Application of bidirectional LSTM deep learning technique for sentiment analysis of COVID-19 tweets: post-COVID vaccination era

Oluwatobi Noah Akande[1]* , Morolake Oladayo Lawrence[1] and Peter Ogedebe[1]

*Correspondence:
oluwatobi.
akande@bazeuniversity.edu.ng

[1] Computer Science Department, Baze University, Abuja, Nigeria

Abstract

Background: Social media platforms, especially Twitter, have turned out to be a major source of data repositories. They have become a platform that citizens can use to voice their concerns about issues that affect them. Most importantly, during the COVID-19 era, the platform was greatly used by governments and health organizations to sensitize people about the safety guidelines that they must adhere to so as to remain safe during the pandemic. As expected, people also used Twitter and other social media platforms to voice their opinions about how governments are handling the COVID-19 pandemic outbreak. Governments and organizations could, therefore, use these social media as a feedback mechanism that can help them know the view of the citizens about their policies. This could help them in making informed decisions about their policies.

Aim: The aim of this paper is to explore the use of BiLSTM deep learning technique for sentiment analysis of COVID-19 tweets.

Methodology: The study retrieved 197,327 tweets from the Nigeria Twitter domain using #COVID or #COVID-19 hashtags as keywords. The dataset was retrieved within the 1st month of COVID-19 vaccination in Nigeria, i.e., March 15–June 15, 2021. BiLSTM deep learning technique was trained using 789,306 sentiment annotated tweets obtained from Kaggle Sentiment140 tweet datasets. The preprocessed case study tweets were then used to evaluate the proposed model. Also, a precision of 78.26% and a recall value of 78.27% were also obtained.

Results: With an accuracy of 78.29%, 98,545 (49.93%) positive sentiments and 98,782 negative sentiments (50.06%) were recorded. Also, a precision of 78.26% and a recall value of 78.27% were also obtained. However, the presence of outliers which are tweets not related to COVID but which used the hashtag was observed.

Conclusion: This study has revealed the strength of BiLSTM deep learning technique for sentiment analysis. The results obtained revealed an almost balanced sentiments toward the pandemic with 49.93% positive disposition to the pandemic as compared to 50.06% negative disposition. This showed affirmed the impact of COVID vaccine in dousing citizen's tension when it was made available for public use. However, the presence of outliers in the classified tweets could be a pointer to the reason why

aspect-based sentiment analysis could be preferred to sentence-based sentiment analysis.

Keywords: Deep learning, BiLSTM, Sentiment analysis, Twitter, COVID-19 tweets, COVID-19 vaccine

Introduction

Social media played a pivotal role in disseminating information about the novel coronavirus when the first reports of the disease became public in December 2019 [1]. As of January 2020, the outbreak of the pandemic became a threat to global health as millions of people contracted the virus with reports of casualties hitting headlines in the international communities. COVID-19 was announced as a global epidemic by the World Health Organization (WHO) on March 11, 2020, thus, causing alteration and halting of day-to-day activities throughout the world, as the government of each nation enacted lockdown, prescription, suspension and prohibition of flights, large gatherings, etc., in an effort to stop the deadly virus from spreading over the world. The consequences arising from these directives include remote working, virtual learning, increased communication via social media and e-mail, online transaction, etc., by industrial workers, employees of large companies, students, and small business owners [2]. Social media has been ingrained in our daily lives. It establishes a connection between people and the outside world. Social media allows us to share our lives in a private, comfortable, and self-directed manner [3]. People are more reliant on posts and tweets shared on social media sites Laor [4]. The explosive growth and invention of various social media platforms and microblogging sites, information were available by the number of online users and users who were active dissemination easier during the pandemic [5]. During the pandemic, social media platforms such as Twitter, Facebook, WhatsApp, Instagram, Telegram, and YouTube were notable sources of information as well as means of communication and discussion among groups of people. Among these social media platforms, Twitter has been the most frequently used to tweet on daily activities all over the world, with over 330 million users [6] and holds a huge amount of data on various topics making waves in the internet; hence, Khattak et al. [7] referred to Twitter as the gold mine for data collection. This is mainly because Twitter has an application programming interface (API) that permits registered developers to collect peoples' tweets and analyze them. Texts used on Twitter are uniquely referred to as tweets and are characterized by hashtags to depict related topics, events, and trends. An example is the #COVID-19 tweets which became the major subject of discussion during the COVID-19 pandemic. Though, tweets are majorly not more than 140 characters accommodating an average of 14–15 words, they are characterized by the presence of emoticons, slang languages, misspellings, and acronymize words. These are acceptable ways of conveying opinions, emotions, sentiments, and expressions. Most importantly, the role of Twitter in the surveillance and monitoring of global crises cannot be underemphasized. It has played a great role in the monitoring of earlier epidemiological diseases such as respiratory syndrome and influenza [8, 9]. One important use of these Twitter data is sentiment analysis [3]. Wankhade et al. [10] defined that sentiment analysis is a technique used in opinion mining to identify the mood or opinion of various textual expressions. Furthermore, Samuel et al. [11] classified opinions into positive, neutral, and negative which play crucial roles in prediction.

Online reviews, comments, and reactions have been applied successfully with sentiment analysis in areas such as business review analysis, financial market survey, recommendation, and effectiveness analysis, etc. [12].

According to COVID-19 vaccination data made available by Mathieu et al. [13], a total number of 5,395,581 total vaccinations were carried out in Nigeria within its 1st month of lease (March 15, 2021–April 5, 2021). Within this month, there were various reactions from Nigerian citizens about the administration and usage of the vaccine. Therefore, this article presents a study aimed at eliciting citizen's opinion about the pandemic during the 1st month of the vaccination period.

