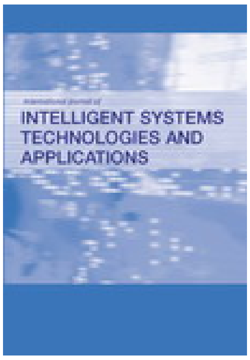


International Journal of Intelligent Systems Technologies and Applications

ISSN online: 1740-8873 - ISSN print: 1740-8865
https://www.inderscience.com/ijista

Integration of deep learning techniques for sentiment and emotion analysis of social media data

H.S. Hota, Dinesh K. Sharma, Nilesh Verma

DOI: 10.1504/IJISTA.2023.10055771

Article History:

Received:	28 February 2022
Last revised:	02 September 2022
Accepted:	08 September 2022
Published online:	27 April 2023

Integration of deep learning techniques for sentiment and emotion analysis of social media data

H.S. Hota*

Department of Computer Science and Application,
Atal Bihari Vajpayee University, India
Email: proffhota@gmail.com
*Corresponding author

Dinesh K. Sharma

Department of Business,
Management and Accounting,
University of Maryland Eastern Shore,
Princess Anne, MD 21853, USA
Email: profdksharma@gmail.com

Nilesh Verma

Department of Computer Science and Application,
Atal Bihari Vajpayee University, India
Email: me@nileshverma.com

Abstract: Sentiment analysis (SA) and emotion analysis (EA) are commonly used to understand people's feelings and opinions on a given topic. COVID-19 is an emerging infectious disease that is rapidly spreading around the world. The mental state of a country's population is more or less the same worldwide. Machine learning (ML) techniques are commonly utilised to analyse human sentiments and emotions. Two popular deep learning (DL) techniques: convolutional neural network (CNN) and long short-term memory (LSTM) are being applied in several areas. In this study, we propose a hybrid of CNN and LSTM to improve the performance of the classification model. The two different models, the sentiment analysis model (SAM) and the emotional analysis model (EAM), were developed using benchmark data, which produces 91.11% and 89.39% accuracy, respectively, by integrating CNN and LSTM. Integration of two or more techniques significantly improves performance by utilising both techniques. The results of the experiments demonstrate that the proposed hybrid technique outperforms other individual DL techniques.

Keywords: convolutional neural network; CNN; sentiment analysis; SA; emotion analysis; EA; COVID-19; deep learning; DL; long short-term memory; LSTM.

Reference to this paper should be made as follows: Hota, H.S., Sharma, D.K. and Verma, N. (2023) 'Integration of deep learning techniques for sentiment and emotion analysis of social media data', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 21, No. 1, pp.1–20.

Biographical notes: H.S. Hota is working as a Professor in the Department of Computer Science and Application of Atal Bihari Bajpayee University, India. He earned MCA and PhD degrees in Computer Science and published more than 50-refereed papers in reputed journals along with one book. He has also delivered talks at international conferences and presented papers at many international and national conferences.

Dinesh K. Sharma is a Professor of Quantitative Methods and Computer Applications in the Department of Business, Management and Accounting at the University of Maryland Eastern Shore. He earned his MS in Mathematics, MS in Computer Science, PhD in Operations Research, and a second PhD in Management. He has published over 250 refereed journal articles and conference proceedings and has won sixteen best paper awards. He is the Editor-in-Chief of the *Journal of Global Information Technology* and *Review of Business and Technology Research*.

Nilesh Verma is a data scientist and AI researcher with more than two years of industry working experience. He has also developed three useful Python libraries with more than 35,000+ downloads. He holds an MSc and BSc degrees in Computer Science and Applications from Atal Bihari Vajpayee University in Bilaspur, India. His research interest is machine learning, deep learning, computer vision, and natural language processing. He has also published research papers in reputed journals.

1 Introduction

Sentiment analysis (SA) has sparked great interest in both academic research and industrial and commercial corporations. SA (Al-Smadi et al., 2018) is a natural language processing (NLP) task that extracts users' sentiments from a piece of text. SA has recently become a popular field of study, mainly when dealing with subjectivity in textual content (Basiri et al., 2020; Jiménez-Zafra et al., 2019). SA encompasses a wide range of disciplines, including finance, business, health sciences, and politics, to name a few (Basiri et al., 2020; Day and Lee, 2016; Liu and Shen, 2020; Zhang et al., 2018; Medhat et al., 2014). SA is used to conduct in-depth investigations of human sentiment in terms of distinct emotions such as happiness, disgust, guilt, melancholy, etc.

Advances in NLP approaches have improved how SA systems mine user-generated data (Basiri et al., 2020). SA entails categorising text into multiple categories, such as positive and negative feelings. Using SA with social media data has become a popular data analysis tool. Twitter is a well-known and widely utilised social media site that can aid in disseminating ideas. It is also open to the public and one of the best data sources for SA. Twitter continues to play an important role in rapidly disseminating information about any social crisis or tragedy to its millions of active users. However, Twitter has been criticised numerous times for whether its data can be considered an authentic source for real-world problem-solving. Tang et al. (2018) have explored and reviewed the relationship between tweets' credibility and many emerging infectious diseases. Furthermore, Lim and Tucker (2019) used Twitter data to present an intelligence system that identifies groups of terms with a causal relationship to real-world enterprise outcomes. Devi and Karthika (2019) have also considered the credibility of tweets associated with natural disasters. They then identified the tweets that spread false information and proposed a credibility-analysis system.

Several authors have investigated machine learning (ML) techniques in various SA-related domains. Deep learning (DL), one of the most well-known ML approaches, is commonly utilised for machine vision as well as sentiment and emotion analysis (EA). Araque et al. (2017) proposed DL methods that apply to SA, with different hybrid techniques in social applications. Brijali et al. (2017) applied ML and semantic SA-based algorithms for suicide sentiment prediction in social networks. Additionally, Bansal and Shrivastava (2018) predicted elections using a hybrid topic-based SA of tweets. Furthermore, Nishar and Yeung (2018), and Grob-Klubmann (2019) used ML methods in the financial domain, specifically for stock market predictions. Katayama et al. (2019) have studied sentiment polarity identification in financial news using DL. Chakraborty et al. (2019) applied DL methods for SA to a list of movie reviews, whereas Abdi et al. (2019) presented DL sentiment classification of evaluative text using multi-feature fusion. In addition, Soumya and Pramod (2020) applied several ML techniques to the SA of Malayalam tweets. Ruz et al. (2020) recently studied ML methods for SA during earthquakes, social movements, and other critical events.

The hybrid approach combines multiple classifiers into a single classifier to increase classifier performance. Researchers frequently employ a hybrid of ML for SA. Pandey et al. (2017), for example, utilised a metaheuristic-based technique for locating appropriate cluster-heads from the sentiment content of the Twitter dataset using K-means and Cuckoo searches. Additionally, to detect Twitter emotions, Felbo et al. (2017) used a DL variant of the long short-term memory (LSTM) model that has been successful in numerous NLP tasks. Next, Ankit (2018) presented an ensemble classification system for SA that employed three classifiers: logistic regression, support vector machine (SVM), and random forest. On several Twitter topics, Hassonah et al. (2020) employed a combination of SVM and an evolutionary technique for SA. Simultaneously, Kumar et al. (2020) presented a hybrid DL model for fine-grained sentiment prediction using textual and visual semiotic modalities in social data. On the other hand, Barkur and Vibha (2020) used Twitter data during the lockdown to examine Indians' sentiments towards the COVID-19 pandemic. Recently, Hota et al. (2021) explored SA by assessing the sentiment of six countries with a Lexicon-based methodology.

