A Deep Learning Approach for Sentiment Classification of COVID-19 Vaccination Tweets

Haidi Said[1], BenBella S. Tawfik[2], Mohamed A. Makhlouf[3]
Department of Information Systems-Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt[1, 2, 3]

***Abstract*—Now-a-days, social media platforms enable people to continuously express their opinions and thoughts about different topics. Monitoring and analyzing the sentiments of people is essential for governments and business organizations to better understand people's feelings and thoughts. The Coronavirus disease 2019 (COVID-19) has been one of the most trending topics on social media over the last two years. Consequently, one of the preventative measures to control and prevent the spread of the virus was vaccination. A dataset was formed by collecting tweets from Twitter for over a month from November 13th to December 31st, 2021. After data cleaning, the tweets were assigned a positive, negative, or neutral label using a natural language processing (NLP) sentiment analysis tool. This study aims to analyze people's public opinion towards the vaccination process against COVID-19. To fulfil this goal, an ensemble model based on deep learning (LSTM-2BiGRU) is proposed that combines long short-term memory (LSTM) and bidirectional gated recurrent unit (BiGRU). The performance of the proposed model is compared to five traditional machine learning models, two deep learning models in addition to state-of-the-art models. By comparing the results of the models used in this study, the results reveal that the proposed model outperforms all the machine and deep learning models employed in this work with a 92.46% accuracy score. This study also shows that the number of tweets that involve neutral, positive, and negative sentiments is 517496 (37%) tweets, 484258 (34%) tweets, and 409570 (29%) tweets, respectively. The findings indicate that the number of people carrying neutral sentiments towards COVID-19 immunization through vaccines is the highest among others.**

Keywords*—COVID-19 vaccination; sentiment analysis; Twitter; machine learning; deep learning; natural language processing (NLP)

I. Introduction

Recently, the massive amount of data generated by users on social media platforms and advances in computational power provide a strong impetus for sentiment analysis to develop and become the top research field under natural language processing (NLP) [1]. Sentiment analysis (SA), also known as sentiment or opinion mining, is the computational analysis of people's opinions and emotions about different topics [2]. This automatic analysis plays a significant role in extracting useful insights for decision-makers in various application domains. SA is a research field that aspires to understand and extract the sentiment of unstructured content. This content can be text, audio, image, or video. Following the technical definition of sentiment analysis introduced by [3], in this sentence, "Mark likes the camera of Samsung S10", here Mark acts as the opinion holder expressing his positive sentiment towards the aspect camera of entity Samsung S10. SA uses NLP, and machine learning (ML) techniques to extract the subjective information of a document and classify it according to its opinion orientation or polarity which can be positive, negative, or neutral. Subjective information carries only sentiments about a particular topic or entity, while objective text contains facts with no sentiments [4]. For instance, "Star Wars is an awesome movie". This sentence has a sentiment (awesome); thus, it is a subjective text, not an objective.

The explosive growth of social media applications (e.g., Facebook, Instagram, Twitter) witnessed today opens the door for a continuous stream of opinions and thoughts [5]. This Internet-generated content enables people to communicate with each other by sharing their sentiments and how they feel in the form of opinions or reviews about any topic. This consequently leads to the generation of an enormous amount of unstructured data. Business organizations need to process and analyze this data to support them in their decision making by gaining deeper insights into user sentiments, which will accordingly improve customer satisfaction. Furthermore, the study of public opinion about a specific topic or issue is very important to governments for the study of human activity and behavior as well as crisis management. Sentiment analysis provides governments with the valuable information necessary to take true actions in time.

Monitoring and analyzing the opinions and sentiments for extracting valuable information from them manually is a challenging task, in terms of time and exerted effort, due to the boom of various internet-based applications and sites. So, automated sentiment analysis or opinion mining systems are required to overcome this problem. Several machine learning and natural language processing-based approaches have been proposed to analyze these sentiments. However, recently deep learning-based approaches have shown significant results and higher performance in the field of sentiment analysis [6].

COVID-19 has dramatically affected every aspect of our lives and was declared a worldwide pandemic by the World Health Organization (WHO) [7]. Vaccination against COVID-19 was introduced as an urgent solution to prevent the spread of the disease among people and reduce death rates [8]. Therefore, analyzing the sentiments of people towards COVID-19 vaccination is essential for governments and health ministries to understand the public mood of people regarding the vaccination process against that virus. Consequently, vaccine campaigns can be developed targeting the hesitant and anti-vaccine groups of people to raise their awareness about the pivotal role of vaccines and boosters in containing the pandemic.

