PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Acta Informatica Pragensia
2023, Volume 12, Issue 2, pp. 225–242
https://doi.org/10.18267/j.aip.206

Article Open Access

Emotion-Based Sentiment Analysis Using Conv-BiLSTM with Frog Leap Algorithms

Sandeep Yelisetti , Nellore Geethanjali

Department of Computer Science Technology, Sri Krishnadevaraya University, Ananthapuram, India

Corresponding author: Sandeep Yelisetti (sandeep.yelisetti@gmail.com)

Abstract

Social media, blogs, review sites and forums can produce large volumes of data in the form of users' emotions, views, arguments and opinions about various political events, brands, products and social problems. The user's sentiment expressed on the web influences readers, politicians and product vendors. These unstructured social media data are analysed to form structured data, and for this reason sentiment analysis has recently received the most important research attention. Sentiment analysis is a process of classifying the user's feelings in different manners such as positive, negative or both. The major issue of sentiment analysis is insufficient data processing and outcome prediction. For this, deep learning-based approaches are effective due to their autonomous learning ability. Emotion identification from the text in natural language processing (NLP) provides more benefits in the field of e-commerce and business environments. In this paper, emotion detection-based text classification is used for sentiment analysis. The data collected are pre-processed using tokenization, stop word discarding, stemming and lemmatization. After performing data pre-processing, the features are identified using term frequency and inverse document frequency (TF-IDF). Then the filtered features are turned into word embeddings by documents as a vector (Doc2Vec). Then, for text classification, a deep learning (DL) based model called convolutional bidirectional long short-term memory (CBLSTM) is used to differentiate the sentiments of human expression into positive or good and negative or bad emotions. The neural network hyper-parameters are optimized with a meta-heuristic algorithm called the frog leap approach (FLA). The proposed CBLSTM with FLA uses four review and Twitter datasets. The experimental results of this study are compared with the conventional approaches LSTM-RNN and LSTM-CNN to prove the efficiency of the proposed model. Compared to LSTM-RNN and LSTM-CNN, the proposed model secures an improved average accuracy of 98.1% for review datasets and 97.5% for Twitter datasets.

Keywords

Emotion; Sentiment analysis; Deep learning; Convolution neural network; LSTM; Frog leap algorithm; Meta-heuristic; TF-IDF.

Citation: Yelisetti, S., & Geethanjali, N. (2023). Emotion-Based Sentiment Analysis Using Conv-BiLSTM with Frog Leap Algorithms. *Acta Informatica Pragensia*, 12(2), 225–242. https://doi.org/10.18267/j.aip.206

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

Copyright: © 2023 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

1 Introduction

Sentiment analysis performs tasks of identifying word polarity in documents into various classes including positive, neutral and negative. It has been used in various applications such as recommendation systems, opinion mining and health informatics. Depending on people's opinions, business people may form their strategic approaches to increase their productivity with increased profits.

Sentiment analysis helps to verify the person's mental ability as good/positive, bad/negative or neutral. This sentiment analysis can be classified into three levels. They are document levels, sentence levels and aspect levels. Document level analysis checks whether the text is a good/positive or bad/negative option for the user. If the text has 40 characters or more, this analysis works well. Sentence level analysis checks whether the expressed opinions are positive, negative or neutral. The aspect level method is better because it analyses large volumes of data and also helps save money and time. To identify one's mental ability, sentiment analysis helps identify the user's positive and negative statements. It monitors the data shared by the user through social media and processes large amounts of text data. It is impossible to read all the feedback manually. With the help of sentiment analysis, user opinions are gathered without reading the entire feedback. Likewise, on social platforms, human psychology and behaviours are analysed based on analysis of their tweets. For analysis of sentiments, researchers make use of various textual documents from Facebook and Twitter posts (Birjali et al., 2021).

Twitter helps users post comments in the form of tweets. This social media platform shares an individual's thoughts about diverse subjects, themes and fields. This is considered a collection of users' thoughts and sentiments across different subjects such as net articles and blogs. The data generated from Twitter are bigger compared to other media and blogging networks. Compared to existing blogging platforms, Twitter's response time is quicker. This emotional analysis has been used by vendors to gain insight into their business activities to recognize new trends on markets (Damarta et al., 2021).