In the literature, convolutional neural networks (CNN), recurrent neural networks (RNN), and auto-encoders have been applied to DL. RNN is the most commonly used algorithm for sentiment data, with modifications such as LSTM. Also, CNN and LSTM have been widely employed in various fields. To combine CNN and LSTM, ensemble models are employed. The activations of their inner layers are utilised as characteristics in subsequent models (Kiran et al., 2020; Hota et al., 2021). The primary goal of ensemble methods is to combine a set of models (base classifiers) to obtain a more accurate and competent model than a single model, which cannot be as precise (Arque et al., 2017). Poria et al. (2017a) used CNN to extract features from video, audio, and text. They subsequently employed multiple-kernel learning (MKL) for SA. Furthermore, Poria et al. (2017b) improved the multimodal SA ensemble of CNN and MKL. Alharbi et al. (2019) employed the ensemble methodology of DL to analyse tweets and user behaviour. However, many authors have identified issues with DL for SA and are working to find better solutions and techniques. Minaee et al. (2019) suggested a DL model for SA based on a CNN and LSTM ensemble. For aspect-based SA, Akhtar et al.

(2020) have done aspect-based SA using the DL technique. Song et al. (2019) used LSTM for aspect-based SA. For aspect-based SA of Arabic hotels' reviews, Al-Smadi et al. (2018) compared deep RNN with SVM, while Heikal et al. (2018) used DL to analyse Arabic tweets for SA. Singh et al. (2018a) evaluated Punjabi text morphology and SA using DL. Yousif et al. (2019) used a recurrent CNN for SA, while Zhang et al. (2018) used CNN for SA as well. DL techniques have been employed for SA in pandemics (Haihong et al., 2019) and EA (Sailunaz and Alhajj, 2019). Kiran et al. (2020) proposed OSLC fit, a DL solution for sentiment polarity classification evaluations. Researchers have also carried out language-specific SA. For example, authors (Ghulam et al., 2019; Mahmood et al., 2020) have studied SA on a Roman Urdu dataset.

Due to the limitations of each model, combining two or more techniques to produce better prediction results is becoming more common. CNN models are typically used to extract features from image data, whereas LSTM models are commonly used to extract features from time series data. LSTM is incapable of extracting features from image data, but CNN is incapable of remembering previous information. As a result, hybridisation of these two is required to capture the strength of both. In this study, we integrate two DL approaches: CNN and LSTM, with the LSTM network, put before the CNN dense layer. In terms of performance metrics, this hybrid model outperforms individual models. The hybrid model is important because it combines features from each individual model.

This study aims to analyse citizens' sentiments and emotions using two DL techniques: CNN and LSTM, combined to maximise the benefits of both for analysing NLP data, such as COVID-19 tweets for SA and EA. The main objective of this study is to build an intelligent system based on a hybrid of CNN and LSTM to enhance the performance of the classification model. The experimental results show that a hybrid of CNN and LSTM outperforms compared to Simple RNN, LSTM, CNN, GRU, and Bidirectional LSTM DL techniques. The proposed system can progress over time without any changes needed. To demonstrate the system's effectiveness, we used COVID-19-related Twitter data for Sentiment and EA in six countries: India, the USA, France, the UK, Spain, and Italy. To better comprehend human emotions, SA was done as positive and negative sentiments, while EA was done as happy, sad, disgust, and so on. Our results show that citizens of all countries face problems due to COVID-19 and their sentiments and emotions are in negative directions due to the pandemic.

The remaining portion of the paper is formed as follows: Section 2 describes the proposed research work. Section 3 explains the experimental data utilised for model construction as well as SA and EA. Section 4 investigates the experimental design and analysis results, and Section 5 ends the research paper.

2 Proposed work

A process flow diagram of the entire research work is depicted in Figure 1. As shown in the figure, the work is divided into four phases: acquisition, pre-processing, model development, and experiment and evaluation. The work started with the acquisition of data from different sources. Data is collected for model development as well as for SA and EA. Twitter data is extracted for SA and EA from the Twitter site. The second phase deals with the pre-processing of textual data as it contains noise in many forms and is necessary to remove to improve classifiers' performance. The third phase explores DL-based model development using a hybrid of two prevalent DL techniques: CNN and

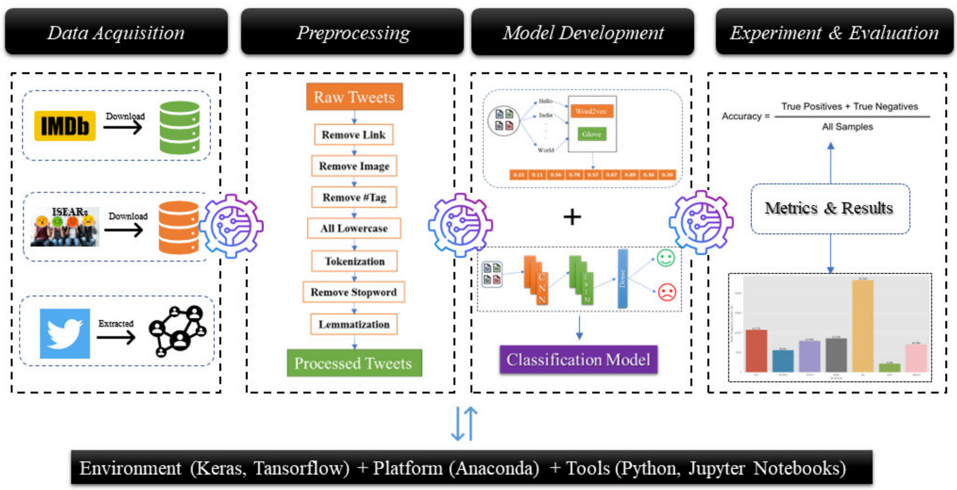
LSTM, with training and testing samples. This phase can be viewed in two folds, as below:

1 Sentiment analysis model (SAM): A binary classification model was developed to classify and analyse sentiment as positive or negative.

2 (EAM): A multi-class classification model was developed to classify emotions as anger, disgust, fear, guilt, sadness, or happiness.

The last phase analyses the models based on various matrices and finally evaluates sentiment and emotion with respective class labels. For the entire experimental work, especially the Keras python library was used. Sections 3 and 4 will explain each phase of the block diagram, as shown in Figure 1 in more detail.

Figure 1 A block diagram of the proposed work for sentiment and emotion analysis (see online version for colours)



3 Material and methods

This section describes the material (data) used for the model development as well as data for SA and EA. As stated in Section 2, two models, namely SAM and EAM, have been developed using benchmark data to classify sentiment and emotions. Finally, Twitter data is used for SA and EA. Apart from this, the section also describes DL techniques in more detail. The details of materials and methods are explained in the following subsections:

3.1 Experimental dataset

In ML, data plays a crucial role, as two different types of analysis with different class levels, binary and multi-class levels, were proposed in this research work. As a result, similar data types are required to develop the model. Following the development of the models using benchmark data, more Twitter data was downloaded for sentiment and EA. Each dataset's specifics are discussed in greater detail below.