A long short-term memory (LSTM) is an extension of recurrent neural network (RNN). LSTMs can handle long-term dependencies between time steps of sequential data [9]. A Bidirectional gated recurrent unit (BiGRU) is used to learn features in both directions to produce a more meaningful output. This study combines the strengths of LSTM and BiGRU models to enhance the accuracy of sentiment classification. To the best of our knowledge, several studies navigated the public views regarding the COVID-19 vaccines before November 2021 [10-13]. However, many important events related to COVID-19 vaccines have occurred after this date, like the emergence of the Omicron variant [14] as well as boosters. Therefore, to fill this gap, we analyzed the tweets related to COVID-19 vaccination during this period of time to fully understand the public mood and perspectives towards COVID-19 vaccination acceptance or hesitancy. The main contributions of this study are as follows:

- Creating and preparing a large dataset of 4,093,986 tweets related to COVID-19 vaccination and vaccines for the sentiment classification task.
- Visualizing and analyzing the sentiments of people about the topic of COVID-19 vaccination.
- Proposing a hybrid approach to determine people's thoughts and opinions towards COVID-19 vaccination. The collected dataset was labeled using Valence Aware Dictionary and sEntiment Reasoner (VADER) lexicon-based approach. Moreover, Conventional machine learning classifiers were trained on the annotated dataset and tested on unseen data to evaluate their performance.
- Improving the classification accuracy by introducing an ensemble model (LSTM-2BiGRU), combining long short-term memory and bidirectional gated recurrent unit.
- A performance comparison of the proposed model with traditional machine learning classifiers, two individual deep learning models (LSTM, BiGRU), and state-of-the-art models to identify the top performer model on our collected dataset.

The rest of the paper is organized as follows: Section II provides an overview of the related work to our study. The subsequent section describes the dataset and methods used in the study. The fourth section presents and discusses the results of the conducted experiments. Finally, the last section concludes and summarizes the paper.

II. Related Work

COVID-19 has seriously affected the daily lives of people in different sectors of life [15]. As a result, several studies were conducted to analyze people's sentiments towards COVID-19 and its related aspects. For example, the study in [16] focused on analyzing and understanding the sentiments of the Canadian people towards social distancing related to COVID-19 using a hybrid approach. They used SentiStrength, a lexicon-based approach, with Support Vector Machine (SVM), a machine learning approach, for sentiment analysis. To perform this study, they collected Twitter data for one month from an open-source publicly available IEEE website. They applied the SVM algorithm by splitting the dataset into 80% for training and 20% for testing the data. Their experiment resulted in 71% accuracy for sentiment classification. By excluding the neutral sentiments from data, the accuracy of the model rose to 81%. Additionally, the accuracy reached 87% by reducing the test data to 10%. The results showed that 40% of the Canadian people have neutral sentiments, followed by 35% expressed negative sentiments and only 25% have positive sentiments towards social distancing. One of the limitations they faced during the study was the insufficient number of training examples.

The authors in [17] compared the effectiveness of two machine learning models, Naïve Bayes (NB), and logistic regression (LR), on Twitter data of varying lengths collected from February to March of 2020. The collected data was cleaned and prepared for analysis using the R programming language and its related packages. NB achieved 91% accuracy with short tweets (less than 77 characters), while the accuracy obtained using the logistic regression model was 74% with short tweets. However, the performance of both models (Naïve Bayes and logistic regression) significantly decreased to 57% and 52%, respectively, in the case of longer tweets. For future work, the authors of this work recommended extending this study with more data and additional methods. This will consequently help policy makers and businesses in understanding the public sentiment and perspectives for making the right decisions in time.

The study in [18] performed a sentiment classification task on 3090 tweets collected from 23 March to 15 July 2020. The collected text was labeled as fear, sad, anger, and joy. Each sentiment was mapped from 0 to 3 (fear:0, sad:1, anger:2, joy:3). They applied a new deep learning model called Bidirectional Encoder Representations from Transformers (BERT). They compared it to three other classification algorithms: LR, SVM, and LSTM. For all the classifiers, they split the dataset into 85% for training and the remaining 15% was used for testing purposes. Their experiment showed that the BERT model outperformed the other models by having 89% accuracy, whereas the other three models scored 75%, 74.775%, and 65%, respectively.

In [19], logistic regression and LSTM learning models were used along with two different word embedding techniques: CountVectorizer and TfidfVectorizer to analyze and study people's attitude towards COVID-19. The dataset used in this study consists of tweets collected from 03/16/2020 to 04/14/2020. The author used only original tweet and label columns from the given columns in the dataset to investigate people's reactions towards COVID-19 during this period of time. The data was initially labeled as five classes (positive, extremely positive, negative, extremely negative, and neutral) but the author manually grouped them into three classes (positive, negative, and neutral) for more precise classification. Then in the featurization stage, the original tweets were pre-processed and vectorized using the previously mentioned two vectorization methods. After that, logistic regression and LSTM models were used to classify the sentiments of the given COVID-19 related tweets. The LSTM/TF-IDF model achieved the best result with an F1 score of 0.85.

The authors in [20] used the keyword "COVID-19" to search tweets on Weibo, one of the largest Chinese social media platforms, and collected 45,987 tweets. They used a lexicon-based approach for sentiment analysis by matching their dataset with words in the sentiment dictionary. BosonNLP sentiment dictionary was selected in this study as it supports the Chinese language and is constructed from millions of labeled data, including Weibo, news, forum data, and others. For future work, they suggest combining their approach with deep learning (DL) or random forest (RF) classifier to achieve more accurate sentiment prediction results. Kuar et al. devised an algorithm called Hybrid Heterogeneous Support Vector Machine (H-SVM) that performs sentiment classification on COVID-19 tweets in study [21]. They compared the performance of their proposed algorithm with RNN and SVM. The H-SVM model outperformed the other two models in terms of precision (86%), recall (69%), and F1 score (77%).

