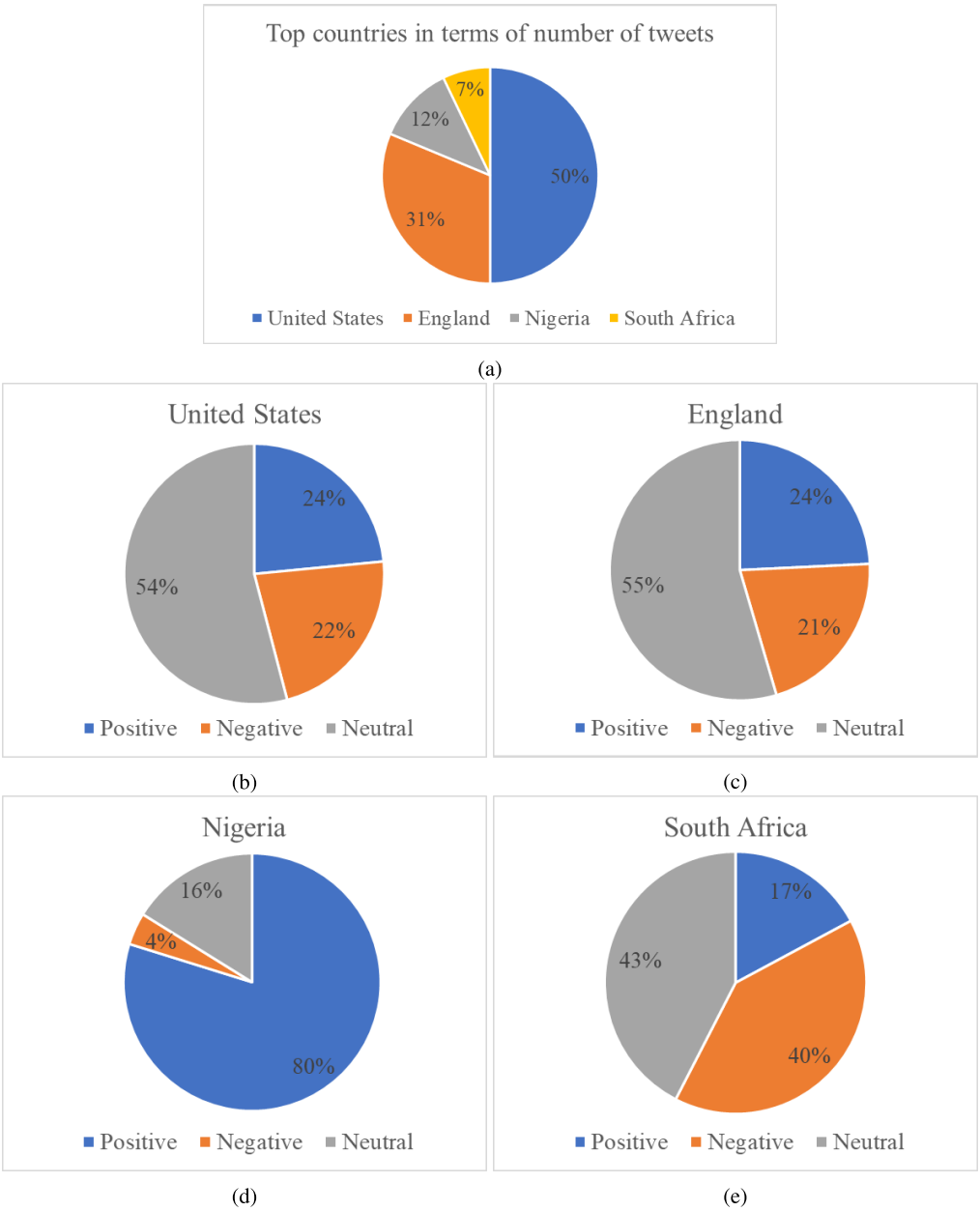

FIGURE 4. Distribution of tweets sentiments in different countries, (a) Ratio of tweets from top four countries in terms of tweets numbers, (b) United States, (c) England, (d) Nigeria, and (e) Republic of South Africa.