However, studies have shown that three levels of sentiment analysis are possible: document-, sentence-, and aspect-based [10, 14, 15]. The entire opinion paper is examined using document-level sentiment analysis, which then speculates on whether the document represents a positive or negative sentiment. In some instances, this may not be always true as it is not always possible to get a single point of view. The opinion document is divided into sentences at the sentence level, and the sentiments of each sentence are then extracted and classified as neutral, negative, or positive polarity. This implies that, compared to the document level, the sentence level provides a more detailed sentiment polarity. However, aspect-level sentiment analysis makes an effort to conduct a more thorough and in-depth study of expressed thoughts that are concealed in each sentence. The aspect level directly analyzes the opinion rather than focusing on language constructs such as documents, sentences, paragraphs, phrases, or clauses. One of the crucial aspects of sentiment analysis is measuring and assigning sentiment scores of polarity and subjectivity. Bordoloi and Biswas [16] defined polarity as a process of identifying sentiment orientations in texts or spoken words and are categorized as positive, neutral, or negative. Sentiment subjectivity and objectivity can also be determined in a text or sentence, such that, an objective sentence or text expresses some factual information while a subjective sentence or text expresses personal beliefs, feelings, opinions, allegations, suspicions, etc. Values are assigned to polarity and subjectivity. Polarity has a range of values of -1–1 (-1, 0, 1), which means, a negative statement has a polarity of -1, a neutral statement has a polarity of 0, and a positive statement has a polarity of 1. Also, the subjectivity score has a floating point value that varies between 0 and 1. When subjectivity is close to 0, the statement is factual whereas as subjectivity increases to 1, the statement is believed to be close to an opinion.

As illustrated in Fig. 1, three major techniques have recently been employed to carry out sentiment analysis. They are the machine learning approach, deep learning approach, and sentiment lexicon approach [17].

Machine learning approach

Machine learning techniques are divided into supervised and unsupervised learning. Unlike supervised learning algorithms, unsupervised learning techniques do not use testing datasets in their process, with a couple of stages: a training phase and a test phase for validation. To train a classifier using supervised learning, a well-labeled corpus is needed. In supervised learning, a variety of algorithms can be used. The biggest difficulty with supervised learning approaches is that they require well-defined labeled data; regression and classification are the two methods of supervised learning available.
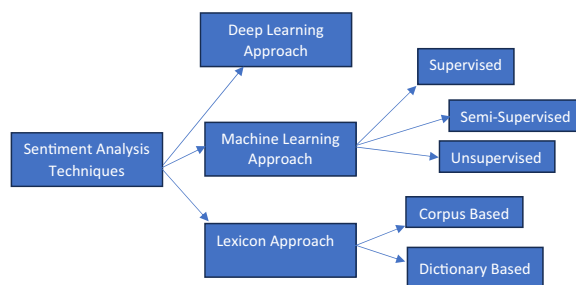
Fig. 1 Approaches to sentiment analysis

Regression involves conditioning labeled datasets, which are then used to repeatedly predict and improve the model using the solutions that are available. Classification aims to assist in identifying the most suitable class labels for predicting favorable, negative, and neutral emotions [18]. In supervised learning, a machine learning algorithm is developed, and it employs labeled data to train and identify tweets while also attempting to anticipate their sentiments. Machine learning or lexicon may be used with unsupervised methods. They do not necessarily necessitate the use of a marked corpus. Only the input dataset is given to the computer using this approach, and the model does not require labeling. Pattern exploration is another term for unsupervised research. Clustering is an example of an unsupervised training approach. A sentiment lexicon is commonly used in an unsupervised approach to sentiment analysis. Semi-supervised learning is a paradigm that bridges the gap between supervised and unsupervised learning. A sequence of labeled and unlabeled data is taken in a semi-supervised learning model, and the purpose of the semi-supervised learning is to identify some of the unlabeled data using labeled knowledge set. Due to the lack of label data, these approaches are applicable in the real world, although they are less effective than supervised methods. The scale of the unlabeled dataset ought to be greater than the labeled dataset, input–output proximity symmetry, relatively simple labeling, and the problem's low dimension are some of the issues with forecasting the sentiment of Twitter datasets.

Deep learning approach

Deep learning is a kind of artificial intelligence-based machine learning technology that uses several layers to gradually extract high-level features gradually. Deep learning is a powerful learning approach that employs neural networks (NN) to complete tasks [19]. In the human brain, biological neurons functions are represented by neural networks, which are analogs. A layer of input, an output layer, and, if desired, a hidden layer are the three layers that make up artificial neural networks. In neural networks, each node is connected to an input value, and each edge is connected to a weight that is initially random but is always fixed. These graphs are completely connected. The weighted sum shown in Eq. (1) can be employed to calculate the weighted sum of a neural network.

$$\text{weighted sum} = \sum w_i x_i + b \tag{1}$$

where w_i is the matrix of a particular weight i, x is the input signal, and b is the bias added to the first layer. To optimize the output, the weighted sum is used as a special

feature. These special functions are known as activation functions, and the activation functions are employed to render the output nonlinear, allowing classification. Relu stands for the rectified linear unit, and it is used to produce only positive and zero values using Eq. (2):

$$\text{Relu} = \max(0, x) \tag{2}$$

where x remains the input signal. Relu also uses a sigmoid function to restrict values to a range of 0–1. It can be generated using Eq. (3):

$$\text{sigmoid} = \frac{1}{(1 + e - x)} \tag{3}$$

where e is the Euler's number.

Deep learning models also use both training and testing datasets. The datasets used for training are collections of related data to train the NN. The NN uses these data to train so as to obtain the intended results because the input solutions have been previously identified. The training accuracy is computed to be the ratio of correctly categorized data instances to the total number of data instances employed as training evidence. The testing dataset in deep learning models is a collection of related data used to put the neural network to the test and see how much it learns to differentiate from the training dataset. The NN is tested on these examples to see if it generates the correct predictions while being trained on new data that are not part of the training collection because the responses to these inputs are already known. The testing accuracy is determined by the ratio of correctly identified data instances to the total number of data instances used as testing evidence. The data fed to the algorithm in supervised learning includes ideal solutions known as labels. Deep learning attempts to learn without being supervised. Some of the popular deep learning techniques are convolutional neural networks, long short-term memory (LSTM), and bidirectional LSTM (BiLSTM) to mention a few.