In [22], the authors proposed a feature extraction technique by concatenating bag-of-words (BoW) and term frequency – inverse document frequency (TF-IDF). They used five machine learning models, such as RF, SVM, decision tree (DT), XGBoost classifier, extra tree classifier (ETC), and LSTM to analyze sentiments of COVID-19 tweets. The results showed that ETC outperformed all other models using their proposed feature extraction technique with an accuracy score of 0.93. The authors stated that the poor performance of the LSTM deep learning model is due to the small dataset. The authors of paper [23] conducted sentiment analysis on 11,960 tweets in the UK related to COVID-19 using three different ensemble models. The dataset was labeled manually by three independent annotators. The stacking classifier (SC) achieved the highest F1 score with 83.5% in comparison with the voting classifier (VC) and bagging classifier (BC) with 83.3% and 83.2% scores, respectively.

Researchers in this study [24] applied sentiment analysis to tweets about COVID-19 different vaccines. They found that 48.49% were neutral, 33.96% were positive, and 17.55% had negative feelings about the coronavirus vaccines. The study used LSTM and bidirectional LSTM (BiLSTM) deep learning models to predict the sentiments of the tweets. According to the study, the resulting accuracy is 90.59% for LSTM and 90.83% for Bi-LSTM. The authors of paper [25] combined convolutional neural network (CNN) with BiLSTM to classify the polarity of COVID-19 tweets as positive, negative, or neutral. They used the hybrid CNN-BiLSTM model with two word embedding pre-trained models: FastText and Global Vectors for Word Representation (GloVe) to obtain higher accuracy levels. The authors conducted experiments on a dataset of 40000 tweets collected from Twitter. The CNN-BiLSTM model with FastText yielded 99.33% accuracy, which is higher than the accuracy score of CNN-BiLSTM with GloVe (97.55%).

The authors in [26] proposed a hybrid feature extraction approach for COVID-19 tweets sentiment classification. They combined TF-IDF with FastText and GloVe to increase the performance of classification. This study applied seven machine learning classifiers in addition to one deep learning model (CNN). It was observed that the best performance was achieved by SVM using the TF-IDF with FastText word embedding proposed technique. This study [27] presented a BiLSTM model for public opinion analysis of COVID-19-related discussions on social media. They performed sentiment classification using four different scenarios on three datasets collected from Twitter and Reddit platforms. One of the limitations they encountered is that the BiLSTM method is time-consuming. Also, they recommended applying multilingual sentiment analysis to get a global understanding of COVID-19 sentiments.

A new hybrid deep learning method is proposed in [28] to find the general sentiment of people in eight countries around the world. Their method is based on the fusion of four deep learning and one machine learning model. Their proposed model shows superior performance over the other models used in the study with 85.8% accuracy. In [29], the authors performed sentiment analysis on Nepali COVID-19 tweets using an ensemble CNN model. They proposed three different CNN models with three feature extraction techniques such as FastText (ft), domain- specific (ds), and domain-agnostic (da). In addition, they combined the three CNN models to form an ensemble CNN model. Their model achieved an accuracy of 68.7% for sentiment classification of tweets in the Nepali language.

This research [30] used three machine learning classifiers to predict people's awareness of COVID-19 preventative measures in Saudi Arabia. For this purpose, they prepared a dataset of Arabic tweets related to COVID-19 preventative procedures. They applied four feature extraction techniques (unigram, unigram TF-IDF, bigram, and bigram TF-IDF) with each classifier. Their results show that the SVM classifier with bigram TF-IDF achieved the highest accuracy of 85%. Furthermore, the people in the south region of Saudi Arabia had the highest level of awareness, where they reacted positively towards COVID-19-related precautionary measures. The sentiments of COVID-19 tweets were analyzed using the BERT model in research [31]. They created two datasets by collecting tweets from India and the whole world from January 20, 2020, to April 25, 2020. The accuracy of their proposed model is approximately 94%.

III. Materials and Methods

This study introduced an ensemble model based on a hybrid approach to analyze the sentiments of people towards COVID-19 vaccination. We collected a large dataset containing tweets related specifically to the COVID-19 vaccines and vaccination process. Tweets were cleaned and pre-processed before feeding them to the machine and deep learning models. VADER, a lexicon-based approach was used to assign positive, negative, or neutral polarity to the collected tweets. Afterwards, the annotated dataset was split for training and testing purposes. The lexicon-based approach is integrated with machine learning and deep learning-based approaches for sentiment classification. Accuracy, precision, recall, and F1 score were used to analyze the performance of the models. Fig. 1 depicts the architecture of the proposed methodology.

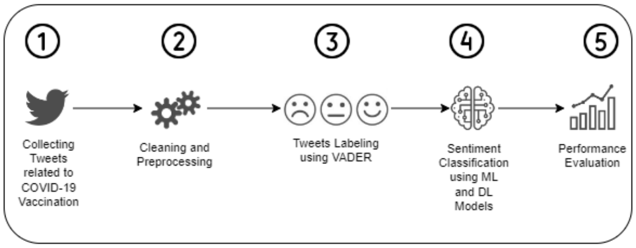

Fig. 1. Architecture of the proposed methodology.