RSA with 54%, 55%, and 43% neutral tweets, respectively. The highest ratio of negative tweets comes from RSA which is 40% of the tweets originated from RSA. On the other hand, the highest number of positive tweets regarding racism originates from Nigeria with 80% of the total tweets from Nigeria. The ratio of positive and negative tweets is approximately similar in the US and the UK. Figure 5, show the word frequency in the dataset through word-cloud.

B. MACHINE LEARNING MODELS RESULTS USING BoW AND TF-IDF

This section contains the results of machine learning models using BoW and TF-IDF features. Table 8 shows the performance of all machine learning models using TF-IDF features and results show that the performance of linear machine models is significantly better as compared to other models. Results indicate that SVM achieves the highest accuracy of 0.97 and LR achieves a 0.96 accuracy score. These models are best performers when the feature set is large as is this study where the TF-IDF feature size is 125,461. These can be appropriate conditions for both SVM and LR models. RF is also good in terms of accuracy with a 0.91 accuracy score. In this study, the RF ensemble model combines 300 DT under majority voting criteria and this ensemble architecture makes RF a significant model in terms of accuracy. KNN is very poor in performance because it is a lazy learner which can perform better when the dataset is small.

racism.png

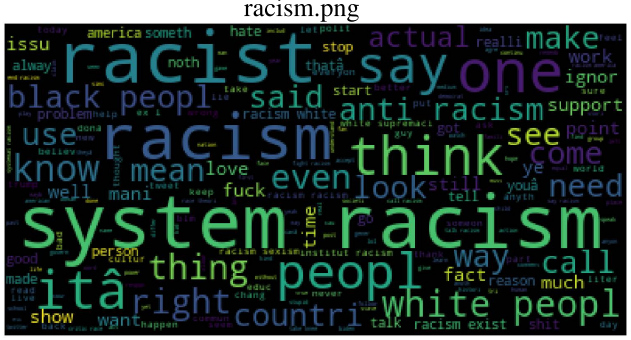

FIGURE 5. Word-cloud for the dataset.

Experimental results of machine learning models using BoW features are given in Table 9. Results suggest that SVM and LR show better performance even when used with BoW features. Both SVM and LR obtain a 0.97 accuracy score which is substantially better than all other models. LR and RF both improve the accuracy by 1% with BoW features as compared to when trained on TF-IDF features. The improvement in the performance is due to simple BoW features which aid in better training of machine learning models. TF-IDF gives a weighted feature set which can be complex when there is a large feature set while BoW gives a simple set that can be more appropriate for training machine learning models. The performance of KNN models is also elevated from 42% accuracy to 52% accuracy which is a significant improvement. On average, the performance of the machine learning models is better using BoW features as compared to their performance when TF-IDF features are used.

C. RESULTS USING DEEP LEARNING MODELS

For performance evaluation and a fair comparison with the proposed ensemble deep learning model, several single deep learning models are implemented as well such as GRU, LSTM, CNN, and RNN. The performance of deep learning models is optimized by setting different structures in terms of the number of layers, loss function, optimizer and number of neurons, etc. Results of all deep learning models are provided in Table 10. Results show that deep by large, learning models perform better than machine learning models. Deep learning is data-intensive, and large dataset gathered for racism detection leads to better training and results for deep learning models. LSTM and RNN perform equally well with an accuracy of 0.95, however, GRU and CNN obtain higher accuracy of 0.97. Consequently, GRU and CNN models are used to make a stacked ensemble with the RNN. RNN is used in the proposed model because RNN is better in terms of computational cost. GCR-NN outperforms all machine learning and deep learning models with a 0.98 accuracy score. This significant performance of the proposed model is due to its stacked ensemble architecture. Processing of data through GRU and CNN provides more appropriate and important features for RNN to make the final prediction. GCR-NN outperforms all other models in terms of all evaluation parameters as it achieves 0.98 scores for accuracy, precision, recall and F1 score.

TABLE 8. Results using machine learning models with TF-IDF features.