Sentiment lexicon approach

In a lexicon-based approach, lexicons are used to analyze data which include positive and negative vocabulary; this method determines the sentence's or text document's polarity. A positive score is given whenever the text has extra positive terms, and a negative score is given if the text contains more negative words [20]. In sentimental review, the lexicon-based approach has certain limitations since the approach is dependent on the scale of the lexicon. The sophistication of analysis increases as the lexicon grows in scale. Unlike machine learning, the lexicon-based approach does not necessitate the storing of a wide corpus of data. It calculates the orientation of a text using a lexicon or dictionary. The indicator of subjectivity and opinion in the text, semantic orientation (SO), takes note of the polarity and frequency of terms or phrases. Both of these terms define the text's overall sentiment orientation. An opinion lexicon can be generated either manually or automatically. The manual method of generating the opinion lexicon takes a lot of time, so it is important to combine it with other automated methods [21]. This distinguishes between two types of manual lexicons: general lexicons and category-specific lexicons. Default sentiment words, split words, negation, and blind negation words constitute all terms having a similar sentiment meaning that can be found in the modern lexicon.

The main focus of this study is to perform Twitter sentiment analysis on COVID-19 tweets, using bidirectional LSTM deep learning technique. In "Introduction" section of the article provides a background information about sentiment analysis of COVID-19 tweets and techniques that could be used for its analysis. In "Related works" section summarizes state-of-the-art works that have been done on sentiment analysis. In "Methodology" section discussed in detail, the methodology behind the work while the results obtained are presented in "Results and discussion" section. In "Conclusion" section presented the conclusion and recommendation for future research.

Related works

In a bid to understand public sentiment during the pandemic, several authors have analyzed different tweets as tools to address critical issues and policy-making in health-care organizations and government agencies [22]. Methods based on deep learning have been heavily utilized to attain these goals in order to derive helpful information from the enormous data available on Twitter [23]. The ability of Twitter to be used to monitor people's concerns about disease outbreaks was explored by Xiang et al. [24]. The proposed system can classify people's tweets into positive and negative. This classification can then be visualized in chart form so that health workers can easily visualize people's opinions for quick and easy interpretation. Boon-Itt and Skunkan [25] explored how sentiment analysis and subject modeling might yield important information regarding patterns of the conversation about the COVID-19 pandemic on social networks as well as various views to explore the COVID-19 problem, which has raised significant awareness among the public. This study demonstrates that Twitter is an effective medium for determining people's interest and perception regarding COVID-19. A recommender system based on users' social media history and sentiments expressed on Twitter was proposed by Khattak et al. [7]. The proposed system assists individuals in compiling a summary of popular sentiment on entities of concern. With this, only tweets of interest will be made available to the users. A study aimed at examining Twitter users' feelings and opinions about COVID-19 was carried out by Xue et al. [26]. A total of 1.9 million Tweets about coronavirus were analyzed using machine learning methods. Afterwards, the retrieved tweets were grouped into topics such as "Updates on verified cases," "COVID-19 associated death," "cases outside China (worldwide)," "COVID-19 outbreak in South Korea," "early Indicators of the outbreak in New York," "Diamond Princess cruise," "economic impact," "Preventive measures," "authorities," and "supply chain." The findings showed that treatment and symptoms tweets were not among the most common topics discussed on Twitter. Alhajji et al. [27] also carried out a sentiment analysis of messages on twitter in Saudi Arabia about the government preventive measures against COVID-19. The Naïve Bayes machine learning technique was used to extract the sentiment polarities of these tweets. Similarly, Dubey [28] collected and analyzed COVID-19-related tweets from twelve states of different nations, between March 11 and March 31, 2020. The purpose of the study was to determine how residents of various nations responded to pandemic outbreaks.

In the same vein [29], he used sentiment analysis to examine the effects of COVID-19 on the global community. He discovered that more positive statements were directed to frontline workers and health practitioners through motivation and appreciation. Also,

positive sentiments were generated on advocacy for healthy diets, heeding precautions, spending enough time with families, etc. However, negative sentiments related to the loss of jobs, number of casualties and infected victims, increased unemployment rates, etc., were also analyzed. Das and Kolya [30] utilized polarity notation, which consists of positive classes as 1 or negative classes as 0, without identifying neutral sentiments. The higher probability combination of a positive or negative cumulation defines the polarity tag of a tweet by the binary output variable *S*, which is the expression of both the polarity generated values. This study explores the use of deep CNN for sentiment analysis of COVID-19 tweets. Neelakandan et al. [31] proposed an effective method for using Twitter data for sentiment analysis. The pre-processing of the Twitter database comprises stemming, tokenization, removal of numbers, stop word removal, and other processes. The preprocessed words are then sent to Hadoop Distributed File System (HDFS), where the MapReduce algorithm is used to decrease the repetition of terms. Both emoticons and non-emoticons are demanded in exchange for functionality. The features generated were prioritized based on their intended meaning. The deep learning modified neural network (DLMNN) was then used to do the classification. To demonstrate the best outcome of the suggested model, the experimental data were analyzed using some other traditional methods such as NN, K-Means, and support vector machine to mention a few. The evaluation's findings demonstrate that DLMNN outperformed all other classifiers in terms of performance. According to Abiola et al. [32], by using TextBlob and the VADER analyzer, the study examines how historical tweets on the coronavirus pandemic (COVID-19) elicited various emotional reactions. The study demonstrates, among other things, how great of a cultural, natural, and financial impact it has on Nigeria. According to the study's findings, social media data can be utilized to guide decisions made by organizations, governments, and countries around the world about how to prevent the harmful impacts of COVID-19 and address the issue of misinformation. This article presents the outcome of a study aimed at examining the use of BiLSTM deep learning technique for sentiment analysis of COVID-19 tweets. A total number of 197,327 tweets were retrieved from Nigeria Twitter domain within the 1st month of COVID-19 vaccination. The tweets were used to validate the proposed deep learning technique.