3.1.1 Data for development of SAM

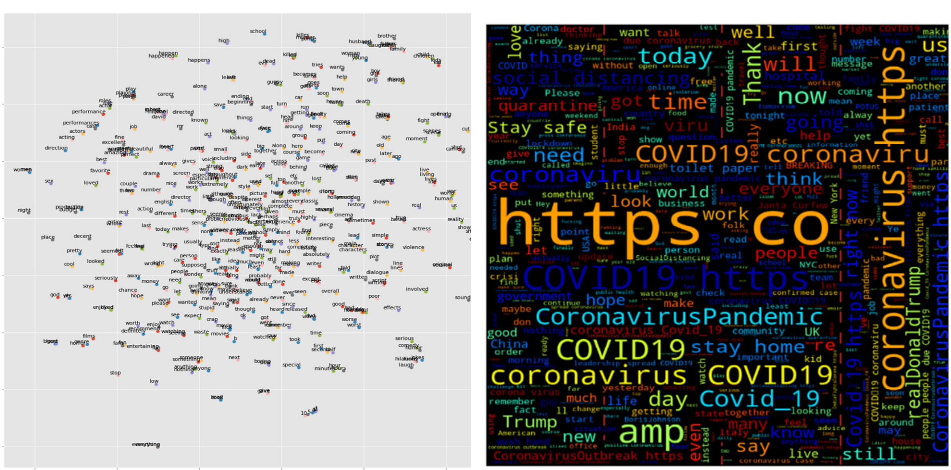
Many benchmark datasets are available to build binary models to analyse sentiment as positive or negative; VADER, STS-Gold, and sentiment140 IMDB are a few. To develop a robust DL-based classifier, and due to computational limitations, we have used the IMDB dataset downloaded from Maas et al. (2011). Many authors used this dataset (Haihong et al., 2019). Other datasets either consist of vast numbers of samples or fewer samples. For example, the sentiment 140 dataset consists of 160,000 data, while STS-Gold consists of only 2034 samples. This dataset consists of 500,000 samples equally divided into positive and negative samples. The dataset is well balanced for developing a DL-based model for SA. As shown in Figure 2, the dataset has two rows: the tweet and its sentiment class as positive or negative. Figure 3 shows the word cloud data.

Figure 2 A sample benchmark data used for development of SAM

	text	target
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive
...	...	...
49995	I thought this movie did a down right good job...	positive
49996	Bad plot, bad dialogue, bad acting, idiotic di...	negative
49997	I am a Catholic taught in parochial elementary...	negative
49998	I'm going to have to disagree with the previou...	negative
49999	No one expects the Star Trek movies to be high...	negative

50000 rows × 2 columns

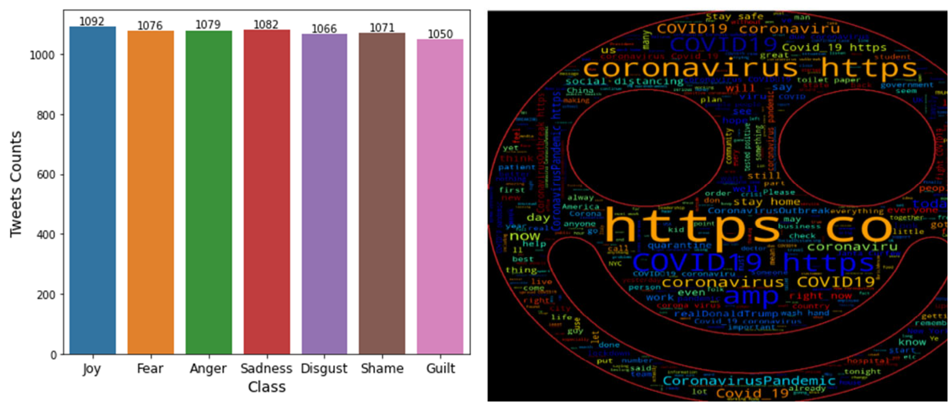
Figure 3 Word cloud of data (see online version for colours)



3.1.2 Data for development of EAM

An emotional analysis-based model is needed in order to analyse the emotions of a human being. The data used for developing this type of model consists of multiple class labels: joy, fear, anger, sadness, disgust, shame, and guilt, and hence the developed model is a multi-class classifier. A popular dataset known as ISEAR, consisting of all the above class labels, is also downloaded (Dan-Glauser and Scherer, 2013). A total of 7,652 label-wise samples and their word clouds are shown in Figure 4. The dataset consists of seven class labels: anger, sadness, disgust, shame, fear, joy, and guilt, with all the most equally divided sample sizes (1091 to 1094 samples in each class).

Figure 4 (a) Detail of samples (left) and (b) word cloud (right) of data used for the development of EAM (see online version for colours)



(a) (b)

3.1.3 Twitter data

Twitter data is widely used for a variety of purposes, including product, service, and (Araque et al., 2017) financial analysis, political forecasting, and SA and EA. People often use Twitter to reflect their opinion on any issue. The COVID-19 pandemic has also attracted Twitter to share sentiments and emotions with the people. # COVID-19 and # coronavirus are two popular hashtags where most opinions about the COVID-19 pandemic are posted. Tweets with these two hashtags are therefore used for SA and EA. A total of 65,329 tweets from the five most affected COVID-19 countries and India from April 15 to May 15, 2020, were extracted from Twitter. Twitter permits a maximum of 240 characters per tweet to anyone, including emoji and special characters.

Figure 5 shows a word cloud of extracted tweets showing the most frequent words in the tweets. On the other hand, Figure 6 shows the country-wise maximum length of the tweet and the country-wise total number of tweets extracted and used for SA and EA. Figure 7 shows the hashtag-wise frequency of tweets; the word coronavirus has the highest frequency in this figure, followed by COVID-19, lockdown, stay home, and so on.

Figure 5 Word cloud of twitter data extracted for SA and EA (see online version for colours)