Model	Accuracy	Class	Precision	Recall	F1 score
DT	0.71	Negative	0.90	0.55	0.69
		Neutral	0.60	0.99	0.75
		Positive	0.87	0.49	0.63
		Macro Avg.	0.79	0.68	0.69
RF	0.91	Negative	0.92	0.88	0.90
		Neutral	0.88	0.98	0.93
		Positive	0.94	0.83	0.88
		Macro Avg.	0.91	0.90	0.90
KNN	0.42	Negative	0.81	0.06	0.11
		Neutral	0.40	0.99	0.57
		Positive	0.86	0.04	0.08
		Macro Avg.	0.69	0.36	0.25
SVM	0.97	Negative	0.96	0.97	0.97
		Neutral	0.99	0.99	0.99
		Positive	0.96	0.96	0.96
		Macro Avg.	0.97	0.97	0.97
LR	0.96	Negative	0.96	0.97	0.96
		Neutral	0.97	0.99	0.98
		Positive	0.96	0.95	0.95
		Macro Avg.	0.96	0.96	0.96

TABLE 9. Results using machine learning models with BoW features.

Model	Accuracy	Class	Precision	Recall	F1 score
DT	0.72	Negative	0.90	0.55	0.69
		Neutral	0.61	0.99	0.76
		Positive	0.89	0.52	0.66
		Macro Avg.	0.80	0.69	0.70
RF	0.91	Negative	0.92	0.87	0.90
		Neutral	0.88	0.99	0.93
		Positive	0.94	0.84	0.89
		Macro Avg.	0.91	0.90	0.90
KNN	0.52	Negative	0.84	0.29	0.43
		Neutral	0.46	0.99	0.63
		Positive	0.89	0.14	0.25
		Macro Avg.	0.73	0.48	0.44
SVM	0.97	Negative	0.96	0.97	0.96
		Neutral	0.99	0.99	0.99
		Positive	0.96	0.96	0.96
		Macro Avg.	0.97	0.96	0.96
LR	0.97	Negative	0.97	0.95	0.96
		Neutral	0.97	0.99	0.98
		Positive	0.97	0.96	0.96
		Macro Avg.	0.97	0.97	0.97

Models' performance is also evaluated in terms of the number of correct predictions (CP) and wrong predictions (WP). SVM gives the highest number of correct predictions using BoW with respect to machine learning models as SVM gives 41,397 correct predictions and 1,103 wrong predictions. SVM also outperforms using the TF-IDF features in terms of correct predictions as it gives 41,361 correct predictions and 1,139 wrong predictions. With respect to both machine learning and deep learning models, the proposed model GCR-NN gives 41,520 correct predictions and 980 wrong predictions which is the highest correct prediction ratio for all the models used in this study.

D. COMPARISON WITH PREVIOUS RESEARCH WORKS ON RACISM

To show the significance of the proposed approach, the results of the proposed GCR-NN are compared with other studies.

TABLE 10. Results using deep learning models.

Model	Accuracy	Class	Precision	Recall	F1 score
GRU	0.97	Negative	0.96	0.96	0.96
		Neutral	0.97	0.98	0.97
		Positive	0.96	0.96	0.96
		Macro Avg.	0.97	0.97	0.97
LSTM	0.95	Negative	0.99	0.87	0.93
		Neutral	0.99	0.98	0.99
		Positive	0.87	0.99	0.93
		Macro Avg.	0.95	0.95	0.95
CNN	0.97	Negative	0.95	0.96	0.96
		Neutral	0.99	0.96	0.98
		Positive	0.94	0.95	0.95
		Macro Avg.	0.97	0.96	0.96
RNN	0.95	Negative	0.93	0.94	0.93
		Neutral	0.98	0.98	0.98
		Positive	0.95	0.92	0.94
		Macro Avg.	0.95	0.95	0.95
GCR-NN	0.98	Negative	0.97	0.97	0.97
		Neutral	0.99	0.99	0.99
		Positive	0.97	0.97	0.97
		Macro Avg.	0.98	0.98	0.98

The study [49] uses the dataset related to racism and hate speech. The dataset has only two target classes of 'racism' and 'no racism' as compared to the current study which uses three classes for experiments. The study leverages XGBoost for racism detection and obtains an accuracy and F1 scores of 0.69 each. The proposed model in this study, on the other hand, achieves a 0.95 accuracy score and whos far better results than previous studies even with the multi-class task. Another dataset related to US airline sentiments is also considered for performance evaluation which is taken from [50]. The proposed model is implemented using the dataset [50] for performance evaluation on a small dataset. Results indicate that GCR-NN performs well on the US airline dataset with 0.81 accuracy.