Methodology

The methodology employed in this research is illustrated in Fig. 2. The steps involved are data collection, data pre-processing, model design and implementation, training and learning annotated data features, model testing and validation, and performance evaluation of the model.

Data collection

This study used two sets of data: contextual investigation data (case study data) and deep learning training and validation tweets datasets. The contextual investigation information is the information on remarks of individuals via a chosen web-based media (Twitter) as regards the state of the COVID-19 pandemic in the country. The model training/testing datasets were obtained from Kaggle Sentiment140 tweet datasets containing a total of 789,306 sentiment annotated tweets of which 591,979 samples were put to use for training, and 197,327 samples were employed
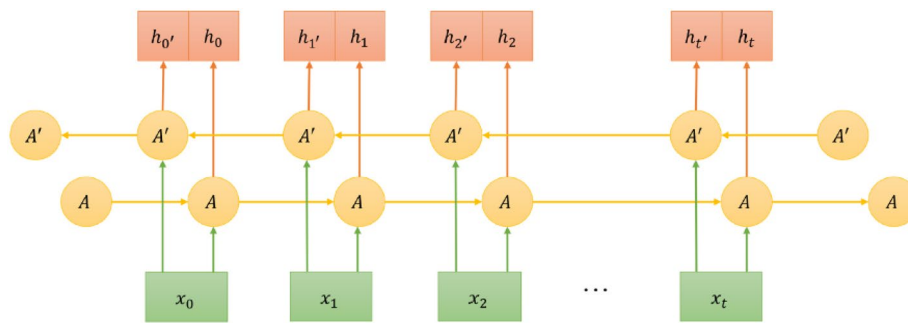
Fig. 2 Bidirectional long short-term memory (BiLSTM) [33]

for testing and validating the model. For the case study data, Twitter API was used to fetch COVID-19-related tweets; however, the API has some limitations such as the inability to fetch tweets that are older than a week, tweets that are incomplete or truncated, as well as repetitive and redundant tweets. Therefore, a web scraping script for additional data received from the Twitter API was added. The technique of web scraping scraps tweets by using a keyword on Twitter; in this case, COVID-19 was used. A total of 197,327 tweets obtained from the Nigeria twitter domain were finally used as the case study data. The data were obtained within the 1st month of administration of COVID-19 vaccine in Nigeria (March 15, 2021–April 5, 2021).

Data pre-processing
After gathering the predefined datasets, utilizing them as they are may not deliver effective outcomes as they contain numerous insignificant components that could hinder the optimum performance of the model. Therefore, pre-processing requires the elimination of unnecessary elements from the sample dataset to eliminate or at least minimize to a degree, the volume of noise that will prevent the optimal performance of the model. Pre-processing of input data is a very important phase. At this point, the dataset is standardized and prepared for the classification algorithm so that the particular algorithm can operate correctly in a minimum of time and provide exact results. All pre-processing techniques were applied using a Python script, which has located and removed stringed patterns of unnecessary and/or unusable elements, such as URLs, HTML tags, emoji, etc. NLTK, a Python module for text processing that removed the English stop words and performed lemmatization of tweets, was used. The tweets were also converted to lowercase representation while unnecessary elements such as usernames, mentions, links, punctuations, hashtags, and retweets were also removed. Counting the words in the tweets was also crucial because tweets with few words in them may not provide the necessary information, or they may even be misleading. Therefore, the numbers of word count in each tweet were calculated and averaged. This word count would serve as the sequence length of each tweet input to the model. Finally, each sentence in the dataset was tokenized and fed as input to the embedding layer of the proposed model.

Model architecture definition

The model used in this paper was developed by using a bidirectional LSTM. It is a variant of recurrent neural networks that carry out sequential processing with two LSTMs. One of the LSTMs processes previous data in the foremost direction while the other processes future data in the opposite direction. It is more effective than a single-model LSTM. Therefore, the improved version of LSTM is BiLSTM. The traditional LSTMs can only learn sequential data in the left-to-right direction. However, BiLSTMs are capable of learning sequential data in both the left-to-right and right-to-left directions. It is advantageous given that a word's meaning can vary based on the context that is being conveyed based on the words that follow or precede it. As shown in Fig. 2, a BiLSTMs output is created by joining the outputs of a left-to-right (h_t) and a right-to-left LSTM ($h_{t'}$). A is a left-to-right LSTM, A' is a right-to-left LSTM, and x_t is an input.

There are five levels in the model's hierarchy:

  i. Embedding layer—it transforms the text sequence into a word embedding matrix.
 ii. BiLSTM layer—this is for modeling the semantic representation in long sequences.
iii. Attention layer—it aims to integrate the final BiLSTM output to obtain significant sentimental polarity information in the sequence after interactive learning.
 iv. Concatenation layer: Inputs are taken by a concatenation layer and combined with a given size. Except for the concatenation dimension, the inputs must be the same size in all dimensions.
  v. Output layer with a Softmax classifier.

The BiLSTM model was implemented using the Keras deep learning library. It is a widely used deep learning platform. All initial word embedding vectors were initialized by the network itself, which is usually initialized to zero and corrected during model training. To find a suitable word vector dimension while training the word vector, the dimension of the word vector was set to 200. The parameters of the BiLSTM neural network model are captured in Table 1.

As illustrated in Fig. 3, the model was compiled using the binary cross-entropy loss function and the Adams optimizer as the loss optimization algorithm. The entire training lasted for 4 epochs. The validation loss is no longer reduced after the 4th epoch and training beyond that epoch would lead to overfitting of the model.

