Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Performance evaluation of machine learning and deep learning approaches for sentiment analysis on COVID-19 sentiments

[1]Mr Brajesh Kumar Shrivash, [2]Dr. Dinesh Verma, [3]Dr. Prateek Pandey

[1]Research Scholar, Jaypee University of Engineering & Technology A-B Road, Raghogarh, Distt. - Guna (M.P.);
[2]Assistant Professor (Senior Grade), Jaypee University of Engineering & Technology A-B Road, Raghogarh, Distt. - Guna (M.P.);
[3]Assistant Professor (Senior Grade), Jaypee University of Engineering & Technology A-B Road, Raghogarh, Distt. - Guna (M.P.);

Abstract: Sentiment Analysis (SA) is a study of people's opinions on products, services, organizations, disastrous events etc. The biggest challenge in SA is the text and the context in which it is written or said, for sentiments vary with context. Recently, existing studies have proposed different ML and DL models to classify the data. However, there are still challenges in dealing with unstructured text, classification; preprocessing, and good accuracy measures are ongoing problems. Researchers in the fields of psychology and sociology have focused a lot on deciphering people's emotional expressions during the pandemic. In this paper, we analyse the people's sentiments posted during the COVID-19 pandemic; a real scenario has been used to validate the effectiveness of the proposed work. The proposed approach uses different feature sets and classifiers to analyse the collected tweets and classify them into positive, negative, neutral, extremely positive, and extremely negative sentiments. The proposed model was trained and tested using ML and DL algorithms like Naive Bayes (NB), Random Forest (RF), Multilayer Perceptron (MLP), Passive Aggressive Classifier (PAC), Support Vector Classifier (SVC), Logistic Regression (LR), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) for SA.

We use evaluation measures (like accuracy, precision, recall, and F1-score) to rate the success of machine learning (ML) and deep learning (DL) classifiers. To measure the performance, the data set was split in the ratio of 70:30, 80:20 and 90:10 for training and testing the model and observed that LSTM outperforms with the highest 88% accuracy, 88% precision, 87% recall, and 87% F-1 score. According to the results, the proposed model could detect people's sentiments and allow domain practitioners to analyze sentiments efficiently in decision-making for public health communication strategies.

Keywords - Sentiment Analysis**,** classification, machine learning, deep learning, LSTM, RNN

1. INTRODUCTION

Sentiment analysis is a study of people's opinions posted on the internet using different web platforms like Twitter, Blogs, Websites, Applications etc. The rise of Web 2.0 tools like blogs, forums, and social networks has made it possible for people to talk and share about anything to express their thoughts online [1]. On the internet, social media platform is most widely used by people to share their opinion [2]. SA is used to analyze these opinions like climate change, immigration, pandemic, healthcare etc. This can help researchers to understand public opinion and inform policy decisions. SA incorporates text mining, Natural Language Processing (NLP), and data mining [3] which makes it an interdisciplinary area and typically performs at the word, sentence, document, and aspect levels [4, 5].

The problem of automatic SA is an expanding field of study. It's a significant field that has different real-world applications; however, it is not a simple task and has some challenges related to NLP. The biggest challenge in SA is that sentiments are different in one context and change in another context. Sentiments may include misspellings, slang, grammatical mistakes, and other things that make it hard to figure out how people feel. Users also do sentence framing in different ways, which makes it difficult to read and understand the text. There may be a lot of uncertainty about how keywords are defined. This means that words may have different meanings depending on how they are used and in what context they are used. It has been observed that text emotions can be affected by the message's syntax and semantics. If the SA method doesn't pay attention to these things, we might misclassify the message. This shows; how hard and time-consuming it is to include emotions and study them. Hence, SA continues to confront theoretical and technological challenges, which reduce the overall accuracy of identifying the polarity [6, 7].

A sophisticated SA method could help to make SA a simpler problem. Existing studies exhibit that many ML approaches have been used to study how people feel [8]. ML and DL techniques are advanced methods that can be used to do SA; these help to fill the gap between humans and machines to make SA more convenient with structured and unstructured data. In this paper, we incorporated the baseline algorithms of both ML and DL techniques to train and test the proposed model, such as deep Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive Classifier, SVC Algorithm, Logistic Regression, RNN and LSTM to identify the sentiment polarity. The performance of the model was validated using accuracy, precision, recall, and F-1 score which are described in the result section of the paper. This work is the latest study which compares the performance of baseline algorithms of ML and DL. The outcomes of the study suggest that classifying the sentiments ML and DL techniques may give the best result. So, the researchers and readers can directly apply the suggested ML and DL techniques to the related dataset and may get appropriate results without wasting time.

The main goal of this paper was to analyze the state of mind of people when a disease spreads?, the study will be effective in determining people's mental well-being if such kind of conditions occur and will also be useful in devising appropriate strategies to manage the situation in the future.

2. RELATED WORK

After the declaration of COVID-19 as a pandemic, different studies were conducted in SA to analyse the people's sentiments posted on social media sites on COVID-19 disease[9, 10]. The perspectives to do the SA include trend analysis, modelling [11], the mindset of the people on social distancing, vaccination [12], and disease surveillance [13, 14]. In [15], the researchers proposed a novel LSTM–CNN-GS model for SA. CNN, LSTM, NN, K-NN, and CNN–LSTM were used to evaluate the model performance using Amazon reviews and IMDB datasets taken from Kaggle. The findings of the study describe that hyperparameter tuning using Grid Search (GS) leads to improvement with 96% accuracy.

In [16], the SentiStrength tool and SVM Classifier were used to perform SA. The study reveals that the performance evaluation of the SVM algorithm resulted in 87% accuracy. Hence, SVM performs well and may be employed to classify the sentiments.

In [17], the researchers proposed an attention-emotion enhanced AE-LSTM method to classify the sentiments for SA. The proposed hybrid DL model performs well with 89% accuracy.

In [18], the findings of this analysis were based on a survey conducted for SA using ML and DL techniques. The study revealed that the performance measures of DL techniques performed better than ML techniques like SVM.

In [19], SA on Twitter data was performed during the COVID-19 pandemic. The main objective of this research was to study the psychological effects of COVID-19 on people. The outcomes of the study reveal that positive tweets were more which shows the positive attitude of the people during the COVID-19 pandemic.

In [20], the main focus of this work was to analyse sentiments on multiple aspects towards a certain entity; the paper demonstrated the use of the proposed hybrid CNN-GRU (Gated Recurrent Networks) model for SA to the Chinese online reviews of hotels and cars data. It has been identified that CNN retrieved local features and GRU's long-term dependency. The robustness of the model was demonstrated through the experimental study of mentioned datasets. The AUC values for subtasks were recoded as 83.06% and 83.17% on the hotel dataset and 86.93% and 77.42% on the car dataset.

In [21], proposed a novel aspect-gated graph convolutional network (AGGCN) for Aspect-Based Sentiment Analysis (ABSA). The model uses both syntactical and sentiment information. The experimental analysis has been carried out using SemEval datasets namely SemEval-2014, SemEval-2015, and SemEval-2016, and compared the performance with different ABSA models to validate the proposed model. The experimental results demonstrate that the proposed model outperforms the existing state-of-the-art ABSA models on three benchmark datasets.