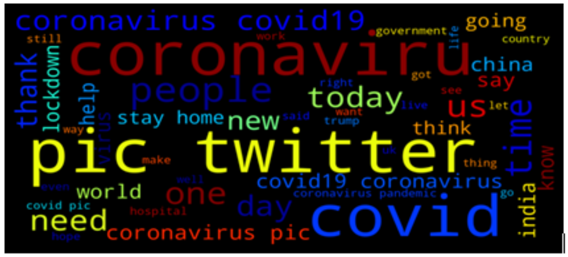
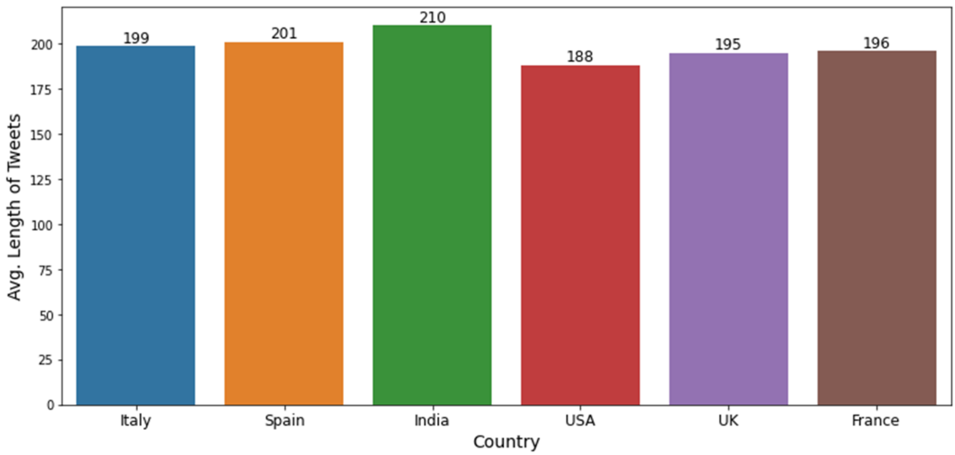
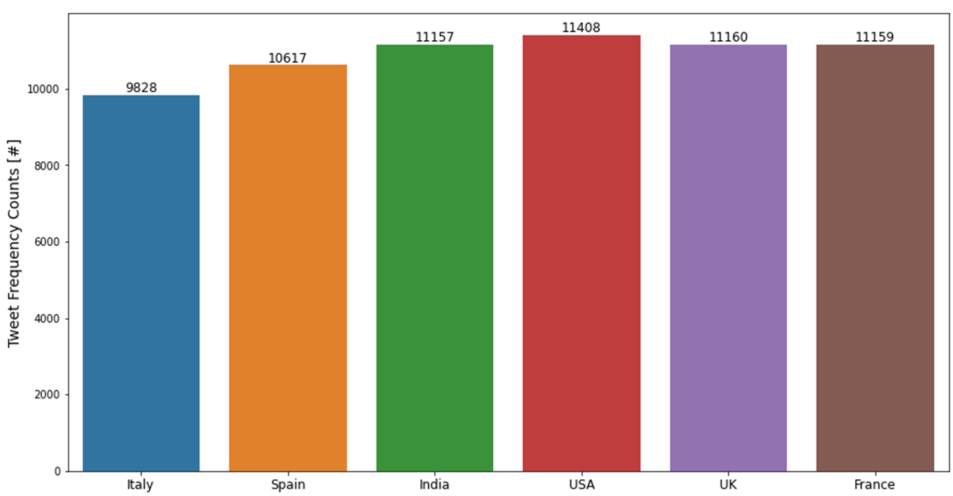


Figure 6 (a) Country wise maximum length of tweet (left) and (b) country wise total number of tweet (right) (see online version for colours)

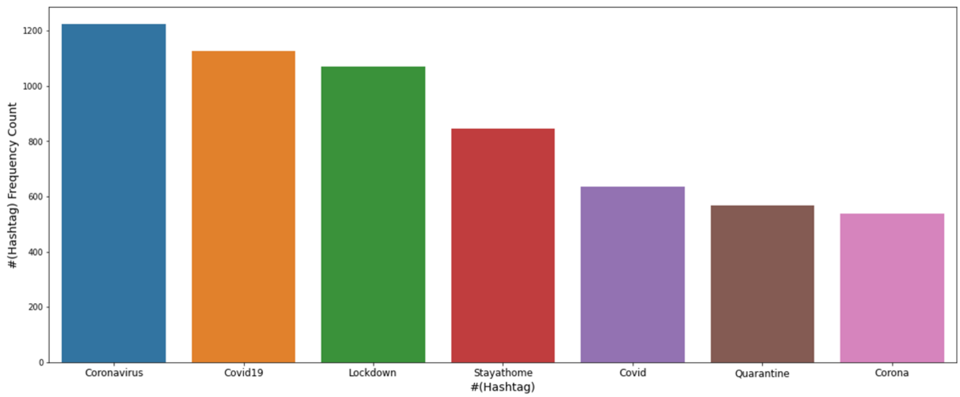


(a)



(b)

Figure 7 #hashtag wise frequency of tweets (see online version for colours)



3.1.4 Evaluation matrix

The following matrices are commonly used to assess the effectiveness of any classification model, where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

1 Accuracy: Accuracy determines the number of points correctly classified out of all points.

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

2 Recall: This is also known as sensitivity, and it represents the probability of correctly identifying a positive sample among all existing positive samples.

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

3 Precision: The precision with which a model's positive prediction is made.

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

4 Specificity: Indicates the frequency of correctly identified negative samples.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

5 F1 – score: It provides a balanced value for precision and recall. The harmonic mean of precision and recall is used to calculate the F1 score.

$$\text{F1 score} = \frac{2\times(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

6 ROC – AUC curve: ROC is an abbreviation for receiver operating characteristic curve, and AUC is an abbreviation for area under the ROC curve. The ROC-AUC

curve measures model performance at various threshold settings. TPR and FPR are two parameters plotted on the ROC curve (Bradley, 1997).

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

3.2 Data pre-processing

Data pre-processing is an important and essential step, especially in the case of Twitter data. Twitter data (Singh and Kumari, 2016) is highly unstructured, small message data containing emoji, hashtags, stop words, unidentified words, symbols, abbreviations, etc. As noise and users create this data with their own shortcut words and spelling, making tweets so complicated, it is challenging to remove these and pre-process them before using them for ML-based classification tasks. Singh and Kumari (2016) investigated the function of text pre-processing in Twitter SA. In addition, data pre-processing was utilised in this study to remove a lot of unnecessary information from Twitter data. Many other studies have revealed that the pre-processing of Twitter data certainly enhances the accuracy of classifiers. In general, five steps (Hota et al., 2021) are: removing noise, changing to lower case, tokenisation, removing stop words, and lemmatisation are used to pre-process textual data.

3.3 Deep learning approach

DL is a widely used artificial neural network (ANN) and also a popular ML technique nowadays. It offers automated feature extraction as well as a more prosperous representation capability and improved performance over conventional feature-based techniques. DL and its variants are used to solve complex problems. While developing the DL-based model, we optimised various DL parameters like dropout layer values, Conv1D filters, kernel size, MaxPooling1D kernel size, LSTM layer, unit size, activation function sigmoid, Relu and learning rate. The two popular techniques of DL, along with its hybrid, are explained below:

3.3.1 CNN

The CNN is the most extensively used ANN and belongs to the feed-forward type of Neural Network. The filtering and encoding of data through transformation is referred to as convolution in CNN. Each network layer functions as a detection filter on the raw data, looking for specific traits or patterns (Kumar et al., 2020). The various layers of CNN are as follow:

3.3.2 Word embedding

The embedding layer connects the input layer and the convolutional layer. As shown in Figure 8, the CNN hierarchically learns feature representation and extraction using word embedding and converts the word into a vector of integers (Kumar et al., 2020). Our data is now in the form of integer numbers, which makes CNN easy to learn. There are many methods of word embedding, but the two most popular methods are Word2Vec and GloVe.

Figure 8 Process of word embedding (see online version for colours)

In this research work, we have used both of them. The Word2Vec method converts the Twitter data into a numeric vector.

- Convolutional layer: Filtering is an essential operation of the convolutional layer. This layer extracts the feature from the original matrix by sliding from top to bottom and left to right (Luan and Lin, 2019).