E. DISCUSSIONS

This study aims at identifying racist content posted in the tweets by performing sentiment analysis. For this purpose, the dataset is annotated into positive, negative, and neutral classes. Positive and neutral classes indicate that racist content is not present in such tweets while negative class indicates that these tweets are racist as they contain negative views related to racism. So a distribution of correct and wrong predictions and accuracy is provided here with respect to the negative class.

TABLE 11. Number of correct and wrong predictions using machine learning models.

Features	Model	CP	WP
BoW	DT	30,490	12,010
	RF	38,551	3,949
	KNN	22,191	20,309
	SVM	41,397	1,103
	LR	41,143	1,357
TF-IDF	DT	30,048	12,452
	RF	38,507	3,993
	KNN	16,650	24,850
	SVM	41,361	1,139
	LR	41,030	1,470
NN	LSTM	40,442	2,058
	GRU	41,152	1,348
	CNN	41,152	1,348
	RNN	41,411	1,089
	GCR-NN	41,520	980

TABLE 12. Comparison with state-of-the-art approaches.

Ref.	Dataset	Model	Accuracy	F1 Score
[49]	Racism Tweets	XGBoost	0.69	0.69
Our study	Racism Tweets	GCR-NN	0.95	0.86
[50]	US Airline Tweets	LR	0.77	0.76
Our study	US Airline Tweets	GCR-NN	0.81	0.81

TABLE 13. Results of machine and deep learning models with respect to negative class.

Features	Model	Accuracy	CP	WP
TF-IDF	DT	0.55	7,454	5,927
	RF	0.87	11,724	1,657
	KNN	0.05	789	12,592
	SVM	0.96	12,899	482
	LR	0.95	12,751	630
BoW	DT	0.55	7,427	5,956
	RF	0.87	11,647	1,736
	KNN	0.28	3,868	9,515
	SVM	0.96	12,929	484
	LR	0.95	12,733	650
NN	LSTM	0.96	12,955	470
	GRU	0.88	11,745	1680
	CNN	0.96	12,921	504
	RNN	0.94	12,608	817
	GCR-NN	0.97	13,073	352

The collected dataset contains a total of 169,999 tweets including 66579, 49887, and 53533 tweets for neutral, positive, and negative tweets, respectively. Tweets containing negative sentiments make 31.49% of the total tweets which is definitely not a small number. Results in Table 13 are provided with respect to 53533 negative tweets. Results indicate that SVM shows the capability of detecting negative tweets with the highest accuracy of 0.96, both for TF-IDF and BoW features which means that 4% of racist tweets are misclassified by SVM. Similarly, LR correctly identifies 95% of the racist techniques but attributes 5% of the racist tweets to non-racist tweets. For racism detection, the performance of the proposed GCR-NN is superior to all models where only 352 of the 13425 racist tweets are misclassified which makes the racism detection accuracy of 0.97. This performance is superior to both machine learning, as well as, deep learning models.

V. CONCLUSION

Racist comments are becoming more frequent on social media platforms like Twitter and should be automatically detected and stopped to avoid further spread. This study considers racism detection from a sentiment analysis perspective and detects racist containing tweets by identifying negative sentiments. For obtaining high-performance sentiment analysis, deep learning is complemented by the ensemble approach where GRU, CNN, and RNN are stacked to form

the GCR-NN model. A large dataset collected from Twitter and annotated using the TextBlob is used for experiments with several machine learning, deep learning, and proposed GCR-NN model. Overall, 31.49% of the collected 169,999 tweets contain racist comments. Results show that deep learning models show substantially better performance than those of machine learning models with the proposed GCR-NN obtaining averaged 0.98 accuracy score regarding the sentiment analysis for positive, negative, and neutral classes. Since the negative class is important to detect racism, a separate analysis indicates that SVM and LR are able to detect 96% and 95%, respectively of racist tweets correctly while 4% and 5% of the racist tweets are misclassified, respectively. The proposed GCR-NN, on the other hand, can correctly detect 97% of the racist tweets with only a 3% misclassification rate.

REFERENCES

[1] K. R. Kaiser, D. M. Kaiser, R. M. Kaiser, and A. M. Rackham, "Using social media to understand and guide the treatment of racist ideology," *Global J. Guid. Counseling Schools, Current Perspect.*, vol. 8, no. 1, pp. 38–49, Apr. 2018.