Table 1 Parameters of the BiLSTM neural network model

Hyper parameters	Value
Tokenizer max word features	10,000
Maximum sequence length	70 (mean average sequence length)
Bidirectional LSTM layer cell size	140 per layer
Attention units	10
Fully connected layer dimension	20
Output layer activation	Sigmoid
Dropout fraction	0.05
Training batch size	100
Training epochs	30 with early stopping monitoring change in the validation loss value

Akande *et al. Journal of Electrical Systems and Inf Technol*   (2023) 10:50

Page 10 of 17
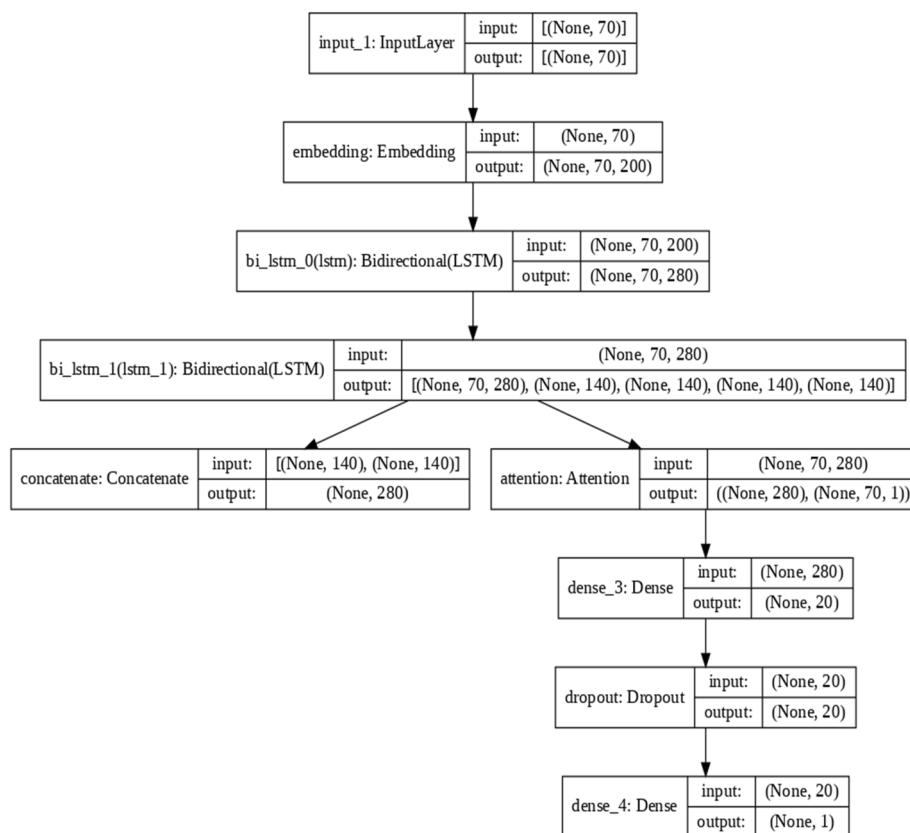
Fig. 3 Architecture of the proposed model

Results and discussion

To evaluate the performance of the proposed model, it was tested and assessed using metrics such as accuracy, precision, recall, and F1-score on both datasets. These were generated from the computed true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values. TP refers to events that have been accurately identified as positive while TN refers to events that are accurately classified as negative. Also, FP is a term used to describe incidents that are incorrectly labeled as positive while FN occurs when negative situations are incorrectly categorized. Table 2 presents the performance evaluation results of the proposed model during training. It shows how the model fits the tweets training data.

Figure 4 illustrates how the model performs on the training data based on the computed accuracy (Fig. 4a), recall (Fig. 4b), precision (Fig. 4c), and AUC values (Fig. 4d). The model was validated and tested with a fraction of the training data. The area under curve metric illustrated in Fig. 4d shows how well (as a fraction between 0 and 1) the model differentiates sequences that belong to a specific class. A value closer to 0 shows an inability to differentiate sentiment polarity, and according to the training data, our model shows high proficiency in distinguishing sentiment polarity labels.

Table 3 presents the results of the performance evaluation of the model on test data. About 78% accuracy, recall, and F-score were achieved. Although not performing badly, the proposed model obtains the same value for all metrics. This can be

Akande *et al. Journal of Electrical Systems and Inf Technol*   (2023) 10:50

Page 11 of 17

Table 2 Performance evaluation results of the proposed model

	TP	FP	TN	FN	Accuracy (%)	Precision (%)	Recall (%)	AUC	Training epochs	Training time (mins)
Training	–	–	–	–	82.09	81.74	82.61	0.9043	4	1309.1s∼21.81
Validation	77,139	21,433	77,349	21,406	78.29	78.26	78.27	0.8655	–	
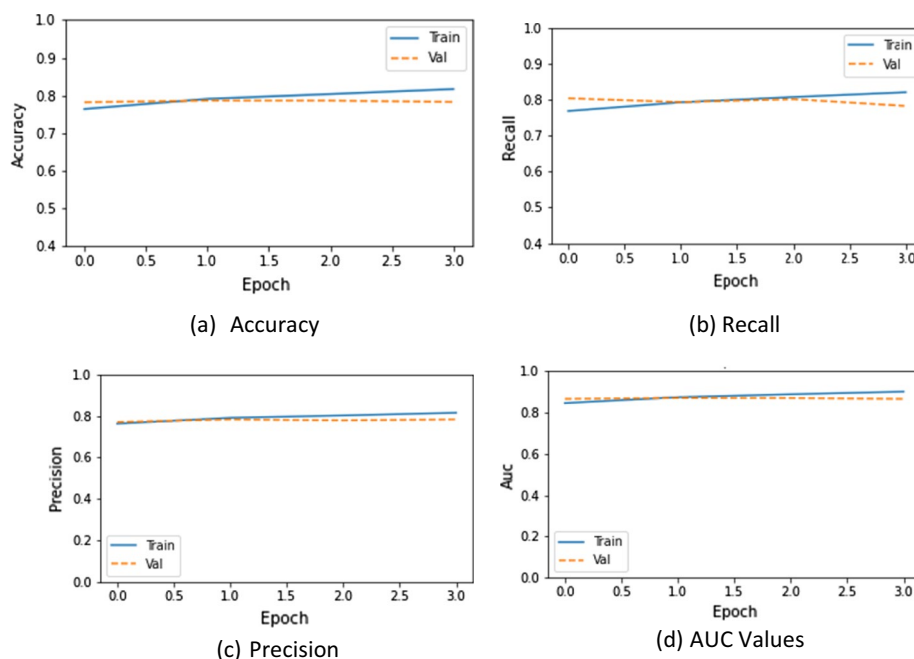
Fig. 4 Performance metrics of the model on the training data