Apart from this, CNN has a fully connected layer, also called a dense layer, a feed forward layer, and is used for classification of data.

3.3.3 LSTM

The LSTM is an RNN that can learn long-term dependencies. To accomplish this, the cell state at the heart of LSTM can selectively enable information to flow through the door mechanism while adding or removing information from cells. A forget gate, an input gate, and an output gate are the three gates that make up an LSTM. The forget gate determines which data should be excluded from the cell state, whereas the input gate determines which data should be included (Luan and Lin, 2019). The cell state can be changed after these two points have been determined. Finally, the output gate decides the ultimate output of the network. NLP tasks have outperformed due to its memory's ability to collect long-term dependencies in sequential data. The output of LSTM has been proven to be superior to that of other methods like Lexicon and ML (Ghulam et al., 2019).

3.3.4 Simple RNN

RNN (Bhowmik et al., 2022) is a feed forward neural network for sequence modelling and data analysis in which the output is determined by the previous state. It keeps track of new state information as it iterates through the sequence of elements and feeds it back to the previous layer to capture the relationship between the current and previous time steps.

3.3.5 LSTM

LSTM is advanced form of RNN which is specifically intended for sequential modelling and is used for text data. There are two type of LSTM: unidirectional LSTM and Bidirectional LSTM. The LSTM is an RNN that can learn long-term dependencies. To accomplish this, the cell state at the heart of LSTM can selectively enable information to flow through the door mechanism while adding or removing information from cells. A forget gate, an input gate, and an output gate are the three gates that make up an LSTM. The forget gate determines which data should be excluded from the cell state, whereas the input gate determines which data should be included (Luan and Lin, 2019). The cell state can be changed after these two points have been determined. Finally, the output gate decides the ultimate output of the network. NLP tasks have outperformed due to its memory's ability to collect long-term dependencies in sequential data. The output of LSTM has been proven to be superior to that of other methods like Lexicon and ML (Ghulam et al., 2019).

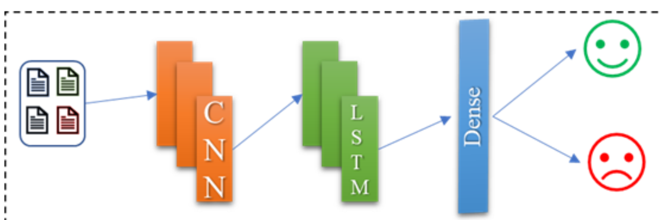
3.3.6 Bidirectional LSTM

The Bi-LSTM is similar to the LSTM architecture in that it works on inputs in two ways: one from left to right (capturing forward context) and the other from right to left (capturing backward context). In sequential modelling, it detects features from both the past and future contexts (Bhowmik et al., 2022).

3.3.7 GRU

Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks. The GRU is similar to a LSTM with a forget gate, but it has fewer parameters because it lacks an output gate.

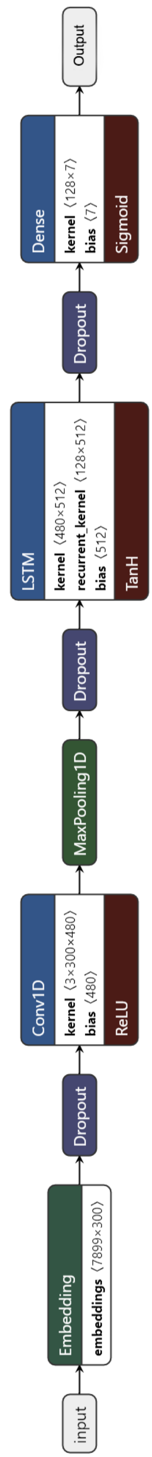
Figure 9 Block diagram of hybrid of CNN and LSTM (see online version for colours)



3.3.8 Hybrid of CNN and LSTM

Combining more than one model as a base classifier is a good practice to improve the classifier's performance, producing an accurate and reliable output compared to the individual one (Araque et al., 2017). This research work proposes a combination of two DL techniques, CNN and LSTM, to improve the performance. A simple block diagram of a hybrid of CNN and LSTM is shown in Figure 9. In which Twitter data is provided to CNN. Then, the convolutional layer results are sent to the LSTM layer, and the last dense layer is used to classify sentiments or emotions into predefined class labels. Figure 10 explains how this hybrid model was used for SA and EA. This model consists of various layers: the embedding layer, the dropout layer, and one 1D convolution layer. The main 1D convolutional layer is used for NLP, while the 2D convolutional layer is used for image processing. This convolution layer is followed by the max polling layer and then again by the dropout layer. The output produced by CNN is then provided to LSTM, and finally, dense layers classify tweets as per desired sentiment and emotion.

Figure 10 Detailed block diagram of hybrid of CNN and LSTM (see online version for colours)



4 Experimental design and result discussion

Experiments were carried out using the Python programming language and many of its libraries, as Python provides a dynamic library for deep learning. The following sections explain the details of the development process of models for sentiment and EA.

4.1 Development of SAM

As explained in Section 2, the IMBD dataset was used to train the model. The dataset is divided into training and testing in an 80%–20% ratio. The vocabulary size was set at 28,059, while 300 dimensions were set for the embedding layer. The model was trained with Word2Vec word embedding and other default hyper-parameters for 25 epochs. The experimental results obtained based on five different DL classifiers and a hybrid of CNN and LSTM are shown in Table 1. The same is also shown in the form of a comparative bar graph in Figure 11. This table shows that the hybrid of CNN and LSTM performs better than the individual classifiers and reflects that the hybrid model produces higher accuracy than others. All the measures in the case of the hybrid model were recorded higher than the others. The developed model was then used to analyse sentiment. The extracted Twitter data was fed into a CNN/LSTM hybrid model for SA. The obtained result in the form of positive and negative is shown in Figure 12.

Table 1 Experimental results in case of SAM

Model	*Accuracy*	*Sensitivity/ recall*	*Specificity*	*ROC_AUC*	*Precision*	*F-score*
Simple RNN	82.16	78.40	86.95	82.67	88.44	83.12
LSTM	89.98	89.73	90.23	89.98	90.13	89.92
CNN	90.09	89.80	90.38	90.09	90.29	90.05
GRU	89.30	88.10	90.55	89.33	90.69	89.38
Bidirectional LSTM	89.19	88.94	89.42	89.19	89.33	89.14
CNN+LSTM	91.11	91.06	91.16	91.11	91.03	91.05

In contrast, country-wise positive and negative sentiments are depicted in Figure 13. The overall negativity and positivity are shown in Figure 12, with 28.96% and 71.04% accuracy. If we observe Figure 13, it reflects that the USA has the highest negativity with 35.28%, followed by the UK (30.91%) and France (30.42%). On the other hand, India has a 27.10% negativity rate, higher than Spain (25.26%) and Italy (23.87%).