In [22], this work focuses on the task of ABSA, which identifies the aspects or features of a product or service that are being referred to in a given text and determines the sentiment associated with each aspect. The proposed approach to ABSA comprises extracting explicit aspects from the text using some set of rules and combining these rules to obtain optimal performance.

In [23], this work presents a study on Twitter SA using a Bidirectional Representation for Transformers (BERT) based pipeline. The proposed pipeline included pre-processing, feature extraction, and classification of Italian tweets. The importance of pre-processing is highlighted as cleaning and normalizing the text data to remove noise and irrelevant information. The feature extraction techniques, including bag-of-words, n-grams, and word embedding were discussed. It has been identified that BERT can be a promising approach for SA due to its ability to capture the context and meaning of words. The results of the study show that the proposed pipeline achieves high accuracy to the F1-score of 80% for the binary classification (positive vs. negative) and 60% for the multi-class classification (positive, negative, or neutral).

In [24], in this paper, ABSA was performed using a neural network (NN) model. To evaluate the performance of the model; it was trained on the datasets Rest14, Rest15, and Rest16. Experimental results demonstrate that the proposed model outperforms as compared to existing ABSA models in terms of accuracy, F1 score, and other evaluation metrics. The paper highlighted that the proposed work has the potential to be used in different domains such as e-commerce, social media analysis, and customer feedback analysis.

In [25], this work presents a novel DL framework for SA that combines CNNs, RNNs, and attention mechanisms. To conduct the study, three Twitter datasets were taken into consideration. The results suggest that the proposed model outperforms existing methods and has the potential to improve SA in different areas. However, further research is needed to evaluate the performance of the proposed approach on larger and more diverse datasets.

In [26], this paper presents a novel approach for SA of Tweets using refined neutrosophic sets. The approach has the potential to overcome some of the limitations of traditional SA techniques and achieve higher accuracy.

In [27], the authors proposed a framework for SA on IMDB, Yelp, and Twitter review datasets using CNN, LSTM, and BERT. The study classified sentiments into two parts: binary and multi-class sentiment classification. For the binary sentiment classification, the performance of CNNs, LSTMs, and BERT on the IMDB dataset was compared and it has been observed that BERT outperforms with an accuracy of 94.2%. For the multi-class sentiment classification, again comparison was performed on the same models on both the Yelp and the Twitter datasets. The results show that BERT again outperforms both datasets with accuracies of 67.2% and 66.1%, respectively.

In [28], this work demonstrates DL techniques for SA, which have gained popularity due to their ability to automatically learn features from raw data. The paper includes a detailed overview of different types of DL models, including CNNs, RNNs, and LSTM networks. The study demonstrated that DL techniques are best fitted to analyse people's emotions.

In continuation with this, the following table summarises SA approaches based on some existing research specific to COVID-19 sentiments along with benefits and limitations.

Table 1: SA methodologies based on existing studies

Author's	**Year**	**Proposed work**	**Dataset**	**Benefits**	**Limitation**
[29]	2020	SA for COVID-19	Twitter API	Provided a real-time perspective on the concerns of users such as economic impact, mental health, misinformation, and healthcare.	The tweets were limited to a short period and contextual information.

[30]	2020	SA for COVID-19	Twitter API	France, Switzerland, the Netherland, and the US show signs of anger and distrust on a bigger scale.	The study was confined to specific countries like Australia, China, India, and USA.
[31]	2020	SA for COVID-19	Twitter API	India has succeeded in controlling the COVID-19. Positivity was higher and concerns of passengers, and quarantine people were recorded.	The tweets were limited to a short period i.e. 25th to 28th March 2020.
[32]	2020	COVID-19 Outbreak Analysis in Europe	Data collected from the European Centre for Disease Prevention and Control	Air travel was certainly not the only determinant of the outbreak dynamics, mobility was a strong contributor to the global spreading of COVID-19	The study was confined to Europe specifically.
[33]	2020	Forecast of COVID-19 spreading in China, Italy and France	Data collected from Johns Hopkins University	A comparative analysis of transmission dynamics, growth rates, and peak periods and forecast of the spreading of COVID-19 in China, Italy, and France.	Data was limited to a short period i.e. 22/02/2020 – 15/03/2020.
[34]	2020	SA on online education during COVID-19 in the Philippines	Survey data from Pangasinan State University students	Shows the emotions of students and aids the faculty members in what intervention should be used in instructional delivery	The study was limited to the BS Business Administration and BS Public Administration students of Pangasinan State University, Lingayen Campus
[35]	2020	COVID-19 impacts on university students, in Bangladesh	Survey data from public and private universities, in Bangladesh	Anxiety and depression symptoms were prevalent among students during the outbreak.	The respondent's size was limited.

[36]	2021	SA for COVID-19 Vaccination in India	Twitter API	The majority of the tweets were positive for vaccines but some of them were negative such as fear and anger.	Sample size, geographical representation.
[37]	2021	SA for COVID-19 Vaccination	All COVID-19 Vaccines Tweets from Kaggle	The study gives a direction to understand the sentiments of people regarding the vaccination process.	RNN and LSTM approaches were used for analysis; ML and other DL approaches may be employed to compare the results.
[38]	2021	SA for COVID-19 Vaccination in India	Twitter API	Pandemics cause vulnerability and anxiety.	The BERT model was used to evaluate the performance; other models may also be employed to compare the results in a better way.
[39]	2021	SA for COVID-19	Twitter API	Optimistic, annoyed and joking tweets generally dominate the tweets with a much lower portion of negative sentiments.	Sample size, geographical representation.
[40]	2021	SA for COVID-19	Twitter API	The study showed predictable sentiments of anger, desperation, and hope.	Studies showed predictable sentiments of anger, desperation, and hope.
[41]	2021	SA for COVID-19	Twitter API	Effectiveness of different supervised machine learning models in accurately determining sentiments from COVID-19-related tweets.	Sentiments were not elaborated.
[42]	2022	SA for COVID-19	COVID-19 tweets	The study addresses the growing significance of comprehending public sentiment during the pandemic.	Noisy features may affect the classification.
[43]	2022	SA for COVID-19	Twitter and Weibo	An automated tool helps to recognise emotional well-being, timely monitoring and decision-making.	Macro nature of the study.

[44]	2022	SA for COVID-19	Twitter data	To identify public opinion through tweets automatically.	The study is validated using the BERT-Bi-LSTM ensemble method only. Other methods may also be employed.
[45]	2023	SA for COVID-19	Twitter data	To understand the varying emotional responses and experiences of African Americans during this challenging time.	The study is limited to African-American community sentiments.
[46]	2023	SA for COVID-19	Survey data	The study focuses on multidisciplinary application areas of sentiment analysis.	Research gaps along with possible solutions may be incorporated.

3. METHODOLOGY

This section describes the overall methodology of the study. The study mainly consists of data collection, data preprocessing, and feature extraction, data splitting into training and testing sets with the ratio of 70:30, 80:20, and 90:10 respectively. A model was proposed; trained, tested and validated using Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive, SVC Algorithm, Logistic Regression, RNN and LSTM for sentiment classification. The classification models were validated using certain evaluation parameters like accuracy, precision, recall, and F-1 score. The overall methodology of the proposed work is as follows:

3.1 Dataset – the dataset of COVID-19 sentiments was considered for the conduction of this study. The data set was collected from the Kaggle stored on Google Drive (https://drive.google.com/drive/folders/1dtu1aYYyvgxKLiUBlGg0hnHIjQDhI-qH?usp=sharing) and processed. Table 2 shows an outline of the sentiment categories that were used to conduct this study.