A. Data Collection

To extract the COVID-19 vaccination-related tweets from Twitter, we applied for a Twitter developer account for the purpose of doing academic research. Accordingly, the Twitter developer team approved our request and enabled us to access Twitter's application programming interface (API), which acts as an intermediary between the developer and the Twitter backend. Ultimately, the Twitter developer team provided us with the required credentials to authenticate access to the Twitter API. We used the Tweepy library in Python to interact with the Twitter API for collecting the required tweets. The data collection process was carried out on a daily basis from 13 November 2021 to 31 December 2021. Different hashtags were used to extract the tweets related to COVID-19 vaccines such as "#vaccinated", "#CovidVaccine", "#CovidVaccination", "#VaccinesWork", "#VaccineSideEffects", "#booster", "#AstraZeneca", "#Pfizer", and "#Moderna". In total, 4,093,986 tweets were extracted about vaccination against the coronavirus disease. The collected dataset shape is 4093986 by 5 columns, including tweet_id, text, favorite_count, created_at, and tweet_location. Table I shows a sample of some tweets from the collected dataset.

TABLE I. A SAMPLE OF TWEETS RELATED TO COVID-19 VACCINATION

Text
IM VACCINATED 😘
I just got vaccinated on Tuesday and it hurts :(
@Dannyoconnor430 Get vaccinated to protect yourself.
So no need to panic ! Get vaccinated !! Protect yourself and others 😇 #omicron #covid #Valimai https://t.co/PAbdvkAGuA
@VishnuNDTV really a ridiculous rule for domestic passenger who r fully vaccinated.
Majority of mumps cases are among the vaccinated, CDC finds https://t.co/WuWIbWy3AY via @nbcnews
@ScottATaylor waiting for the non vaccinated to get wise, so I can go out and have fun again
Saw a kid go up and thank the medical professional who vaccinated him today and my heart remembered the world has good in it
@FlutoShinzawa Good thing they are all double vaccinated.
@ABC Do what you want with your own body...But get vaccinated or lose your job

B. Data Cleaning and Preprocessing

After data collection and exploration, tweet texts were cleaned by removing noise or words that are irrelevant to our experiment. While extracting data from Twitter, all retweets were ignored by applying -filter: retweets to our search query. In this study, we are interested only in analyzing the sentiments of tweets, so all the columns other than text were discarded. This led to having 4,093,986 rows and 1 column. In addition, duplicated tweets and rows with null values were removed. After these operations, the ultimate shape of the dataset became 1,411,324 rows by 1 column. Subsequently, irrelevant data elements to the sentiment analysis process were removed including hashtags (#), hyperlinks, mentions (@username), special characters, numbers, and punctuation marks. Table II shows some tweets before and after pre-processing operations. After labeling the tweets, we removed stop words and applied tokenization, case-folding, and stemming to avoid misleading the models.

TABLE II. A SAMPLE OF TWEETS BEFORE AND AFTER THE CLEANING PROCESS

Original raw tweets	Cleaned tweets
Please get vaccinated. https://t.co/YfZnFLRTSd	Please get vaccinated
Egypt bans unvaccinated people from public institutions#Egypt#vaccination#covid#COVID19#coronavirus…https://t.co/0rvWPbaL8P	Egypt bans unvaccinated people from public institutions

C. Data Labeling using a Lexicon-based Approach

VADER is a lexicon-based approach used for text sentiment analysis. It depends on a dictionary comprising nearly 7517 features (words) with corresponding valence scores [32]. VADER sentiment analyzer returns a tuple of (neg, neu, pos, compound) for each sentence. A compound score is computed for each text, ranging from -1 (most negative feeling) to 1 (most positive feeling). It is a normalized score calculated by adding the scores of each word present in the sentence. If the compound score is greater than or equal to 0.05 then the tweet sentiment is considered positive and if it is smaller than or equal to -0.05 then the sentiment of the tweet is negative, otherwise the tweet is regarded as neutral. Some tweets with their polarity orientation are shown in Table III.

TABLE III. A SAMPLE OF TWEETS WITH THEIR CORRESPONDING SENTIMENTS

Tweets	Polarity
I just got vaccinated on Tuesday and it hurts	Negative
The city is starting to enforce the vaccine requirement to enter most indoor public spaces You MUST be vaccinated	Neutral
Good thing they are all double vaccinated	Positive

D. Sentiment Classification using ML and DL Models

This study used traditional machine learning in addition to deep learning models to classify the sentiments of tweets. LR, NB, DT, RF, and K-Nearest Neighbors (KNN) classifiers are implemented using scikit-learn library in Python. For our sentiment analysis task, two individual deep learning models, LSTM and BiGRU, were used in addition to our proposed ensemble model (LSTM-2BiGRU) to achieve higher levels of accuracy. The proposed model combines LSTM and BiGRU to obtain better results. In LSTM-2BiGRU, the output of the LSTM is fed to a stack of two sequential BiGRU layers for tweets sentiment classification.

E. Performance Evaluation

We used precision, recall, accuracy, and F1-Score to assess the performance of the sentiment classification models. These performance metrics are calculated using the following formulas:

$$\text{Precision}= \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall}= \frac{TP}{TP+FN} \tag{2}$$

$$\text{Accuracy}= \frac{TP+TN}{TP+TN+ FP+FN} \tag{3}$$

$$\text{F1-score}=2* \frac{\text{Precision*Recall}}{\text{Precision+ Recall}} \tag{4}$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative instances respectively.