4.2 Development of EAM

Similar to Section 4.1, another model named EAM was developed to analyse emotions more intensely. For this purpose, the dataset explained in Section 3.1.2 was used with all default hyper-parameters, but GloVe word embedding was used instead of Word2Vec along with the transfer learning function. The activation function is taken as sigmoidal, the batch size is 512, and the epochs are 25. For training, 80% of the samples were used, while the remaining 20% were used to test the model. The experimental results obtained after successful training and testing of the models are shown in Table 3, reflecting that the hybrid of CNN and LSTM performs better than others, with an accuracy of 89.39% and the least loss of 0.295%. Based on these results, the hybrid model was selected for EA. Once again, the same Twitter data is provided to the model for EA. The forms of various emotions are shown in Figure 14 for all the countries together, while country-wise emotions are also shown in Figure 15. Of the seven emotions, six (anger, sadness, disgust, shame, fear, joy, and guilt) can be considered negative emotions, and the remaining one can be considered a positive emotion. The comparative bar graph of Figure 14 also reflects that all the countries' sentiments are positive in the form of joy rather than negative. For example, say, Italy and Spain have a higher percentage of joy, while the USA and the UK have a lower percentage of joy.

Figure 11 Comparative results of deep learning-based classification models for SA (see online version for colours)

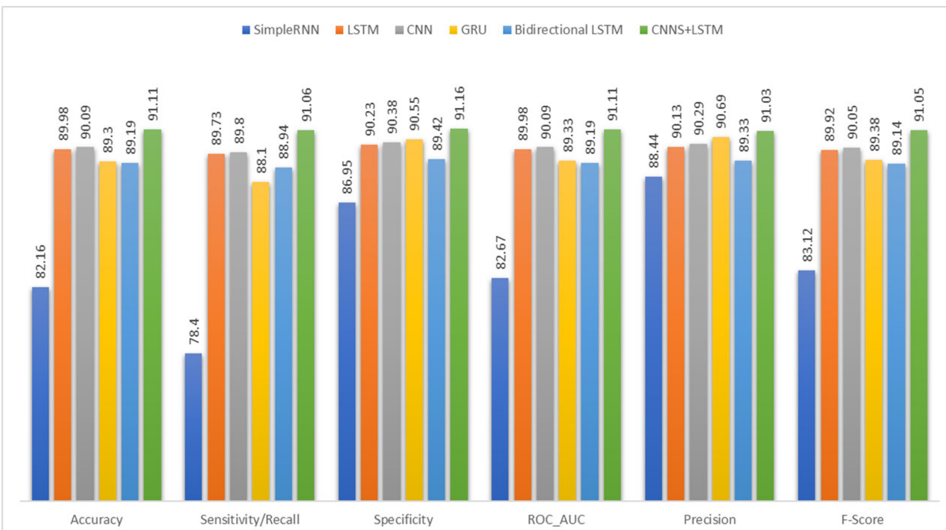


Figure 12 Overall sentiment as negative and positive (see online version for colours)

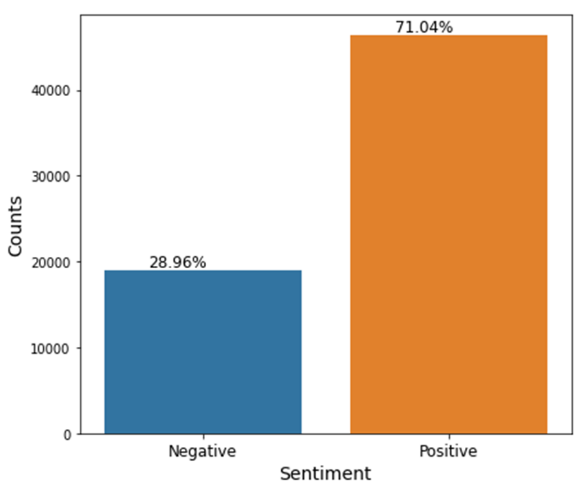


Figure 13 Country wise sentiment as negative and positive (see online version for colours)

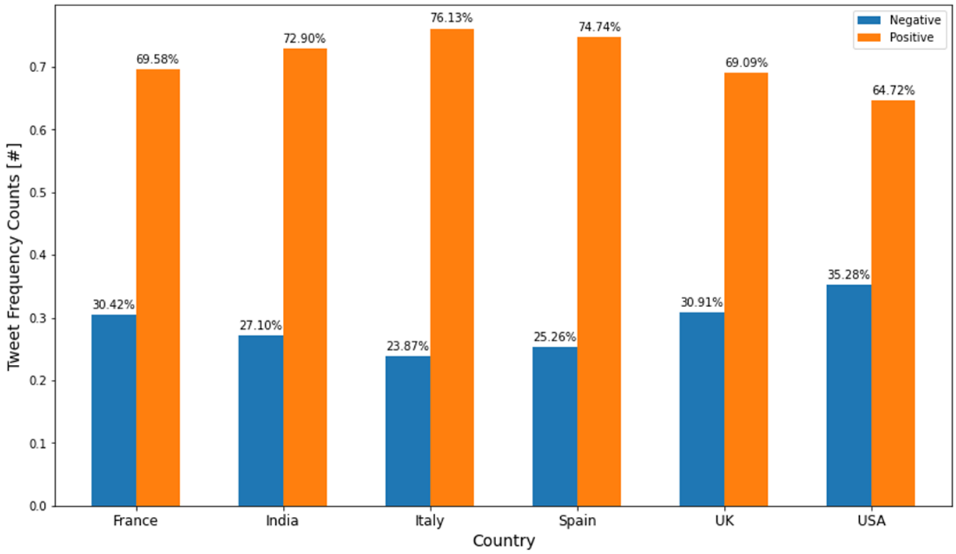


Table 2 Experimental results in case of EAM

Model	*Accuracy*	*Loss*
Simple RNN	86.07%	0.39311
LSTM	89.03%	0.30431
CNN	88.67%	0.33298
GRU	86.23%	0.37479
Bidirectional LSTM	89.17%	0.30694
CNN+LSTM	*89.39 %*	*0.29564*

Figure 14 Overall emotion of all the countries together (see online version for colours)