Table 2: Sentiments categories in the dataset	
Positive	11422
Negative	9917
Neutral	7713
Extremely Positive	6624
Extremely Negative	5481
Total	**41157**

This dataset contains a total of 41157 records of which 11422 were positive & 6624 extremely positive, 9917 were negative & 5481 extremely negative and 7713 were neutral tweets. Figure 1 shows the sentiment categories in graphical form to understand them well.

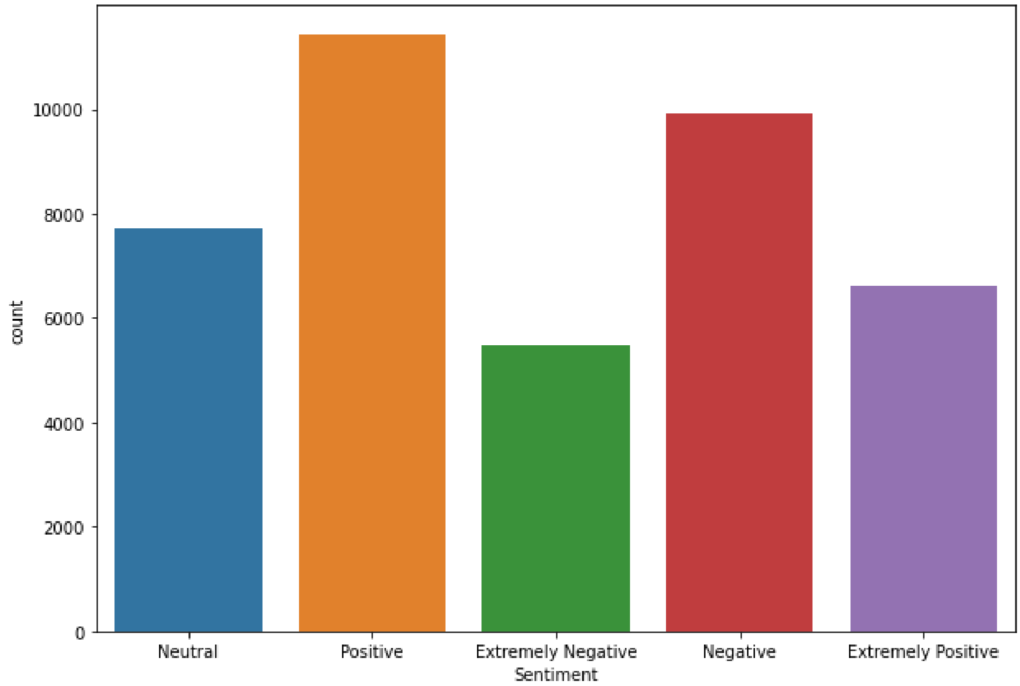

Figure 1: Sentiments Classification

	UserName	ScreenName	Location	TweetAt	OriginalTweet	Sentiment
0	3799	48751	London	16-03-2020	@MeNyrbie @Phil_Gahan @Chrisitv https://t.co/i...	netural
1	3800	48752	UK	16-03-2020	advice Talk to your neighbours family to excha...	positive
2	3801	48753	Vagabonds	16-03-2020	Coronavirus Australia: Woolworths to give elde...	positive
3	3802	48754	NaN	16-03-2020	My food stock is not the only one which is emp...	positive
4	3803	48755	NaN	16-03-2020	Me, ready to go at supermarket during the #COV...	negative
...	...	...	...	...	...	...
41152	44951	89903	Wellington City, New Zealand	14-04-2020	Airline pilots offering to stock supermarket s...	netural
41153	44952	89904	NaN	14-04-2020	Response to complaint not provided citing COVI...	negative
41154	44953	89905	NaN	14-04-2020	You know itÂ□s getting tough when @KameronWild...	positive
41155	44954	89906	NaN	14-04-2020	Is it wrong that the smell of hand sanitizer i...	netural
41156	44955	89907	i love you so much || he/him	14-04-2020	@TartiiCat Well new/used Rift S are going for ...	negative

41157 rows × 6 columns

	UserName	ScreenName	Location	TweetAt	OriginalTweet	Sentiment	label
0	3799	48751	London	16-03-2020	@MeNyrbie @Phil_Gahan @Chrisitv https://t.co/i...	netural	0
1	3800	48752	UK	16-03-2020	advice Talk to your neighbours family to excha...	positive	1
2	3801	48753	Vagabonds	16-03-2020	Coronavirus Australia: Woolworths to give elde...	positive	1
3	3802	48754	NaN	16-03-2020	My food stock is not the only one which is emp...	positive	1
4	3803	48755	NaN	16-03-2020	Me, ready to go at supermarket during the #COV...	negative	2
5	3804	48756	Ã□T: 36.319708,-82.363649	16-03-2020	As news of the regionÂ□s first confirmed COVID...	positive	1
6	3805	48757	35.926541,-78.753267	16-03-2020	Cashier at grocery store was sharing his insig...	positive	1
7	3806	48758	Austria	16-03-2020	Was at the supermarket today. Didn't buy toile...	netural	0
8	3807	48759	Atlanta, GA USA	16-03-2020	Due to COVID-19 our retail store and classroom...	positive	1
9	3808	48760	BHAVNAGAR,GUJRAT	16-03-2020	For corona prevention,we should stop to buy th...	negative	2

Figure 2: Transformation of original Tweets into sentiment polarity and labelled

The data set was divided into training and test data. The data set was explored to determine the relationship between data variables, the structure of the dataset, the presence of outliers, and the data value distribution. After transformation, the performance was measured on a different machine and deep learning algorithms.

3.2 Data preprocessing – data preprocessing is a significant approach to preparing the dataset before actual is used. Data cleaning refers to the removal of noises from the data and making the text in a standard form that is predictable and analyzable for the machine and significant results can be obtained. In our dataset, we convert raw messages into vectors. To do this, we often started by removing web links, numbers, punctuations, and stopwords and converting text to lowercase, which is useful later while parsing. After this; we returned with a list of clean words. In line with this, we did tokenize a text which entails separating characters into tokens and deleting punctuation and stopwords at the same time. The stemming and lemmatization processes are then performed.

3.3 Feature extraction - Following tokenisation, we perform a Term Frequency-Inverse Document Frequency (Tfidf) Vectorizer for feature extraction. TFIDF is a method of transforming textual data to numeric form and Vectorization is the process of turning words into numbers.
The relative term frequency can be calculated for each term within each document as –

$$\text{TF(t,d)} = \frac{\text{number of the times term(t) appears in the document(d)}}{\text{total number of terms in the document(d)}}$$

Inverse Document Frequency quantifies how significant a word is to distinguish each document. We can derive this using the below-mentioned formula –

$$\text{IDF(t,D)} = \log\left(\frac{\text{total no of document(D)}}{\text{number of documents with the term(t) in it}}\right)$$

Overall, TFIDF is calculated as:

$$X_{i,wj} = \frac{1 + log(t_{i,wj})}{1 + log(\sum_i^N t_{i,wj})} * log(\frac{N}{\sum_{wj} t_{i,wj}})$$

Where t_i, wj is the frequency of word w_j appears in document i [37].