[2] A. Perrin and M. Anderson. (2018). *Share of U.S. Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018*. [Online]. Available: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

[3] M. Ahlgren. *40C Twitter Statistics & Facts*. Accessed: Sep. 1, 2021. [Online]. Available: https://www.websitehostingrating.com/twitter-statistics/

[4] D. Arigo, S. Pagoto, L. Carter-Harris, S. E. Lillie, and C. Nebeker, "Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery," *Digit. Health*, vol. 4, Jan. 2018, Art. no. 205520761877175.

[5] A.-M. Bliuc, N. Faulkner, A. Jakubowicz, and C. McGarty, "Online networks of racial hate: A systematic review of 10 years of research on cyber-racism," *Comput. Hum. Behav.*, vol. 87, pp. 75–86, Oct. 2018.

[6] M. A. Price, J. R. Weisz, S. McKetta, N. L. Hollinsaid, M. R. Lattanner, A. E. Reid, and M. L. Hatzenbuehler, "Meta-analysis: Are psychotherapies less effective for black youth in communities with higher levels of anti-black racism?" *J. Amer. Acad. Child Adolescent Psychiatry*, 2021, doi: 10.1016/j.jaac.2021.07.808. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0890856721012818

[7] D. Williams and L. Cooper, "Reducing racial inequities in health: Using what we already know to take action," *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, p. 606, Feb. 2019.

[8] Y. Paradies, J. Ben, N. Denson, A. Elias, N. Priest, A. Pieterse, A. Gupta, M. Kelaher, and G. Gee, "Racism as a determinant of health: A systematic review and meta-analysis," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0138511.

[9] J. C. Phelan and B. G. Link, "Is racism a fundamental cause of inequalities in health?" *Annu. Rev. Sociol.*, vol. 41, no. 1, pp. 311–330, Aug. 2015.

[10] D. R. Williams, "Race and health: Basic questions, emerging directions," *Ann. Epidemiol.*, vol. 7, no. 5, pp. 322–333, Jul. 1997.

[11] Z. D. Bailey, N. Krieger, M. Agénor, J. Graves, N. Linos, and M. T. Bassett, "Structural racism and health inequities in the USA: Evidence and interventions," *Lancet*, vol. 389, no. 10077, pp. 1453–1463, Apr. 2017.

[12] D. R. Williams, J. A. Lawrence, B. A. Davis, and C. Vu, "Understanding how discrimination can affect health," *Health Services Res.*, vol. 54, no. S2, pp. 1374–1388, Dec. 2019.

[13] C. P. Jones, "Levels of racism: A theoretic framework and a gardener's tale," *Amer. J. Public Health*, vol. 90, no. 8, p. 1212, 2000.

[14] S. Forrester, D. Jacobs, R. Zmora, P. Schreiner, V. Roger, and C. I. Kiefe, "Racial differences in weathering and its associations with psychosocial stress: The CARDIA study," *SSM-Population Health*, vol. 7, Apr. 2019, Art. no. 100319.

[15] B. J. Goosby, J. E. Cheadle, and C. Mitchell, "Stress-related biosocial mechanisms of discrimination and African American health inequities," *Annu. Rev. Sociol.*, vol. 44, no. 1, pp. 319–340, Jul. 2018.

[16] A. T. Geronimus, M. Hicken, D. Keene, and J. Bound, "'Weathering' and age patterns of allostatic load scores among blacks and whites in the United States," *Amer. J. Public Health*, vol. 96, no. 5, pp. 826–833, 2006.

[17] Z. Papacharissi, "Affective publics and structures of storytelling: Sentiment, events and mediality," *Inf., Commun. Soc.*, vol. 19, no. 3, pp. 307–324, Mar. 2016.

[18] G. Bouvier, "How journalists source trending social media feeds: A critical discourse perspective on Twitter," *Journalism Stud.*, vol. 20, no. 2, pp. 212–231, Jan. 2019.

[19] M. KhosraviNik, "Social media critical discourse studies (SM-CDS)," in *The Routledge Handbook of Critical Discourse Studies*. London, U.K.: Routledge, 2017, pp. 582–596.