F. Architecture of the Proposed Model

The structure of the proposed model is composed of seven layers, as given in Fig. 2 below. The first layer of the proposed model is an embedding layer, where the number of unique words in the vocabulary is 5000. Each word is represented by a vector of 100 dimensions. A dropout layer with a 0.5 dropout rate follows the embedding layer. After the dropout layer, an LSTM layer is applied with 128 units. The output of the previous layer is fed to a stack of two BiGRU layers with 100 and 64 units for each layer, respectively. The output of the first BiGRU layer is processed by the subsequent BiGRU layer to improve the accuracy of the model. BiGRU traverses the input data two times through a forward and a backward pass. Lastly a dense layer is used with three neurons based on the number of target classes (positive, negative, or neutral) and a softmax activation function. Before the previous dense layer, a dropout layer is placed with a rate of 0.5. The dropout layers are employed to reduce the complexity of the model as well as its propensity for overfitting. The LSTM-2BiGRU model was trained using 21 epochs and a batch size of 128. Early stopping was used to prevent overfitting of the model. The hyperparameters used to train our proposed model are listed in Table IV below.

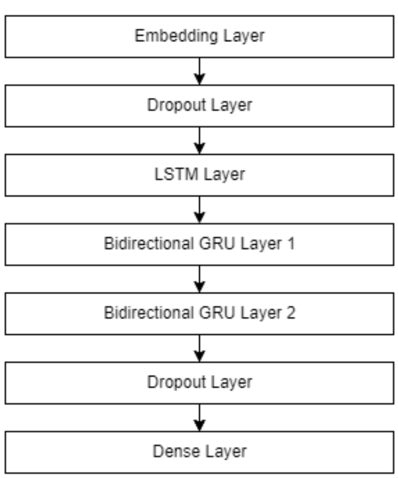

Fig. 2. Architecture of the proposed model.

TABLE IV. HYPERPARAMETERS USED FOR THE PROPOSED MODEL

Hyperparameter	Value
Optimizer	Adam
Loss	Categorical cross-entropy
Batch Size	128
Epoch	21
Dropout	0.5
Activation	Softmax

The pseudo-code below represents the steps taken to develop and evaluate the proposed model.

Algorithm 1: The pseudocode of the proposed model

Import Keras library for creating a deep learning model (M)
Read the collected dataset (D_i)
Clean the tweets from unwanted symbols like #, @, etc.
Label tweets using VADER tool
    If the compound score >= 0.05
        Return "Positive"
    Else If compound score <= -0.05
        Return "Negative"
    Else
        Return "Neutral"
    End If
Pre-process cleaned tweets by removing stop words, applying case-folding, stemming, and tokenization
Initialize hyperparameters by setting input_dim=5000, output_dim=100, batch_size=128, epochs= 21
For no_epochs and batch_size **do**
    Train the model (M)
End For
Predict the sentiment of unlabeled tweets using M
Evaluate the model and calculate the metrics accuracy, precision, recall, and F1-score
Plot the ROC curve of the model

IV. RESULTS AND DISCUSSIONS

This section demonstrates the results of sentiment analysis on COVID-19 vaccination tweets using machine learning and deep learning models. The collected dataset in this study is split into 85% for training the models and 15% for testing purpose on unseen and unlabeled data. The number of tweets in the training set is 1,199,625 while the testing set consists of about 211699 tweets. Fig. 3 shows a word cloud of the most frequent terms used in all the tweets. Fig. 4 to Fig. 6 present the key terms used in the positive, negative, and neutral tweets respectively.

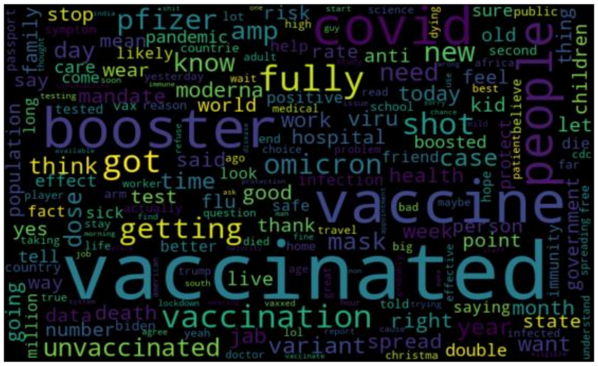

Fig. 3. Word cloud of all tweets.

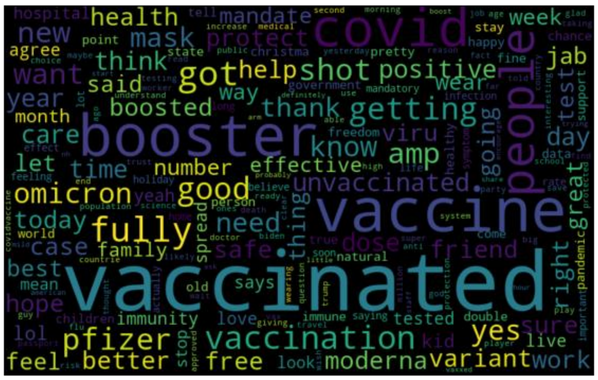

Fig. 4. Word cloud of positive tweets.