3.4 Applying machine and deep learning algorithms – after feature extraction, we did text classification using ML and DL classification models and the performance on different parameters was evaluated. A brief description of deep learning RNN and LSTM approaches is as –

LSTM - Long Short-Term Memory Networks, also known as "LSTMs" are a type of RNN that can learn about long-term dependencies. LSTMs are made to avoid the problem of long-term dependencies. They are almost always good at remembering things for a long time. When you make an artificial intelligence system, it will always be made up of the same parts of a neural network that repeat. There will be only one tanh layer in standard RNNs for this repeating module.

LSTM is a kind of neural network that repeats itself. In RNN, the yield from the last phase is used as an input for the next phase. Hochreiter and Schmidhuber came up with the idea for LSTM [19, 22]. An RNN can't predict what words are in its long-term memory. However, this is an effective approach to predicting the data with more accuracy. RNN can no longer do a good job if the gap length grows. In this case, LSTM can be considered as a better approach because this can have the data for a long time. It is used to make, classify, and predict things based on time series data. LSTM is composed of four distinct neural networks and a specialized form of memory storage unit known as a cell. These cells are in charge of storing information, while gates are in charge of changing the way the memory works. The Forget gate ensures that information that is no longer helpful is deleted from the cell by ensuring that it is no longer present. With the input gate, you can enter information that will be valuable to the cell later on. It is employed to obtain useful information from cells.

Recurrent Neural Networks - "A recurrent neural network (RNN) is a type of artificial neural network where connections between nodes form a directed graph that moves in a certain way over time". This allows it to change its behaviour over time [24]. The main job of an RNN is to process information that comes in a sequence based on the internal memory that the directed cycles build up. Unlike traditional neural networks, RNNs can keep track of how much information they've already worked with and apply it to the next input in the sequence.

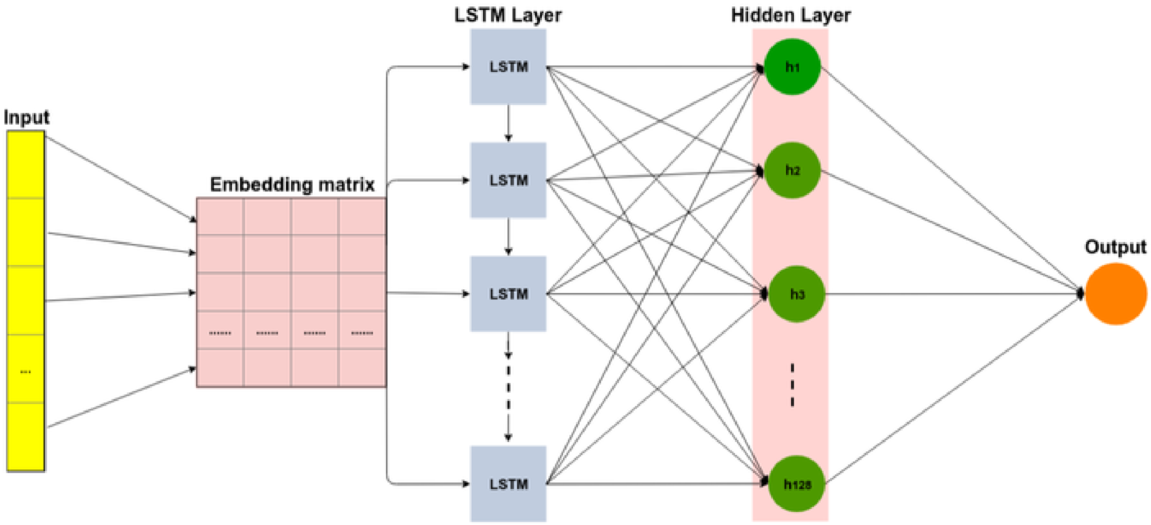

Figure 3: Recurrent Neural Networks [36]

When an RNN is made, a special type called long short-term memory (LSTM) is used. This type of RNN can use long memory as input for activation functions in the hidden layer. The figure below shows the architecture of the LSTM architecture. To make the embedding matrix, input data is first reshaped. Here, LSTM is the next layer, and the final layer is a fully connected layer that may be used to categorise text in a variety of ways [27].

3.5 Proposed model of the study – the following figure describes the model of the proposed work.

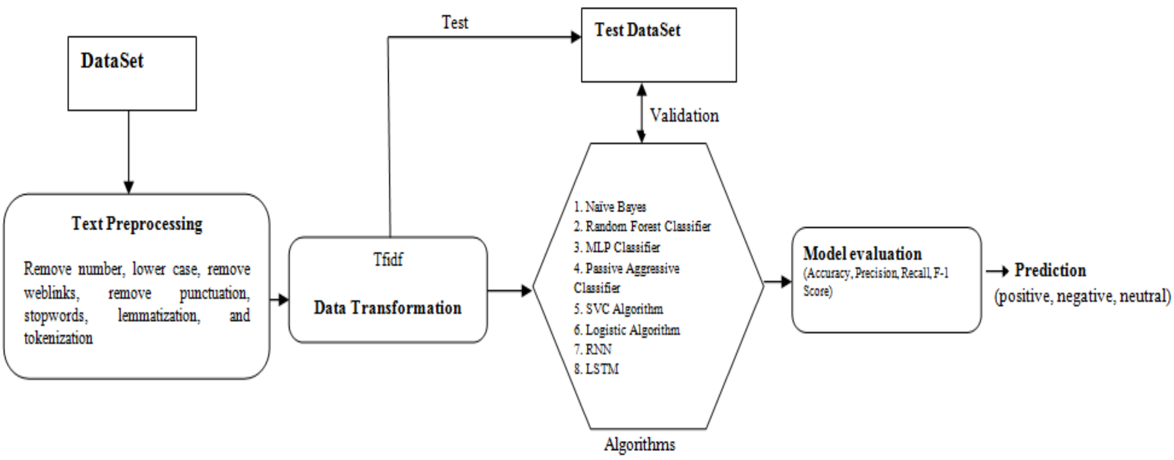

Figure 4: Model of the proposed work

Figure 4 represents the process flow of the proposed model. The collected text data was prepared and preprocessed for the model building. The preprocessing of data includes the removal of numbers, conversion of text into lower case, removal of web links and punctuations, stopwords removal, and splitting of text into smaller units by lemmatization & tokenization. After text preprocessing and preparation; data transformation begins using Tfidf (Term frequency-inverse document frequency) text vectorizer. Term frequency (TF) refers to the frequency or number of times a certain word appears in a document. It quantifies the significance of a particular word in a document. TF encodes each text in the dataset as a matrix, where the rows reflect the number of documents and the columns represent the number of unique phrases found in all the documents. Inverse document frequency (IDF) is a measure of the significance of a term. It seeks to diminish the importance of a phrase if the term's occurrences are scattered throughout documents.