[20] T. T. Nguyen, S. Criss, A. M. Allen, M. M. Glymour, L. Phan, R. Trevino, S. Dasari, and Q. C. Nguyen, "Pride, love, and Twitter rants: Combining machine learning and qualitative techniques to understand what our tweets reveal about race in the US," *Int. J. Environ. Res. Public Health*, vol. 16, no. 10, p. 1766, May 2019.

[21] D. T. Goldberg, *Are we all Postracial Yet?*. Hoboken, NJ, USA: Wiley, 2015.

[22] I. Ashraf, S. Hur, and Y. Park, "Application of deep convolutional neural networks and smartphone sensors for indoor localization," *Appl. Sci.*, vol. 9, no. 11, p. 2337, 2019.

[23] M. Umer, I. Ashraf, S. Ullah, A. Mehmood, and G. S. Choi, "COVINet: A convolutional neural network approach for predicting COVID-19 from chest X-ray images," *J. Ambient Intell. Humanized Comput.*, pp. 1–13, Jan. 2021. [Online]. Available: https://doi.org/10.1007/s12652-021-02917-3

[24] M. Khalid, I. Ashraf, A. Mehmood, S. Ullah, M. Ahmad, and G. S. Choi, "GBSVM: Sentiment classification from unstructured reviews using ensemble classifier," *Appl. Sci.*, vol. 10, no. 8, p. 2788, Apr. 2020.

[25] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, "Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model," *IEEE Access*, vol. 9, pp. 78621–78634, 2021.

[26] S. Ranjan and S. Sood, "Social network investor sentiments for predicting stock price trends," *Int. J. Sci. Res. Rev.*, vol. 7, no. 2, pp. 90–97, 2019.

[27] M. S. Neethu and R. Rajasree, "Sentiment analysis in Twitter using machine learning techniques," in *Proc. 4th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2013, pp. 1–5.

[28] K. Perifanos and D. Goutsos, "Multimodal hate speech detection in Greek social media," *Multimodal Technol. Interact.*, vol. 5, no. 7, p. 34, 2021.

[29] I. Aljarah, M. Habib, N. Hijazi, H. Faris, R. Qaddoura, B. Hammo, M. Abushariah, and M. Alfawareh, "Intelligent detection of hate speech in arabic social network: A machine learning approach," *J. Inf. Sci.*, vol. 47, no. 3, May 2020, Art. no. 0165551520917651.

[30] S. Goswami, M. Hudnurkar, and S. Ambekar, "Fake news and hate speech detection with machine learning and NLP," *PalArch's J. Archaeol. Egypt/Egyptol.*, vol. 17, no. 6, pp. 4309–4322, 2020.

[31] R. Alshalan and H. Al-Khalifa, "A deep learning approach for automatic hate speech detection in the Saudi Twittersphere," *Appl. Sci.*, vol. 10, no. 23, p. 8614, Dec. 2020.

[32] L. Ketsbaia, B. Issac, and X. Chen, "Detection of hate tweets using machine learning and deep learning," in *Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2020, pp. 751–758.

[33] T. Putri, S. Sriadhi, R. Sari, R. Rahmadani, and H. Hutahaean, "A comparison of classification algorithms for hate speech detection," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 830, Apr. 2020, Art. no. 032006.

[34] U. Bhandary, "Detection of hate speech in videos using machine learning," M.S. thesis, Dept. Comput. Sci., San Jose State Univ., San Jose, CA, USA, 2019.

[35] B. Vidgen and T. Yasseri, "Detecting weak and strong Islamophobic hate speech on social media," *J. Inf. Technol. Politics*, vol. 17, no. 1, pp. 66–78, Jan. 2020.

[36] J. Salminen, M. Hopf, S. A. Chowdhury, S.-G. Jung, H. Almerekhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Hum.-Centric Comput. Inf. Sci.*, vol. 10, no. 1, pp. 1–34, Dec. 2020.

[37] M. A. Al-garadi, M. R. Hussain, N. Khan, G. Murtaza, H. F. Nweke, I. Ali, G. Mujtaba, H. Chiroma, H. A. Khattak, and A. Gani, "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.