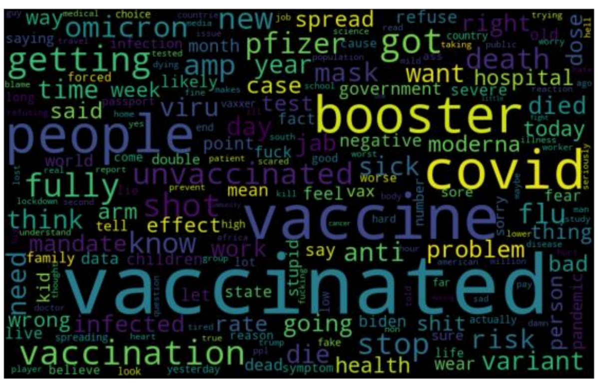

Fig. 5. Word cloud of negative tweets.

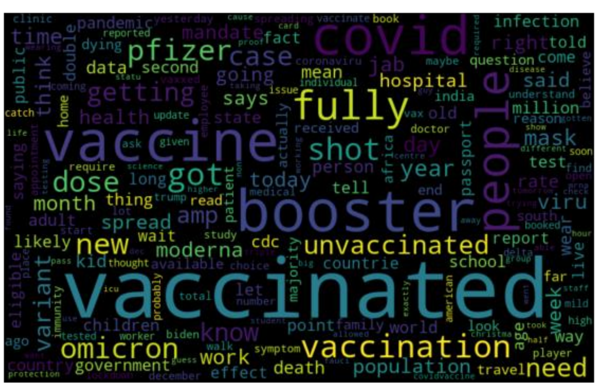

Fig. 6. Word cloud of neutral tweets.

A. Sentiment Analysis using VADER

Fig. 7 shows that among 1,411,324 tweets, 517496, 484258 and 409570 tweets were labeled with the help of VADER as neutral, positive, and negative tweets respectively. The percentage of neutral, positive, and negative tweets is 37%, 34%, and 29% respectively, as shown in Fig. 8 below. According to the VADER approach, most tweets belong to the neutral class, followed by the positive tweets, whereas the percentage of negative tweets is the least among them.

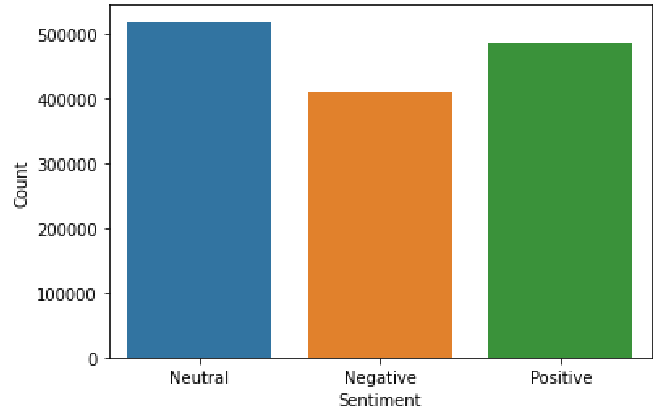

Fig. 7. The bar graph shows the total number of tweets in each class (sentiment).

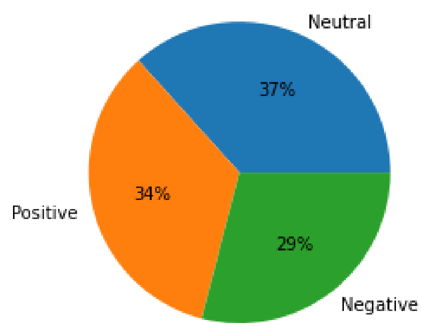

Fig. 8. The pie chart shows the proportion of sentiments present in the analyzed tweets.

B. Sentiment Analysis using Machine Learning Models

Table V shows the results of machine learning models used for sentiment analysis besides the lexicon-based approach, VADER. It can be noted that both RF and LR achieved the highest accuracy of about 87%. The accuracy of Decision Tree is nearly 83% followed by NB with an accuracy score of around 80%. On the other hand, KNN showed the lowest accuracy score, roughly 57%. To improve the performance of sentiment classification, a majority voting ensemble model was used which comprises the most accurate classifiers, including RF, LR, and DT. The ensemble model outperformed RF and LR models and recorded a test accuracy score of almost 88% as well as a superior performance in terms of precision, recall, and F1 score. Table VI lists the hyperparameters used for fine-tuning the machine learning models.