In line with this, in the next step data was split into training and test datasets. The trained data was validated using machine and deep learning approaches like Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive, SVC Algorithm, Logistic Algorithm, RNN, and LSTM algorithms. In the next step, the proposed model was validated using accuracy, precision, recall, and F-1 score.

3.6 Evaluating the model - the dataset was split into training and testing set on three different ratios; 70:30,80:20 and 90:10 and the performance of the model was evaluated. The training dataset was used for building up the model, and a test dataset was used to validate it. The evaluation parameters considered for the study were accuracy, precision, recall, and F-1 score.

4. Experimental results and discussion

The performance evaluation of Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive, SVC Algorithm, Logistic Algorithm, RNN, and LSTM algorithms have been considered with evaluation parameters like accuracy, precision, recall, and F-1 score. The performance evaluation of the model with 70:30, 80:20 and 90:10 were described in Table 3, Table 4, and Table 5 respectively.

Table 3: Performance evaluation on Run 1 of the model with 70 % testing and 30 % test data

Models	**Accuracy**	**Precision**	**Recall**	**F-1 Score**
Naive Bayes	67%	73%	67%	69%
Random Forest	69%	71%	69%	70%
MLP Classifier	71%	71%	71%	71%
Passive Aggressive Classifier	74%	74%	74%	74%
SVC Algorithm	76%	77%	76%	76%
Logistic Algorithm	77%	78%	77%	77%
RNN	81%	82%	80%	81%
LSTM	**86%**	**88%**	**85%**	**86%**

From Table 3, it has been observed that the performance measures for the Naive Bayes were; accuracy 67%, precision 73%, recall 67%, and F-1 score 69%. For the Random Forest, the accuracy was 69%; precision 71%, recall 69% and F-1 score 70%. MLP Classifier was better performed as compared to Random Forest with 71% accuracy, precision, recall and F-1 score values. In comparison with Random Forest &MLP Classifier; the performance evolution for the Passive-Aggressive exhibits a better performance with 74% accuracy, precision, recall and F-1 score values. The performance evaluation for the Support Vector Classifier (SVC) exhibits better than Passive Aggressive Classifier; for the SVC, the accuracy was 76%, precision 77%, recall 76% and F-1 score 76%.

Performance measures for the logistic regression classification were observed with 77% accuracy, 78% precision, 77% recall and 77% F-1 score. The performance evolution for the Logistic regression classification is

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

better than the Support Vector Classifier. From the table, it has been observed that the Recurring Neural Network (RNN) outperform with 81% accuracy, 82% precision, 80% recall and 81% F-1 score. The performance evolution for the RNN classification was recorded better than Logistic Regression. The highest accuracy **86%,** precision **88%,** recall **85%, and** F-1 score value **86% was recorded for the LSTM classifier.**

Hence, it is stated that the performance evaluation for the LSTM technique exhibits greater than the Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive Classifier, SVC Algorithm, Logistic Algorithm, and RNN.

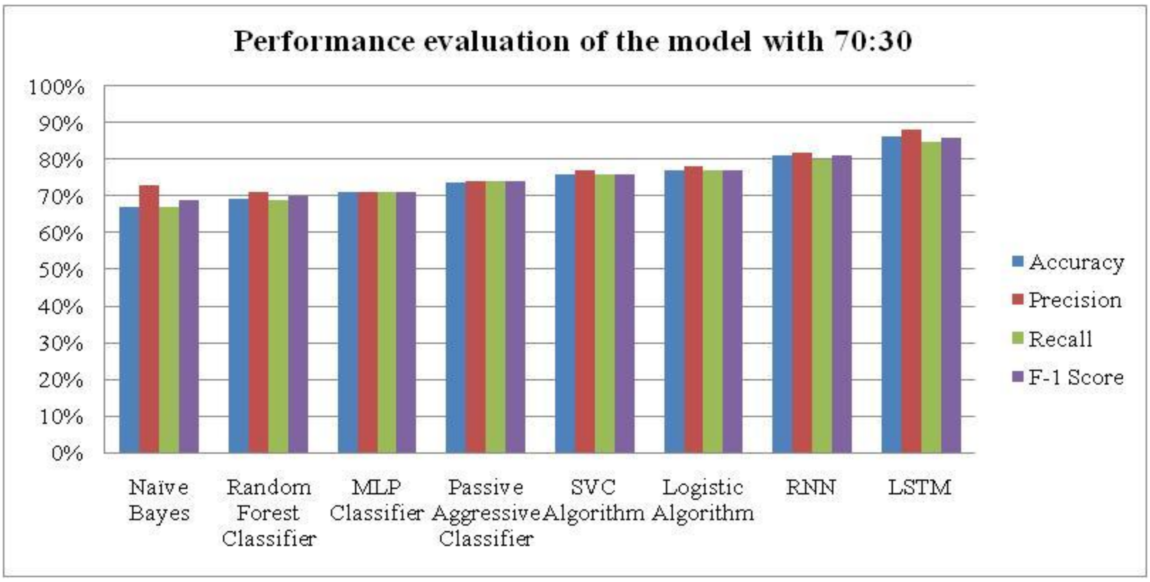

Figure 5: Model accuracy on Run 1 of the model

The graphical representation of the outcome in Figure 5 is showing that the performance evaluation of LSTM on evaluation parameters like accuracy, precision, recall and F-1 score was higher while Naive Bayes observed with least accuracy, precision, recall and F-1 score.

Table 4: Performance evaluation on Run 2 of the model with 80 % testing and 20 % test data

Models	**Accuracy**	**Precision**	**Recall**	**F-1 Score**
Naive Bayes	66%	68%	66%	72%
Random Forest	69%	71%	69%	70%
MLP Classifier	72%	73%	72%	72%
Passive Aggressive Classifier	74%	74%	74%	74%
SVC Algorithm	76%	77%	76%	76%
Logistic Algorithm	77%	78%	77%	77%
RNN	80.90%	82%	79%	81%
LSTM	**87.30%**	**89%**	**86%**	**87%**

Table 4 is showing the outcomes of the second run of the model with 80% testing and 20% testing. It has been observed that the performance measures for the Naive Bayes were; accuracy 66%, precision 68%, recall 66%, F-1 score 72%. For the Random Forest, the accuracy was 69%; precision 71%, recall 69% and F-1 score 70%. MLP Classifier was better performed as compared to Random Forest with 72% accuracy, 73% precision, 72% recall and 72% F-1 score. In comparison with Random Forest & MLP Classifier; the performance evolution for the Passive-Aggressive exhibits a better performance with 74% accuracy, precision, recall and F-1 score values.

The performance evaluation for the Support Vector Classifier (SVC) exhibits better than Passive Aggressive Classifier; for the SVC, the accuracy was 76%, precision 77%, recall 76% and F-1 score 76%. Performance measures for the Logistic Regression classification were observed with 77% accuracy, 78% precision, 77% recall and 77% F-1 score. The performance evolution for the Logistic regression classification is better than the Support Vector Classifier. From the table, it has been observed that the Recurring Neural Network (RNN) outperform with 81% accuracy, 82% precision, 79% recall and 81% F-1 score. The performance evolution for the RNN classification was recorded better than Logistic Regression. The highest accuracy **of 87.3%,** precision of **89%,** recall **of 86%, and** F-1 score **of 87% were recorded for the LSTM classifier.**

Hence, it is stated that the performance evaluation for the LSTM technique exhibits greater than the Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive Classifier, SVC Algorithm, Logistic Algorithm, and RNN.