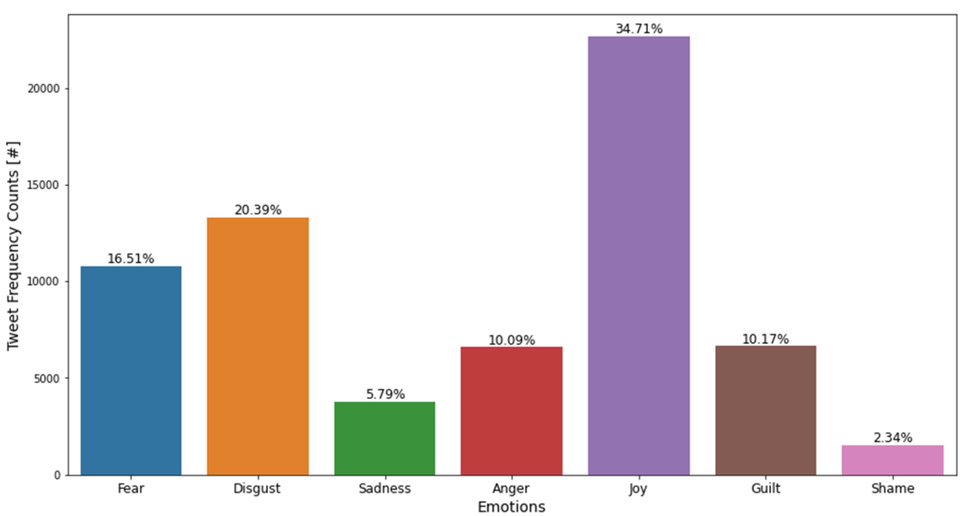
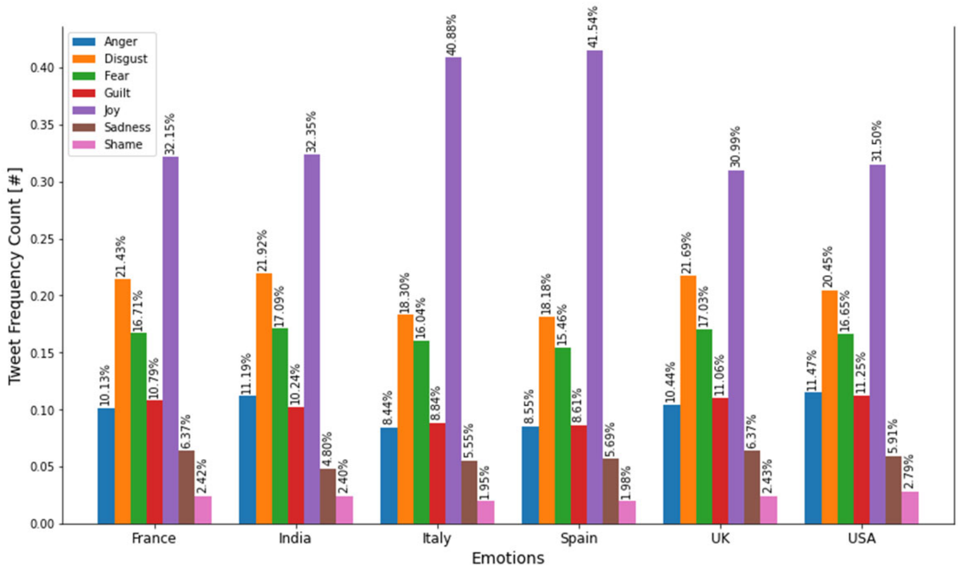


Figure 15 Country wise emotion (see online version for colours)



5 Concluding remarks

The COVID-19 pandemic has changed human thought processes. The ever-increasing number of positive COVID-19 instances has altered human behaviour and sentiment in all countries. This research investigates the possibilities of applying DL techniques for automatically extracting embedding features from Twitter data and analysing human sentiment and emotion in an appropriate class label. Various DL approaches, including a hybrid of CNN and LSTM, are being examined for experimentation. It has been discovered that a hybrid of CNN and LSTM outperforms single DL-based classifiers. Few authors have pointed out and used the combination of CNN and LSTM for SA. Publicly available datasets are used to develop two sentiment and EA models. In addition, the sentiment of six countries worldwide most affected by COVID-19 was considered for sentiment and EA. This research was conducted using Twitter data obtained from the Twitter website. The citizens of France, the USA, and the UK were found to be more negative than positive. It can also be concluded that sentiment and EA through Twitter data and DL techniques are more powerful than other models. The model developed here has self-explanatory and automatic feature extraction capabilities, making it more powerful, especially for Twitter data analysis.

Acknowledgements

The authors are grateful to the Editor-in-Chief and an anonymous reviewer for their constructive feedback.

References

Abdi, A., Shamsuddin, S.M., Hasan, S. and Piran, J. (2019) 'Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion', *Inf. Process Manag.*, Vol. 56, pp.1245–1259, https://doi.org/10.1016/j.ipm.2019.02.018.

Akhtar, M.S., Garg, T. and Ekbal, A. (2020) 'Multi-task learning for aspect term extraction and aspect sentiment classification', *Neurocomputing*, Vol. 398, pp.247–256, https://doi.org/10.1016/j.neucom.2020.02.093.

Alharbi, A.S.M. and de Doncker, E. (2019) 'Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioural information', *Cognitive Systems Research*, May, Vol. 54, pp.50–61, https://doi.org/10.1016/j.cogsys.2018.10.001.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M. et al. (2018) 'Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews', *J. Comput. Sci.*, Vol. 27, pp.386–393, https://doi.org/10.1016/j.jocs.2017.11.006.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y. and Gupta, B. (2018) 'Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews', *Journal of Computational Science*, Vol. 27, pp.386–393, https://doi.org/10.1016/j.jocs.2017.11.006.

Ankit, N.S. (2018) 'An ensemble classification system for Twitter sentiment analysis', *International Conference on Computational Intelligence and Data Science (ICCIDS 2018), Procedia Computer Science*, Vol. 132, pp.937–946.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J.F. and Iglesias, C.A. (2017) 'Enhancing deep learning sentiment analysis with ensemble techniques in social applications', *Expert Systems with Applications*, Vol. 77, pp.236–246.

Bansal, B. and Srivastava, S. (2018) 'On predicting elections with hybrid topic based sentiment analysis of tweets', *Procedia Comput. Sci.*, Vol. 135, pp.346–353, https://doi.org/10.1016/j.procs.2018.08.183.

Barkur, G. and Vibha, K.G.B. (2020) 'Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India', *Asian J. Psychiatr.*, Vol. 51, p.102089, https://doi.org/10.1016/j.ajp.2020.102089.

Basiri, M.E., Abdar, M., Cifci, M.A. et al. (2020) 'A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques', *Knowledge-Based Syst.,* Vol. 198, p.105949, https://doi.org/10.1016/j.knosys.2020.105949.

Bhowmik, N.R., Arifuzzaman, M. and Mondal, M.R. (2022) 'Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms', *Array*, Vol. 13, p.100123, https://doi.org/10.1016/j.array.2021.100123.

Birjali, M., Beni-Hssane, A. and Erritali, M. (2017) 'Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks', *Procedia Computer Science*, Vol. 113, pp.65–72, https://doi.org/10.1016/j.procs.2017.08.290.

Bradley, A.P. (1997) 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition*, Vol. 30, No. 7, pp.1145–1159, https://doi.org/10.1016/s0031-3203(96)00142-2.

Chakraborty, K., Bhattacharyya, S., Bag, R. and Hassanien, A.A. (2019) 'Sentiment analysis on a set of movie reviews using deep learning techniques', *Social Network Analytics*, pp.127–147.

Dan-Glauser, E.S. and Scherer, K.R. (2013) 'The difficulties in emotion regulation scale (DERS): factor structure and consistency of a French translation', *Swiss Journal of Psychology*, Vol. 72, No. 1, pp.5–11.

Day, M.Y. and Lee, C.C. (2016) 'Deep learning for financial sentiment analysis on finance news providers', in *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM 2016, IEEE, pp.1127–1134.

Devi, P.S. and Karthika, S. (2019) 'Cyclone Gaja-rank based credibility analysis system in social media during the crisis', *Procedia Comput. Sci.,* Vol. 165, pp.684–690, https://doi.org/10.1016/j.procs.2020.01.064.

Felbo, B., Mislove, A., Søgaard, A. et al. (2017) 'Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm', in *EMNLP 2017 – Conference on Empirical Methods in Natural Language Processing, Proceedings*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp.1615–1625.