TABLE V. SENTIMENT ANALYSIS RESULTS USING MACHINE LEARNING CLASSIFIERS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	87	87	87	87
Logistic Regression	87	87	87	87
Decision Tree	83	83	83	83
Naïve Bayes	80	81	80	80
KNN	57	68	56	55
Voting Classifier	88	88	88	88

TABLE VI. HYPERPARAMETERS SELECTED FOR MACHINE LEARNING MODELS

Algorithm	Hyperparameter
Logistic Regression	solver = 'saga'
Multinomial Naïve Bayes	default setting
Decision Tree	criterion = 'gini'
Random Forest	n_estimators = 100
KNN	n_neighbors = 3, weights = 'uniform'

C. Sentiment Analysis using Deep Learning Models

To attain higher levels of accuracy for our sentiment analysis task, two individual deep learning models are used including LSTM and BiGRU as well as our proposed ensemble model. The results of the deep learning models are provided in Table VII. The proposed LSTM-2BiGRU model shows a superior performance over the traditional machine learning and deep learning models used in the study. It achieved the highest accuracy score of 92.46%, followed by LSTM and BiGRU with accuracy scores of about 91%. It is important to mention that using an ensemble of LSTM and BiGRU improved the classification accuracy than using each model individually. Fig. 9 shows the receiver operating characteristic (ROC) curve for the proposed model. The micro-average value of the area under the curve (AUC) for the LSTM-2BiGRU model is 0.98.

TABLE VII. SENTIMENT ANALYSIS RESULTS USING DEEP LEARNING MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LSTM	91.98	92.70	91.28	91.97
BiGRU	91.97	92.59	91.43	92.00
LSTM-2BiGRU	92.46	92.97	91.94	92.44

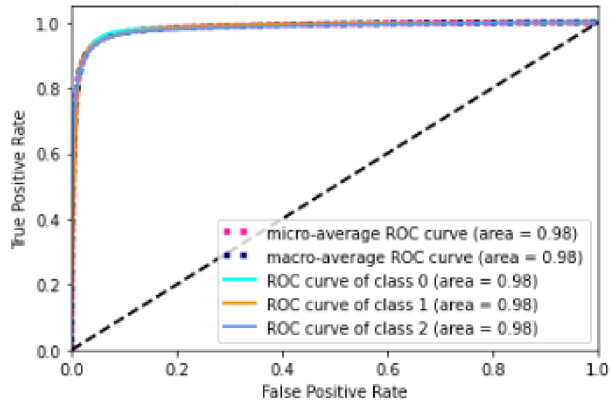

Fig. 9. ROC curve for the proposed model.

D. Comparison with State-of-the-Art Studies

To assess the performance of the proposed LSTM-2BiGRU model with respect to recent studies, a performance comparison was conducted on four cutting-edge studies related to sentiment analysis. These experiments were performed using the collected dataset in this study. In addition, all the steps involved in the proposed approach were applied to the models in the selected studies to guarantee a fair and accurate comparison. The study [33] introduced a stacked BiLSTM model to determine the polarity of deep fake tweets. The authors of paper [34] performed sentiment analysis of tweets related to COVID-19 vaccination using LSTM model. Likely, Reshi et al. presented an ensemble model called LSTM-GRNN for carrying out the same task in research [35]. In addition, paper [36] combined LSTM and GRU to analyze sentiments and detect emotions of tweets related to cryptocurrency. Table VIII compares the result of our proposed model to the findings of cutting-edge studies relevant to our task. The output results show that the proposed model outperforms other models in recent studies with 0.92 accuracy.

TABLE VIII. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART STUDIES

Reference	Model	Accuracy (%)
[33]	Stacked Bi-LSTM	91.94
[34]	LSTM	91.67
[35]	LSTM-GRNN	90.99
[36]	LSTM-GRU	90.75
This study	LSTM-2BiGRU	92.46

V. Conclusion

This study performs sentiment analysis on tweets related to the COVID-19 vaccination. To analyze the sentiments of the tweets, this study uses a hybrid approach based on two techniques: natural language processing and machine learning. For conducting experiments, tweets were extracted from Twitter, and the dataset was annotated using VADER. Accordingly, the proportion of people having neither positive nor negative feelings towards the COVID-19 vaccines is the highest among others. Considering the machine learning side of the hybrid approach, the dataset was split into 85% for training and 15% for testing. The sentiments of the tweets were analyzed using five machine learning classifiers: LR, NB, DT, RF, and KNN. To obtain higher accuracy for sentiment classification, two individual deep learning models (LSTM, BiGRU) were used in addition to our proposed ensemble model LSTM-2BiGRU. The experimental results show that the LSTM-2BiGRU model performs significantly better than all the traditional machine learning and deep learning models used in this study. With a 92.46% accuracy score, the proposed model proves its superiority in predicting the sentiments of tweets that are related to the vaccination aspect of COVID-19. However, it is important to point out that BiGRU is slow and requires more time for training due to its complex nature. In future, we aim to perform multimodal sentiment analysis as social media users express their feelings and opinions in multimedia forms, such as videos or images, rather than text solely.

References

[1] R. Jagdale, V. Shirsath, and S. Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques: Proceeding of CISC 2017," 2019, pp. 639-647.

[2] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.

[3] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," in Mining Text Data, C. C. Aggarwal and C. Zhai Eds. Boston, MA: Springer US, 2012, pp. 415-463.

[4] H. Kaur, V. Mangat, and Nidhi, "A survey of sentiment analysis techniques," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 10-11 Feb. 2017 2017, pp. 921-925, doi: 10.1109/I-SMAC.2017.8058315.

[5] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," Electronics, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.

[6] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," Artificial Intelligence Review, 2021/03/03 2021, doi: 10.1007/s10462-021-09973-3.