[38] Q. Al-Maatouk, M. S. Othman, A. Aldraiweesh, U. Alturki, W. M. Al-Rahmi, and A. A. Aljeraiwi, "Task-technology fit and technology acceptance model application to structure and evaluate the adoption of social media in academia," *IEEE Access*, vol. 8, pp. 78427–78440, 2020.
[39] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in social networks: A survey on multilingual corpus," in *Proc. 6th Int. Conf. Comput. Sci. Inf. Technol.*, vol. 10, 2019, pp. 1–19.
[40] A. Schmidt and M. Wiegand, "A survey on hate speech detection using natural language processing," in *Proc. 5th Int. Workshop Natural Lang. Process. Social Media*, 2017, pp. 1–10.
[41] A. Alrehili, "Automatic hate speech detection on social media: A brief survey," in *Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2019, pp. 1–6.
[42] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, "Tweets classification on the base of sentiments for US airline companies," *Entropy*, vol. 21, no. 11, p. 1078, Nov. 2019.
[43] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis," *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0245909.
[44] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. A. Reshi, and I. Ashraf, "Sentiment analysis and topic modeling on tweets about online education during COVID-19," *Appl. Sci.*, vol. 11, no. 18, p. 8438, Sep. 2021.
[45] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, "Classification of Shopify app user reviews using novel multi text features," *IEEE Access*, vol. 8, pp. 30234–30244, 2020.
[46] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
[47] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Trans. Geosci. Electron.*, vol. GE-15, no. 3, pp. 142–147, Jul. 1977.
[48] V. Rupapara, F. Rustam, A. Amaar, P. B. Washington, E. Lee, and I. Ashraf, "Deepfake tweets classification using stacked bi-LSTM and words embedding," *PeerJ Comput. Sci.*, vol. 7, p. e745, Oct. 2021.
[49] B. Devi, V. G. Shankar, S. Srivastava, K. Nigam, and L. Narang, "Racist tweets-based sentiment analysis using individual and ensemble classifiers," in *Micro-Electronics and Telecommunication Engineering*. Singapore: Springer, 2021, pp. 555–567.
[50] M. T. H. K. Tusar and M. T. Islam, "A comparative study of sentiment analysis using NLP and different machine learning techniques on US airline Twitter data," 2021, *arXiv:2110.00859*.

ERNESTO LEE is currently working as a Professor with the Department of Computer Science, Broward College, Broward County, Florida, USA. His research interests include block-chain, the IoT, and data mining, mainly working on machine learning and deep learning-based IoT and text mining tasks.

FURQAN RUSTAM received the master's degree in computer science from the Islamia University of Bahawalpur, in 2017, and the M.S. degree from the Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2020. He served as a Research Assistant and a Lecturer for the Fareed Computing and Research Center, KFUEIT. His current research interests include data mining, mainly working on machine learning and deep learning-based IoT, text mining, and bioinformatics tasks.

PATRICK BERNARD WASHINGTON is currently working as a Professor with the Division of Business Administration and Economics, Morehouse College, Atlanta, GA, USA. His research interests include block-chain, the IoT, and data mining, mainly working on machine learning and deep learning-based IoT and text mining tasks.

FATIMA EL BARAKAZ received the Engineering Diploma degree in computer science engineering from the National Institute of Statistics and Applied Economics, Morocco, in June 2015. She is currently pursuing the Ph.D. degree in machine learning/data mining with the LAROSERI Laboratory, Department of Computer Science, Chouaib Doukkali University El Jadida, Morocco. She is also a temporary Teacher in C/C++ classrooms. Her research interests include data mining, mainly working on machine learning and deep learning and text mining tasks.

WAJDI ALJEDAANI received the bachelor's degree in software engineering from the Athlone Institute of Technology, Ireland, in 2014, and the master's degree in software engineering from the Rochester Institute of Technology, New York, in 2016. He is currently pursuing the Ph.D. degree in computer science and engineering with the University of North Texas. He worked as a Lecturer at the Al-Khari College of Technology, Saudi Arabia, from 2017 to 2020. His research interests include software engineering, mining software repository, accessibility, machine learning, and text mining.

IMRAN ASHRAF received the M.S. degree in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, Gyeongsan, South Korea, in 2018. He has worked as a Postdoctoral Fellow at Yeungnam University. He is currently working as an Assistant Professor with the Information and Communication Engineering Department, Yeungnam University. His research interests include indoor positioning and localization, indoor location-based services in wireless communication, smart sensors (LIDAR) for smart cars, and data mining.