Ghulam, H., Zeng, F., Li, W. and Xiao, Y. (2019) 'Deep learning-based sentiment analysis for Roman Urdu text', *Procedia Comput. Sci.*, Vol. 147, pp.131–135, https://doi.org/10.1016/j.procs.2019.01.202.

Groß-Klußmann, A., König, S. and Ebner, M. (2019) 'Buzzwords build momentum: global financial Twitter sentiment and the aggregate stock market', *Expert Syst. Appl.,* Vol. 136, pp.171–186, https://doi.org/10.1016/j.eswa.2019.06.027.

Haihong, E., Yingxi, H., Haipeng, P. et al. (2019) 'Theme and sentiment analysis model of public opinion dissemination based on generative adversarial network', *Chaos, Solitons and Fractals*, Vol. 121, pp.160–167, https://doi.org/10.1016/j.chaos.2018.11.036.

Hassonah, M.A., Al-Sayyed, R., Rodan, A. et al. (2020) 'An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter', *Knowledge-Based Syst*., Vol. 192, p.105353, https://doi.org/10.1016/j.knosys.2019.105353.

Heikal, M., Torki, M. and El-Makky, N. (2018) 'Sentiment analysis of Arabic Tweets using deep learning', *Procedia Comput. Sci*., Vol. 142, pp.114–122, https://doi.org/10.1016/j.procs.2018.10.466.

Hota, H.S., Sharma, D.K. and Verma, N. (2021) 'Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad', *Data Science for COVID-19*, pp.275–295, Academic Press, London wall, UK.

Jiménez-Zafra, S.M., Martín-Valdivia, M.T., Molina-González, M.D. and Ureña-López, L.A. (2019) 'How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain', *Artif. Intell. Med*., Vol. 93, pp.50–57, https://doi.org/10.1016/j.artmed.2018.03.007.

Katayama, D., Kino, Y. and Tsuda, K. (2019) 'A method of sentiment polarity identification in financial news using deep learning', *Procedia Comput. Sci.*, Vol. 159, pp.1287–1294, https://doi.org/10.1016/j.procs.2019.09.298.

Kiran, R., Kumar, P. and Bhasker, B. (2020) 'Oslcfit (organic simultaneous LSTM and CNN fit): a novel deep learning based solution for sentiment polarity classification of reviews', *Expert Systems with Applications*, Vol. 157, No. 2020, p.113488.

Kumar, A., Srinivasan, K., Cheng, W.H. and Zomaya, A.Y. (2020) 'Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data', *Inf. Process Manag.*, Vol. 57, p.102141, https://doi.org/10.1016/j.ipm.2019.102141

Lim, S. and Tucker, C.S. (2019) 'Mining Twitter data for causal links between tweets and real-world outcomes', *Expert Syst. with Appl. X.*, Vol. 3, p.100007, https://doi.org/10.1016/j.eswax.2019.100007.

Liu, N. and Shen, B. (2020) 'ReMemNN: a novel memory neural network for powerful interaction in aspect-based sentiment analysis', *Neurocomputing*, Vol. 395, pp.66–77, https://doi.org/10.1016/j.neucom.2020.02.018.

Luan, Y. and Lin, S. (2019) 'Research on text classification based on CNN and LSTM', *2019 IEEE International Conference on Artificial Intelligence and Computer Applications* (ICAICA), 2019, pp.352–355, DOI: 10.1109/ICAICA.2019.8873454.

Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C. (2011) 'Learning word vectors for sentiment analysis', *The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*.

Mahmood, Z., Safder, I. and Nawab, R.M.A. (2020) 'Deep sentiments in Roman Urdu text using recurrent convolutional neural network model', *Inf. Process Manag.,* Vol. 57, p.102233, https://doi.org/10.1016/j.ipm.2020.102233.

Medhat, W., Hassan, A. and Korashy, H. (2014) 'Sentiment analysis algorithms and applications: a survey', *Ain Shams. Eng. J.*, Vol. 5, pp.1093–1113, https://doi.org/10.1016/j.asej.2014.04.011.

Minaee, S., Azimi, E. and Abdolrashidi, A.A. (2019) *Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models*, arXiv.

Nisar, T.M. and Yeung, M. (2018) 'Twitter as a tool for forecasting stock market movements: a short-window event study', *J. Financ. Data Sci.,* Vol. 4, pp.101–119, https://doi.org/10.1016/j.jfds.2017.11.002.

Pandey, M., Williams, R., Jindal, N. and Batra, A. (2017) 'Sentiment analysis using Lexicon based approach', *IITM Journal of Management and IT*, Vol. 10, No. 1.

Poria, S., Chaturvedi, I., Cambria, E. and Hussain, A. (2017a) 'Convolutional MKL based multimodal emotion recognition and sentiment analysis', *Proc. – IEEE Int. Conf. Data Mining*, ICDM, pp.439–448, https://doi.org/10.1109/ICDM.2016.178.

Poria, S., Peng, H., Hussain, A. et al. (2017b) 'Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis', *Neurocomputing*, Vol. 261, pp.217–230, https://doi.org/10.1016/j.neucom.2016.09.117.

Ruz, G.A., Henríquez, P.A. and Mascareño, A. (2020) 'Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers', *Futur. Gener. Comput. Syst*., Vol. 106, pp.92–104, https://doi.org/10.1016/j.future.2020.01.005.

Sailunaz, K. and Alhajj, R. (2019) 'Emotion and sentiment analysis from Twitter text', *J. Comput. Sci.,* Vol. 36, p.101003, https://doi.org/10.1016/j.jocs.2019.05.009.

Singh, J., Singh, G., Singh, R. and Singh, P. (2018a) 'Morphological evaluation and sentiment analysis of Punjabi text using deep learning classification', *J. King Saud Univ. – Comput. Inf. Sci.*, https://doi.org/10.1016/j.jksuci.2018.04.003.

Singh, T., and Kumari, M. (2016) 'Role of text pre-processing in twitter sentiment analysis', *Procedia Computer Science*, Vol. 89, pp.549–554, https://doi.org/10.1016/j.procs.2016.06.095.

Song, M., Park, H., Shin, K.S.B.N.B. (2019) 'Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean', *Inf. Process Manag.,* Vol. 56, pp.637–653, https://doi.org/10.1016/j.ipm.2018.12.005.

Soumya, S. and Pramod, K.V. (2020) 'Sentiment analysis of Malayalam tweets using machine learning techniques', *ICT Express*, Vol. 6, pp.300–305, https://doi.org/10.1016/j.icte.2020.04.003.

Tang, L., Bie, B., Park, S.E. and Zhi, D. (2018) 'Social media and outbreaks of emerging infectious diseases: a systematic review of literature', *Am. J. Infect Control*, Vol. 46, pp.962–972, https://doi.org/10.1016/j.ajic.2018.02.010.

Youssef, M., and El-Beltagy, S. R. (2018) 'MoArLex: an Arabic sentiment lexicon built through automatic Lexicon expansion', *Procedia Computer Science*, Vol. 142, pp.94–103, https://doi.org/10.1016/j.procs.2018.10.464.

Zhang, L., Wang, S. and Liu, B. (2018) 'Deep learning for sentiment analysis: a survey', *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, Vol. 8, https://doi.org/10.1002/widm.1253.