Table 3 Performance metrics of the model on testing data

Performance metric	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	Support
Positive classes	78	78	78	78	98,545
Negative classes		78	78	78	98,782
Macro average on both sentiment labels		78	78	78	197,327
Weighted average on both sentiment labels		78	78	78	

interpreted as the model showing no biases to either polarity classification. This could have resulted from:

    i.   Balanced dataset having similar fractions of belonging to both sentiment polarity labels

   ii.   Attention mechanism regularization overriding aggressive weight adjustment during backpropagation

 iii.   The model learns too fast given the small epoch amount before training terminates.

Support is the number of samples belonging to a specific polarity in the test data. In the field of sentiment analysis, the macro average and weighted average are distinct methodologies employed to assess the overall effectiveness of a given model. The macro average is a statistical measure that computes the average performance across all classes, without considering the support or the proportion of test data belonging to each class. The algorithm assigns equal weight to each class in the calculation of the average. Conversely, the

weighted average incorporates the consideration of the level of support for each class. The average performance is calculated by taking into account the relative representation of each class within the dataset. When dealing with an imbalanced dataset, it is possible to assign greater credence to specific forecasts by utilizing the weighted average technique. Nevertheless, in order to ensure equitable treatment of each class throughout the evaluation of your model's performance, the macro average might be employed.

Furthermore, since the model was trained on binary classification, it only outputs predictions as probabilities between 0 and 1 where values closer to 0 would be more negative and vice versa for positive predictions. Hence, a natural threshold for separating sentiment polarity labels would be 0.5. Using this threshold on the predictions on the case study data, the model classified 98,545 tweet instances as positive and 98,782 as negative. As illustrated in Fig. 5, there appeared to be an almost balanced opinion about citizens emotions to the pandemic in Nigeria within the 1st month of vaccination. This further affirmed the ability of BiLSTM deep learning technique in accurately analyzing sentiments.

A closer and subjective look at the classifications presented in the analysis text file available in Fig. 6 showed the presence of some outliers. Those are tweets whose contents are not related to COVID-19 but were captured because of the hashtags used. The presence of outliers due to the hashtags used and their accurate classifications further shows the ability of the proposed technique in text classification. However, it revealed a new research endeavor that should aim at removing such outliers during sentiment analysis so that only tweets related to the topic of interest will be analyzed. This could also be seen as a limitation of sentence approach to sentiment analysis and could reveal why aspect-based approach is preferable.

Despite the presence of these outliers, the proposed model showed remarkable ability in classifying the polarity of sentiments in the COVID-19 case study data file. From objective cross-examination of the model's performance on the case study data, it is safe to infer that the model attention mechanism focuses on the presence of sentiment
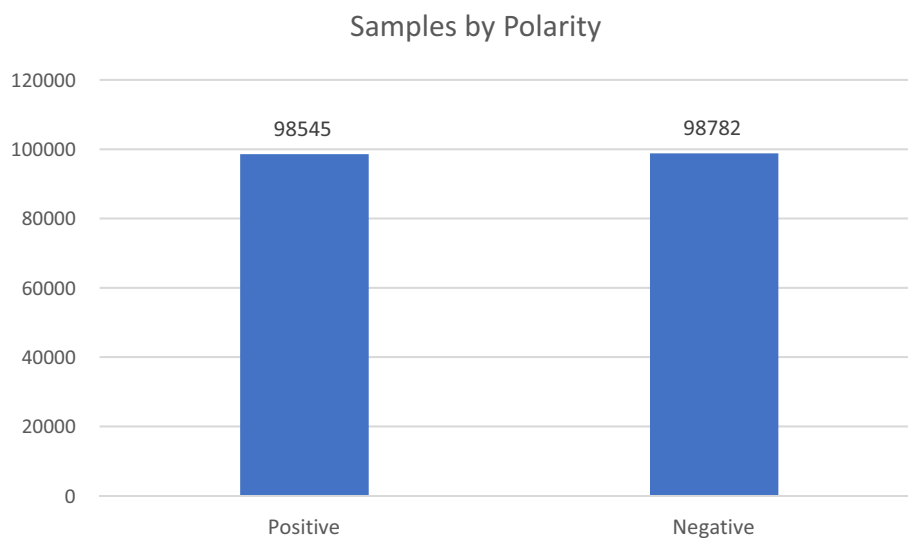
Fig. 5 Sentiment polarity of the test data
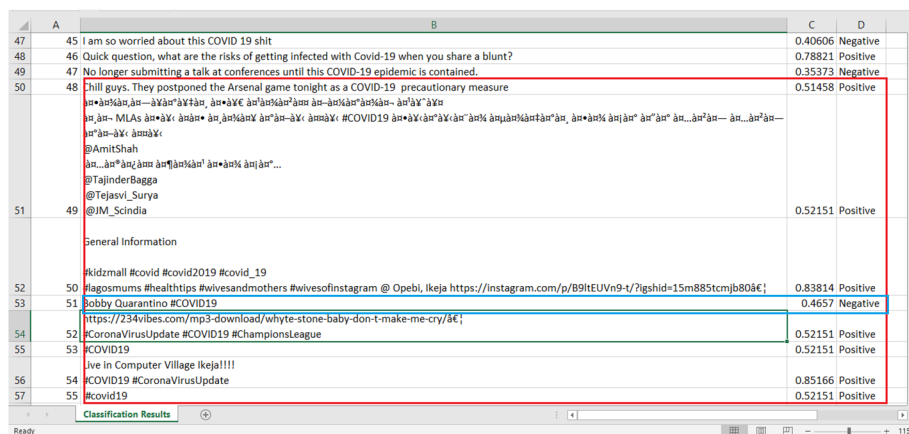
Fig. 6 The presence of outliers in the sentiment result data file