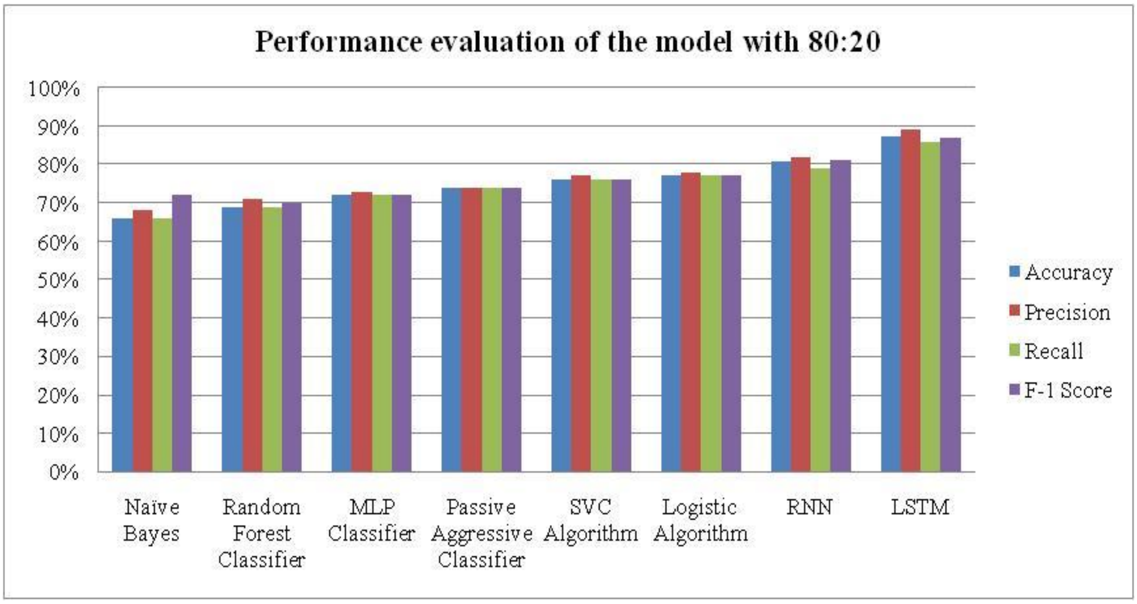

Figure 6: Model accuracy on Run 2 of the model

Figure 6 demonstrates the results when the model was run with 80% testing and 20% test dataset. It has been observed that the performance evaluation of LSTM on evaluation parameters like accuracy, precision, recall and F-1 score was higher while Naive Bayes observed with least accuracy, precision, recall and F-1 score. The accuracy rate on Run 2 of the model has increased as compared to Run 1.

Table 5: Performance evaluation on Run 3 of the model with 90 % testing and 10 % test data

Models	**Accuracy**	**Precision**	**Recall**	**F-1 Score**
Naive Bayes	66%	72%	66%	68%
Random Forest	68%	70%	68%	69%
MLP Classifier	74%	74%	74%	74%
Passive Aggressive Classifier	76%	76%	76%	76%
SVC Algorithm	76%	77%	76%	76%
Logistic Algorithm	78%	78%	77%	77%
RNN	82%	84%	80%	82%
LSTM	**88%**	**88%**	**87%**	**87%**

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Table 5 shows the outcome of the third run of the model with 90% testing and 10% test data. It has been observed that the performance measures for the Naive Bayes were; accuracy 66%, precision 72%, recall 66%, F-1 score 68%. For the Random Forest, the accuracy was 68%; precision 70%, recall 68% and F-1 score 69%. MLP Classifier was better performed as compared to Random Forest with 74% accuracy, precision, recall and F-1 score values. In comparison with Random Forest & MLP Classifier; the performance evolution for the Passive-Aggressive exhibits a better performance with 76% accuracy, precision, recall and F-1 score values.

The performance evaluation for the Support Vector Classifier (SVC) exhibits better than the Passive Aggressive Classifier on precision value with an increment of 1%; for the SVC, the accuracy was 76%, precision 77%, recall 76% and F-1 score 76%. Performance measures for the Logistic Regression classification were observed with 78% accuracy, 78% precision, 77% recall and 77% F-1 score. The performance evolution for the Logistic Regression classification is better than the Support Vector Classifier. From the table, it has been observed that the Recurring Neural Network (RNN) outperform with 82% accuracy, 84% precision, 80% recall and 82% F-1 score. The performance evolution for the RNN classification was recorded better than Logistic Regression. The highest accuracy **of 88%,** precision of **88%,** recall **of 87%, and** F-1 score **of 87% were recorded for the LSTM classifier.**

Hence, it is stated that the performance evaluation for the LSTM technique exhibits greater than the Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive Classifier, SVC Algorithm, Logistic Algorithm, and RNN.

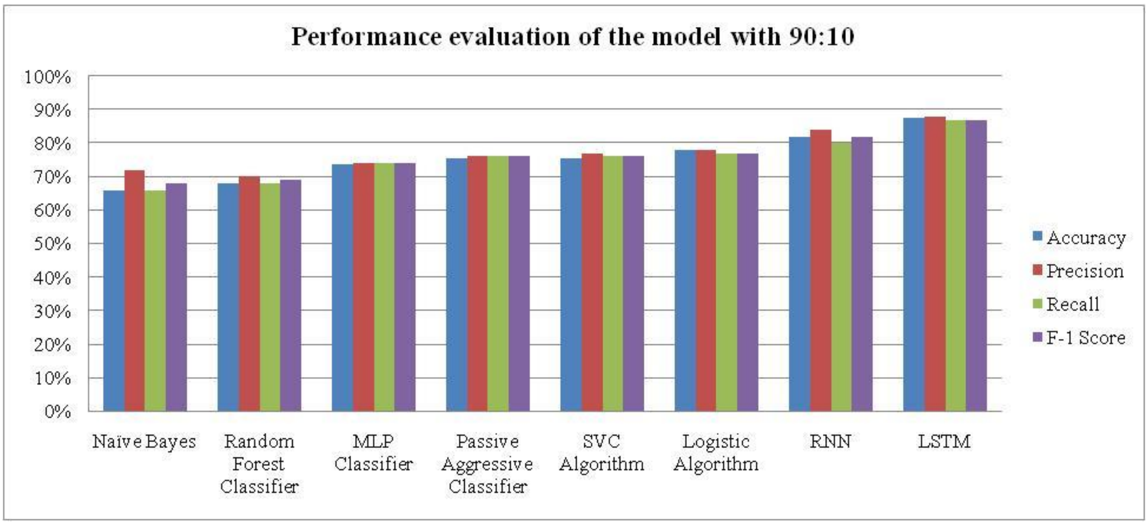

Figure 7: Model accuracy on Run 3 of the model

The graphical representation of Run 3 of the model is shown in Figure 7. Figure 7 demonstrates the results when the model was run with 90% testing and 10% test dataset. The performance evaluation of LSTM on evaluation parameters like accuracy, precision, recall and F-1 score was recorded as higher values while Naive Bayes observed with least accuracy, precision, recall and F-1 score. The accuracy rate on Run 3 of the model was better than Run 2 of the model.