[7] S. A. Lone and A. Ahmad, "COVID-19 pandemic – an African perspective," Emerging Microbes & Infections, vol. 9, no. 1, pp. 1300-1308, 2020/01/01 2020, doi: 10.1080/22221751.2020.1775132.

[8] Z. Mukandavire, F. Nyabadza, N. J. Malunguza, D. F. Cuadros, T. Shiri, and G. Musuka, "Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios," PloS one, vol. 15, no. 7, p. e0236003, 2020.

[9] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," Neural computation, vol. 31, no. 7, pp. 1235-1270, 2019.

[10] H. Yin, X. Song, S. Yang, and J. Li, "Sentiment analysis and topic modeling for COVID-19 vaccine discussions," World Wide Web, vol. 25, no. 3, pp. 1067-1083, 2022.

[11] Z. B. Nezhad and M. A. Deihimi, "Twitter sentiment analysis from Iran about COVID 19 vaccine," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 16, no. 1, p. 102367, 2022.

[12] E. T. Khalid, E. B. Talal, M. K. Faraj, and A. A. Yassin, "Sentiment analysis system for COVID-19 vaccinations using data of Twitter," Indonesian Journal of Electrical Engineering and Computer Science, vol. 26, no. 2, pp. 1156-1164, 2022.

[13] R. Marcec and R. Likic, "Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines," Postgraduate Medical Journal, vol. 98, no. 1161, pp. 544-550, 2021, doi: 10.1136/postgradmedj-2021-140685.

[14] Y. Fan, X. Li, L. Zhang, S. Wan, L. Zhang, and F. Zhou, "SARS-CoV-2 Omicron variant: recent progress and future perspectives," Signal transduction and targeted therapy, vol. 7, no. 1, p. 141, 2022.

[15] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, "Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes," Information, vol. 12, no. 5, p. 204, 2021.

[16] C. Shofiya and S. Abidi, "Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data," International Journal of Environmental Research and Public Health, vol. 18, no. 11, p. 5993, 2021.

[17] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," Information, vol. 11, no. 6, p. 314, 2020.

[18] N. Chintalapudi, G. Battineni, and F. Amenta, "Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models," (in eng), Infect Dis Rep, vol. 13, no. 2, pp. 329-339, Apr 1 2021, doi: 10.3390/idr13020032.

[19] X. Zhou, "Sentiment Analysis of COVID-19 Tweets," 2021.

[20] X. Yu, C. Zhong, D. Li, and W. Xu, "Sentiment analysis for news and social media in COVID-19," Proceedings of the 6th ACM SIGSPATIAL International Workshop on Emergency Management using GIS, 2020.

[21] H. Kaur, S. U. Ahsaan, B. Alankar, and V. Chang, "A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets," Information Systems Frontiers, vol. 23, no. 6, pp. 1417-1429, 2021.

[22] 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, p. e0245909, 2021.

[23] M. Rahman and M. N. Islam, "Exploring the performance of ensemble machine learning classifiers for sentiment analysis of covid-19 tweets," in Sentimental Analysis and Deep Learning: Springer, 2022, pp. 383-396.

[24] K. N. Alam et al., "Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data," Computational and Mathematical Methods in Medicine, vol. 2021.

[25] T. T. Mengistie and D. Kumar, "Deep Learning Based Sentiment Analysis On COVID-19 Public Reviews," in 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 13-16 April 2021 2021, pp. 444-449, doi: 10.1109/ICAIIC51459.2021.9415191.

[26] Y. Didi, A. Walha, and A. Wali, "COVID-19 Tweets Classification Based on a Hybrid Word Embedding Method," Big Data and Cognitive Computing, vol. 6, no. 2, p. 58, 2022.

[27] M. Arbane, R. Benlamri, Y. Brik, and A. D. Alahmar, "Social media-based COVID-19 sentiment classification model using Bi-LSTM," Expert Systems with Applications, vol. 212, p. 118710, 2023/02/01/ 2023, doi: https://doi.org/10.1016/j.eswa.2022.118710.

[28] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharrya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," Knowledge-Based Systems, vol. 228, p. 107242, 2021.

[29] C. Sitaula, A. Basnet, A. Mainali, and T. B. Shahi, "Deep learning-based methods for sentiment analysis on Nepali covid-19-related tweets," Computational Intelligence and Neuroscience, vol. 2021, 2021.

[30] S. S. Aljameel et al., "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent

COVID-19 outbreaks in Saudi Arabia," International journal of environmental research and public health, vol. 18, no. 1, p. 218, 2021.

[31] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," Social Network Analysis and Mining, vol. 11, no. 1, pp. 1-11, 2021.

[32] E. Saad et al., "Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums," IEEE Access, vol. 9, pp. 85721-85737, 2021.

[33] 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 Computer Science, vol. 7, p. e745, 2021.

[34] R. R. Aryal and A. Bhattarai, "Sentiment Analysis on Covid-19 Vaccination Tweets using Naïve Bayes and LSTM," Advances in Engineering and Technology: An International Journal, vol. 1, no. 1, pp. 57-70, 2021.

[35] A. A. Reshi et al., "COVID-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset," in Healthcare, 2022, vol. 10, no. 3: MDPI, p. 411.

[36] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," IEEE Access, vol. 10, pp. 39313-39324, 2022.