Table 4 Results comparison with existing works

S/N	References	Proposed methodology	Accuracy (%)	Precision (%)	Recall (%)	AUC values (%)
1	Anitha and Metilda [34]	SVM	70.66	71.45	71.33	–
		NB	66.97	67.31	67.33	–
		RNN	69.34	70.04	70.05	–
2	Basiri et al. [35]	CNN	81.60	–	–	–
		BiGRU	79.7	–	–	–
3	Qi and Shabrina [21]	Random forest	45.00	53.00	47.00	–
		MultinomialNB	62.00	61.33	58.00	–
		SVC	71.00	69.00	69.67	–
4	Parveen et al. [6]	*GARN	96.12	93.90	94.34	–
		BiLSTM	94.59	91.99	92.27	–
		**BiGRU	95.79	93.59	94.06	–
5	Mahadevaswamy Mohamad Sham and Mohamed [36]	BiLSTM	91.40	–	–	–
6	Minaee et al. [37]	LSTM and CNN	90.00	–	–	–
7	Senthil and Malarvizhi [38]	LSTM-CNN	98.60	73.00	83.00	98.60
		LSTM	89.50	–	–	89.5
		SAMF-BiLSTM	83.30	–	–	–
	Proposed method	BiLSTM	78.29%	78.26%	78.27%	86.55

*Gated attention recurrent network

**Bidirectional gated recurrent unit

opinioned words in a tweet, allotting weights to its context and position in the tweet to determine the overall sentiment of a tweet.

To examine how the proposed BiLSTM technique works with other techniques reported in the literature, a comparison of results obtained in the existing literature was compared with ours. Table 4 summarizes these results. The accuracy obtained with BiLSTM (78.29%) seems to be the lowest so far when compared with others. This could be the reason why most researchers prefer a hybrid BiLSTM algorithm. Nevertheless,

the accuracy level obtained could be subjective to the type of dataset used among other factors.

Conclusion

This study has demonstrated the prowess of bidirectional attention network model as an effective deep learning technique that can be used for sentiment analysis of COVID-19 tweets. A total number of 197,327 tweets were collected from the Nigerian Twitter domain using #COVID or #COVID-19 hashtags as keywords. A sentiment analysis of 98,545 (49.93%) positive sentiments and 98,782 negative sentiments (50.06%) was recorded. This shows that the proposed technique could accurately differentiate between the positive and negative sentiments. However, the presence of outliers in the sentiment result data file could be seen as a limitation of the document sentiment approach employed. Yet, the deep learning technique was able to accurately classify the outliers correctly. Nevertheless, it can be conjectured that Nigerian citizens expressed a mixed feelings about the global pandemic during the 1st month of the vaccination period. The study once again supports the call for the use of social media platform as a medium of eliciting the opinion of the populace about government policies. Therefore, it is recommended that sentiment analysis and opinion mining be given more attention and funding so that they might be used as assessment tools in times of crisis and unrest.

Declarations

Competing interests
All authors agree to the content of the article and thereby declare no competing interest.

References
1. Ashish K, Safi UK, Ankur K (2020) COVID-19 pandemic: a sentiment analysis: a short review of the emotional effects produced by social media posts during this global crisis. Eur Heart J 41(39):3782–3783. https://doi.org/10.1093/eurheartj/ehaa597
2. Eboibi FE, Robert E (2020) Global legal response to coronavirus (COVID-19) and its impact: perspectives from Nigeria, the United States of America and the United Kingdom. Commonwealth Law Bull 47(4):593–624. https://doi.org/10.1080/03050718.2020.1835507
3. Khattak A, Zubair M, Ishaq Z, Haider W, Hameed IA (2021) Enhanced concept-level sentiment analysis system with expanded ontological relations for efficient classification of user reviews. Egypt Inform J. https://doi.org/10.1016/j.eij.2021.03.001
4. Laor T (2022) My social network: group differences in frequency of use, active use, and interactive use on Facebook, Instagram and Twitter. Technol Soc 68:101922. https://doi.org/10.1016/j.techsoc.2022.101922

5. Cao G, Shen L, Evans R, Zhang Z, Bi Q, Huang W, Yao R, Zhang W (2021) Analysis of social media data for public emotion on the Wuhan lockdown event during the COVID-19 pandemic. Comput Methods Programs Biomed 212:106468. https://doi.org/10.1016/j.cmpb.2021.106468

6. Parveen N, Chakrabarti P, Hung BT, Shaik A (2023) Twitter sentiment analysis using hybrid gated attention recurrent network. J Big Data. https://doi.org/10.1186/s40537-023-00726-3

7. Khattak AM, Batool R, Satti FA, Hussain J, Khan WA, Khan AM, Hayat B (2020) Tweets classification and sentiment analysis for personalized tweets recommendation. Complexity 2020:1–11

8. Edo-osagie O, La BD, Lake I, Edeghere O (2020) A scoping review of the use of Twitter for public health research. Comput Biol Med 122:103770. https://doi.org/10.1016/j.compbiomed.2020.103770

9. Jia Xue, Chen J, Chen C, Zheng C, Li S, et al. (2020) Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLOS ONE 15(9): e0239441. https://doi.org/10.1371/journal.pone.0239441

10. Wankhade M, Rao ACS, Kulkarni C (2022) A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55(7):5731–5780. https://doi.org/10.1007/s10462-022-10144-1

11. Samuel JIM, Rahman M, Ali GGN, Han P, Chong JOO, Member S (2020) Feeling positive about reopening ? New normal scenarios from COVID-19 US reopen sentiment analytics. IEEE Access 8:142173–142190