Hence, the performance evaluation for the LSTM exhibits better than the other models. The reason is that LSTM is specifically designed to capture and process sequential information. Traditional RNNs struggle with learning long-term dependencies due to the vanishing gradient problem; LSTMs overcome this issue by gating mechanism that allows the network to selectively remember or forget information over long periods. LSTMs use gates (input, forget, and output gates) to control the flow of information through the network. These gates enable LSTMs to selectively update and discard information, allowing them to focus on important aspects of the input sequence while ignoring irrelevant details. This makes them more effective at capturing semantic relationships over extended sequences. Sentences and text data have a natural sequential structure where the meaning of a word often depends on the context of the preceding words. LSTMs can handle sequences of different lengths without the need for fixed-size input. This adaptability is crucial for processing text data where the length of sentences can vary significantly.

LSTMs have a memory cell that enables the network to store and retrieve information over long periods. The cell state can carry relevant information across multiple time steps, helping the model retain context and make better predictions for sentiment analysis. LSTMs can be combined with pre-trained word embeddings (e.g., Word2Vec, GloVe) to initialize the network with meaningful word representations. This helps the model capture semantic relationships between words and provides a good starting point for learning sentiment-related patterns.

DISCUSSION

The main goal of this research work was to examine people's opinions during an unfavorable situation like lockdown, pandemic, competitive anxiety etc. Furthermore, the proposed framework may have compared different ML and DL models to see which ones performed better in analyzing and classifying the sentiment on COVID-19. Hence, the research work could contribute to understanding public sentiments, concerns, and attitudes during the pandemic or unfavourable conditions, which could have implications for public health communication strategies and policy-making.

CONCLUSION

The outcomes of the study reveal that when we ran the model with 70% testing and 30% test data; the LSTM performed well with 86%, accuracy, 88% precision, 85% recall, and 86% F-1 score in comparison with Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive Classifier, SVC Algorithm, Logistic Regression and RNN. When we apply Run 2 on the proposed model with 80% testing and 20% test data; again, LSTM performed well with 87.3% accuracy, 89% precision, 86% recall, and 87% F-1 score in comparison with Naive Bayes, Random Forest, MLP Classifier, Passive Aggressive Classifier, SVC Algorithm, Logistic Regression and RNN. From the second run of the model, it was observed the performance measures on the mentioned parameters were increased. During Run 3 of the proposed model with 90% testing and 10% test data; again, LSTM performed well with 88% accuracy, 88% precision, 87% recall, and 87% F-1 score. It has been recorded that on Run 3 of the model, the performance measures on the mentioned parameters were increased.

The results of the proposed model indicate that the performance of the model was increased during the second and third run respectively. Overall, the LSTM approach is found to be efficient for sentiment analysis on text data and can be easily implemented for sentiment classification of COVID-19 or related reviews. This study has both theoretical and practical implications. In terms of theoretical implications, this study employs machine and deep learning approaches for sentiment analysis on COVID-19 sentiments. The implications of this study are especially visible in recent COVID-19 cases when sentiments communicated on social media have influenced public mood [47].

The future work aligns with the parameter tuning on the LSTM technique to enhance the performance and validate the results. The LSTM technique can also be extended to sentiment classification to overcome decision-making problems in different industries dealing with text reviews or sentiments; including marketing, government, service, and academia. This study demonstrates that the proposed deep learning approach can be used and adapted to achieve a high degree of accuracy, especially when considering the intricacies involved in textual analysis.

Declarations:

Funding: No funding is received for this research.

Conflicts of interest: No conflicts of interest regarding the research or submitted research article.

Availability of data and material: Data is available within the manuscript/ any data/ code used in this may be made available by the authors in electronic form whenever required.

Code availability: (software application or custom code): Not Applicable

Authors' contributions: Mr Brajesh Kumar Shrivash[1] arranged the data and proposed the model, Dr. Dinesh Verma[1] compared the models, and Dr. Prateek Pandey[1] was responsible for model validation.

REFERENCES

1. Dang, Nhan Cach, María N. Moreno-García, and Fernando De la Prieta. "Sentiment analysis based on deep learning: A comparative study." *Electronics* 9, no. 3 (2020): 483.

2. Ainapure, Bharati Sanjay, Reshma Nitin Pise, Prathiba Reddy, Bhargav Appasani, Avireni Srinivasulu, Mohammad S. Khan, and Nicu Bizon. "Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches." *Sustainability* 15, no. 3 (2023): 2573.

3. Cambria, Erik, Soujanya Poria, Alexander Gelbukh, and Mike Thelwall. "Sentiment analysis is a big suitcase." *IEEE Intelligent Systems* 32, no. 6 (2017): 74-80.

4. Farhadloo, Mohsen, and Erik Rolland. "Fundamentals of sentiment analysis and its applications." *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence* (2016): 1-24.

5. Liang, Bin, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks." *Knowledge-Based Systems* 235 (2022): 107643.

6. Cambria, Erik, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. "Affective computing and sentiment analysis." *A practical guide to sentiment analysis* (2017): 1-10.

7. Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." *Journal of King Saud University-Engineering Sciences* 30, no. 4 (2018): 330-338.

8. Jin, Zhigang, Xiaofang Zhao, and Yuhong Liu. "Heterogeneous graph network embedding for sentiment analysis on social media." *Cognitive Computation* 13 (2021): 81-95.

9. Ahmed, Md Shoaib, Tanjim Taharat Aurpa, and Md Musfique Anwar. "Detecting sentiment dynamics and clusters of Twitter users for trending topics in COVID-19 pandemic." *Plos one* 16, no. 8 (2021): e0253300.

10. Chakraborty, Arunava Kumar, Sourav Das, and Anup Kumar Kolya. "Sentiment analysis of COVID-19 tweets using evolutionary classification-based LSTM model." In *Proceedings of Research and Applications in Artificial Intelligence: RAAI (2020)*, pp. 75-86. Singapore: Springer Singapore, 2021.

11. Chandrasekaran, Ranganathan, Vikalp Mehta, Tejali Valkunde, and Evangelos Moustakas. "Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study." *Journal of Medical Internet research* 22, no. 10 (2020): e22624.

12. Aygün, Irfan, Buket Kaya, and Mehmet Kaya. "Aspect Twitter sentiment analysis on vaccination and vaccine types in COVID-19 pandemic with deep learning." *IEEE Journal of Biomedical and Health Informatics* 26, no. 5 (2021): 2360-2369.

13. Tsai, Meng Hsiu, and Yingfeng Wang. "Analyzing Twitter data to evaluate people's attitudes towards public health policies and events in the era of COVID-19." *International Journal of Environmental Research and Public Health* 18, no. 12 (2021): 6272.

14. Crocamo, Cristina, Marco Viviani, Lorenzo Famiglini, Francesco Bartoli, Gabriella Pasi, and Giuseppe Carrà. "Surveilling COVID-19 emotional contagion on Twitter by sentiment analysis." *European Psychiatry* 64, no. 1 (2021): e17.

15. Priyadarshini, Ishaani, and Chase Cotton. "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis." *The Journal of Supercomputing* 77, no. 12 (2021): 13911-13932.

16. Shofiya, Carol, and Samina Abidi. "Sentiment analysis on COVID-19-related social distancing in Canada using Twitter data." *International Journal of Environmental Research and Public Health* 18, no. 11 (2021): 5993.

17. Huang, Faliang, Xuelong Li, Changan Yuan, Shichao Zhang, Jilian Zhang, and Shaojie Qiao. "Attention-emotion-enhanced convolutional LSTM for sentiment analysis." *IEEE Transactions on neural networks and learning systems* 33, no. 9 (2021): 4332-4345.

18. Shrivash, Brajesh Kumar, Dinesh Kumar Verma, and Prateek Pandey. "An Analysis on Machine Learning Approaches for Sentiment Analysis." In *Smart Systems: Innovations in Computing: Proceedings of SSIC 2021*, pp. 499-513. Springer Singapore, 2022.

19. Priyadarshini, Ishaani, Pinaki Mohanty, Raghvendra Kumar, Rohit Sharma, Vikram Puri, and Pradeep Kumar Singh. "A study on the sentiments and psychology of Twitter users during COVID-19 lockdown period." *Multimedia Tools and Applications* 81, no. 19 (2022): 27009-27031.

20. Zhao, Narisa, Huan Gao, Xin Wen, and Hui Li. "Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis." *IEEE Access* 9 (2021): 15561-15569.

21. Lu, Qiang, Zhenfang Zhu, Guangyuan Zhang, Shiyong Kang, and Peiyu Liu. "Aspect-gated graph convolutional networks for aspect-based sentiment analysis." *Applied Intelligence* 51, no. 7 (2021): 4408-4419.
22. Tubishat, Mohammad, Norisma Idris, and Mohammad Abushariah. "Explicit aspects extraction in sentiment analysis using optimal rules combination." *Future Generation Computer Systems* 114 (2021): 448-480.
23. Pota, Marco, Mirko Ventura, Rosario Catelli, and Massimo Esposito. "An effective BERT-based pipeline for Twitter sentiment analysis: A case study in Italian." *Sensors* 21, no. 1 (2020): 133.
24. Srividya, Kotagiri, and A. Mary Sowjanya. "WITHDRAWN: NA-DLSTM–A neural attention-based model for context aware Aspect-based sentiment analysis." (2021).
25. Basiri, Mohammad Ehsan, Shahla Nemati, Moloud Abdar, Erik Cambria, and U. Rajendra Acharya. "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis." *Future Generation Computer Systems* 115 (2021): 279-294.
26. Kandasamy, Ilanthenral, W. B. Vasantha, Jagan M. Obbineni, and Florentin Smarandache. "Sentiment analysis of tweets using refined neutrosophic sets." *Computers in Industry* 115 (2020): 103180.
27. Dang, Nhan Cach, María N. Moreno-García, and Fernando De la Prieta. "Sentiment analysis based on deep learning: A comparative study." *Electronics* 9, no. 3 (2020): 483.
28. Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, no. 4 (2018): e1253.
29. Abd-Alrazaq, Alaa, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. "Top concerns of tweeters during the COVID-19 pandemic: infoveillance study." *Journal of medical Internet research* 22, no. 4 (2020): e19016.
30. Dubey, Akash Dutt. "Twitter sentiment analysis during COVID-19 outbreak." *Available at SSRN 3572023* (2020).
31. Barkur, Gopalkrishna, and Giridhar B. Kamath. "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India." *Asian journal of psychiatry* 51 (2020): 102089.
32. Linka, Kevin, Mathias Peirlinck, Francisco Sahli Costabal, and Ellen Kuhl. "Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions." *Computer methods in biomechanics and biomedical engineering* 23, no. 11 (2020): 710-717.
33. Fanelli, Duccio, and Francesco Piazza. "Analysis and forecast of COVID-19 spreading in China, Italy and France." *Chaos, Solitons & Fractals* 134 (2020): 109761.
34. Pastor, Cherish Kay Landicho. "Sentiment analysis on synchronous online delivery of instruction due to extreme community quarantine in the Philippines caused by COVID-19 pandemic." *Asian Journal of Multidisciplinary Studies* 3, no. 1 (2020): 1-6.
35. Dhar, Bablu Kumar, Foster Kofi Ayittey, and Sabrina Maria Sarkar. "Impact of COVID-19 on Psychology among the University Students." *Global Challenges* 4, no. 11 (2020): 2000038.
36. Dubey, Akash Dutt. "Public sentiment analysis of COVID-19 vaccination drive in India." *Available at SSRN 3772401* (2021).
37. Alam, Kazi Nabiul, Md Shakib Khan, Abdur Rab Dhruba, Mohammad Monirujjaman Khan, Jehad F. Al-Amri, Mehedi Masud, and Majdi Rawashdeh. "Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data." *Computational and Mathematical Methods in Medicine* 2021 (2021).
38. Chintalapudi, Nalini, Gopi Battineni, and Francesco Amenta. "Sentimental analysis of COVID-19 tweets using deep learning models." *Infectious disease reports* 13, no. 2 (2021): 329-339.
39. Chandra, Rohitash, and Aswin Krishna. "COVID-19 sentiment analysis via deep learning during the rise of novel cases." *PloS one* 16, no. 8 (2021): e0255615.
40. Storey, Veda C., and Daniel E. O'Leary. "Text analysis of evolving emotions and sentiments in COVID-19 Twitter communication." *Cognitive Computation* (2022): 1-24.
41. Rustam, Furqan, Madiha Khalid, Waqar Aslam, Vaibhav Rupapara, Arif Mehmood, and Gyu Sang Choi. "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis." *Plos one* 16, no. 2 (2021): e0245909.

42. Singh, Chetanpal, Tasadduq Imam, Santoso Wibowo, and Srimannarayana Grandhi. "A deep learning approach for sentiment analysis of COVID-19 reviews." *Applied Sciences* 12, no. 8 (2022): 3709.
43. Wang, Jianghao, Yichun Fan, Juan Palacios, Yuchen Chai, Nicolas Guetta-Jeanrenaud, Nick Obradovich, Chenghu Zhou, and Siqi Zheng. "Global evidence of expressed sentiment alterations during the COVID-19 pandemic." *Nature Human Behaviour* 6, no. 3 (2022): 349-358.
44. Wankhade, Mayur, and Annavarapu Chandra Sekhara Rao. "Opinion analysis and aspect understanding during covid-19 pandemic using BERT-Bi-LSTM ensemble method." *Scientific Reports* 12, no. 1 (2022): 17095.
45. Chaudhary, Meghna, Kristin Kosyluk, Sylvia Thomas, and Tempestt Neal. "On the use of aspect-based sentiment analysis of Twitter data to explore the experiences of African Americans during COVID-19." *Scientific Reports* 13, no. 1 (2023): 10694.
46. Bordoloi, Monali, and Saroj Kumar Biswas. "Sentiment analysis: A survey on design framework, applications and future scopes." *Artificial Intelligence Review* (2023): 1-56.
47. Depoux, Anneliese, Sam Martin, Emilie Karafillakis, Raman Preet, Annelies Wilder-Smith, and Heidi Larson. "The pandemic of social media panic travels faster than the COVID-19 outbreak." Journal of travel medicine 27, no. 3 (2020): taaa031.