12. Rodríguez-Ibánez M, Casánez-Ventura A, Castejón-Mateos F, Cuenca-Jiménez PM (2023) A review on sentiment analysis from social media platforms. Expert Syst Appl 223:119862. https://doi.org/10.1016/j.eswa.2023.119862

13. Mathieu E, Ritchie H, Ortiz-Ospina E et al (2021) A global database of COVID-19 vaccinations. Nat Hum Behav. https://doi.org/10.1038/s41562-021-01122-8

14. Akande ON, Enemuo SN, Akande HB, Vincent O, Balogun A, Ayoola J (2022) TWEERIFY: a webbased sentiment analysis system using rule and deep learning techniques. In: Chaki N et al (eds) Lecture notes on data engineering and communications technologies, proceedings of international conference on computational intelligence and data engineering, vol 99, 978–981–16–7181–4, 511945_1_En (Chapter 7)

15. Zhu L, Xu M, Bao Y, Xu Y, Kong X (2022) Deep learning for aspect-based sentiment analysis: a review. PeerJ Comput Sci 8:e1044. https://doi.org/10.7717/peerj-cs.1044

16. Bordoloi M, Biswas SK (2023) Sentiment analysis: a survey on design framework, applications and future scopes. Artif Intell Rev. https://doi.org/10.1007/s10462-023-10442-2

17. Kaur G, Sharma A (2023) A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data. https://doi.org/10.1186/s40537-022-00680-6

18. Salama ES, El-khoribi RA, Shoman ME, Wahby MA (2020) A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition. Egypt Inform J 22:1–10

19. Colón-ruiz C, Segura-bedmar I (2020) Comparing deep learning architectures for sentiment analysis on drug reviews. J Biomed Inform 110:103539. https://doi.org/10.1016/j.jbi.2020.103539

20. Ojeda-Hernandez M, Lopez-Rodriguez D, Mora N (2023) Lexicon-based sentiment analysis in texts using formal concept analysis. Int J Approx Reason 155:104–112

21. Qi Y, Shabrina Z (2023) Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. Soc Netw Anal Min 13(1):31. https://doi.org/10.1007/s13278-023-01030-x

22. Sharma D, Sabharwal M, Goyal V, Vij M (2020) sentiment analysis techniques for social media data : a review sentiment analysis techniques for social media data : a review. Springer, Singapore. https://doi.org/10.1007/978-981-15-0029-9

23. Paramés-Estévez S, Carballosa A, Garcia-Selfa D, Munuzuri A (2023) Artificial intelligence techniques used to extract relevant information from complex social networks. Entropy 25(3):507. https://doi.org/10.3390/e25030507

24. Xiang J, Soon AC, James G (2013) Monitoring public health concerns using twitter sentiment classifications. In: IEEE international conference on healthcare informatics, pp 335–344. https://doi.org/10.1109/ICHI.2013.47

25. Boon-Itt S, Skunkan Y (2020) Public perception of the COVID-19 pandemic on twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveil 6(4):e21978. https://doi.org/10.2196/21978

26. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, Zhu T (2020) Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 22(11):e20550

27. Alhajji M, Al Khalifah A, Aljubran M, Alkhalifah M (2020) Sentiment analysis of tweets in saudi arabia regarding governmental preventive measures to contain COVID-19. Preprints.org. https://doi.org/10.20944/preprints202004.0031.v1

28. Dubey AD (2020) Twitter sentiment analysis during COVID19 outbreak. SSRN Electron J. https://doi.org/10.2139/ssrn.3572023

29. Yadav A, Vishwakarma DK (2021) A language-independent network to analyze the impact of COVID-19 on the world via sentiment analysis. ACM Trans Internet Technol 22(1):1–30. https://doi.org/10.1145/3475867

30. Das S, Kolya AK (2021) Predicting the pandemic: sentiment evaluation and predictive analysis from large-scale tweets on Covid-19 by deep convolutional neural network. Evol Intell 15(3):1913–1934. https://doi.org/10.1007/s12065-021-00598-7

31. Neelakandan S, Paulraj D, Ezhumalai P, Prakash M (2022) A deep learning modified neural network (DLMNN) based proficient sentiment analysis technique on Twitter data. J Exp Theor Artif Intell. https://doi.org/10.1080/0952813x.2022.2093405

32. Abiola O, Abayomi-Alli A, Tale OA, Misra S, Abayomi-Alli O (2023) Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser. J Electr Syst Inf Technol 10(1):1–20. https://doi.org/10.1186/s43067-023-00070-9

33. Pasupa K, Seneewong T, Ayutthaya N (2019) Thai sentiment analysis with deep learning techniques : a comparative study based on word embedding, POS-tag, and sentic features. Sustain Cities Soc 50:101615. https://doi.org/10.1016/j.scs.2019.101615

34. Anitha S, Metilda M (2022) Apache Hadoop based effective sentiment analysis on demonetization and covid-19 tweets. Glob Transit Proc 3:338–342

Akande *et al. Journal of Electrical Systems and Inf Technol*   (2023) 10:50

Page 17 of 17

35. Basiri ME, Nemati S, Abdar M, Asadi S, Acharrya UR (2021) A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. Knowl Based Syst. https://doi.org/10.1016/j.knosys.2021.107242
36. Mahadevaswamy Mohamad Sham N, Mohamed A (2022) Climate change sentiment analysis using lexicon, machine learning and hybrid approaches. Sustainability 14(8):4723. https://doi.org/10.3390/su14084723
37. Minaee S, Azimi E, Abdolrashidi A (2019) Deep-sentiment: sentiment analysis using ensemble of CNN and Bi-LSTM models. https://arxiv.org/abs/1904.04206v1 [cs.CL]
38. Senthil Kumar NK, Malarvizhi N (2020) Bi-directional LSTM–CNN combined method for sentiment analysis in part of speech tagging (PoS). Int J Speech Technol. https://doi.org/10.1007/s10772-020-09716-9

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.