Sentiments consist of various featured values such as bi-grams and tri-grams in the sense of polarities and their combinations. Therefore, feelings are processed through numerous machine-learning algorithms. To compute the label belongingness, neural networks are implemented. Context level data extraction among the edges and acyclic graph nodes is executed using a Bayesian network. To optimize the words, sentences and learning rate, the data accuracy was increased. These optimization methods are used to reduce the error to obtain an increased level of precision of social media data (Dang et al., 2020). In the recent past, machine learning (ML) approaches have been improved and provided enhanced solutions for sentiment analysis. ML is connected with statistical analysis, which creates predictions using digital computers. Sentiment analysis of Twitter and review data uses various approaches where deep learning (DL) models produce great results in emotion recognition (Dashtipour et al., 2020; Datta et al., 2021). In the field of speech recognition and computer vision, DL-based approaches provide better results (De Haro-Garcia et al., 2020; Dessì et al., 2020). Recurrent neural network (RNN) based approaches obtain better results in the field of natural language processing (NLP) deep networks (Eusuff et al., 2020). Meta-heuristics (MH) based neural network parameter optimization is not researched more but increases the classification accuracy of the model.

This research focuses on the following objectives:

- This paper uses convolutional and bi-directional LSTM for analysing sentiments using two datasets as long and short reviews of Twitter datasets which have not been used in the same research area to the best of our knowledge.
- The DL model performs better in gathering contextual data from the datasets and sentiment analysis to differentiate user emotions into good or bad compared with techniques used by Pimpalkar et al. (2022), who performed sentiment classification into positive or negative with an accuracy rate of 92.3% by using convolutional multi-layer Bi-LSTM.

- The MH-based optimized model improves the performance of the classifier to classify tweets and reviews into positive or negative emotions.

Major contribution of this work:

- The input data are pre-processed to remove special symbols, stop words, numerals, and so on.
- From the pre-processed data, features are selected using the TF-IDF approach. The selected features are fed as input to Doc2Vec for word embeddings.
- Here, we implement convolutional Bi-LSTM with FLA to classify opinions into good or negative emotions. The good emotions include fun, love, neutrality, enthusiasm, happiness, surprise and relief, while the bad emotions include worry, emptiness, anger, sadness and boredom.
- The proposed emotion-based sentiment analysis using CBLSTM-FLA is implemented using review and Twitter datasets. The effectiveness of this model is compared with conventional models and the results show better performance in analysing sentiments based on emotions.

The paper is structured as follows: the related research papers are reviewed in Section 2. The DL model with MH-based sentiment analysis is proposed in Section 3. The proposed model performance is experimented with and evaluated in Section 4. The future effectiveness of the proposed model is reviewed in Section 5 as a conclusion.

2 Related Work

This section discusses related research work on sentiment analysis. Birjali et al. (2021) developed a DL-based sentiment analysis for Nepali COVID-19 tweets. They proposed three feature extraction approaches to representing the tweets, namely fastest-based, domain-specific and domain-agnostic. The domain-specific and agnostic methods are novel. They used three convolutional neural networks (CNN) to execute the extracted features. The ensemble CNN was evaluated using the Nepali Twitter sentiment dataset and produces better results for sentiment classification. Dashtipour et al. (2020) developed emotion recognition along with text analysis using LSTM and recurrent neural network (RNN) models. A multitude of tweets was classified into positive and negative emotions. This hybrid model secured 89.1% for positive emotion classification and 91.3% for negative emotion classification.

Imran et al. (2020) proposed an unsupervised aspect-based sentiment analysis (ABSA) method to handle various language tweets. The aspect word and polarity lexicons were selected from the datasets. The expectation maximization (EM) approach was used to detect sentiment probabilities. To find the polarity, the documents are broken down and this model is evaluated using the Persian and English language dataset. The performance is good. Imran et al. (2020) introduced an aspect-level neural network called CNN with Bi-LSTM for sentiment analysis. The phrases were removed using the embedding layer of the CNN and Bi-LSTM was used to enhance the content coding and retain the global semantics of the local features. This model experimented with the Twitter dataset.

Jadidoleslam et al. (2012) reviewed 32 DL-based sentiment analysis papers and analysed different DL models such as deep neural networks, RNN and CNN using eight datasets. The DL models mostly used approaches to sentiment analysis among the 32 approaches. They used both word embedding and TF-IDF to prepare the features for classifiers. The processing time of RNN was larger than that for CNN. Their analysed DL models included CNN, LSTM (tree, discourse, attention and Bi-LSTM), gated recurrent unit, RNN, latent rating NN, co-attention MemNet, simple RNN, and recurrent neural tensor network. Jain et al. (2020) proposed a DL model for demonetization tweets. They developed various stages of processing such as pre-processing, polarity feature extraction, aspect extraction and classification. The weights of the polarity score were optimized with two MH algorithms such as the firefly algorithm and multi-verse optimization algorithm. They developed a firefly multi-verse optimizer with RNN for sentiment classification.

Karakoyun et al. (2015) used the DNet model for sentiment analysis with gated CNN. The aspect data are extracted from the unstructured text using stacked gated CNN. This model was evaluated using Access Control List (ACL14) and Semeval 2014 Twitter datasets. Kastrati et al. (2021) developed an automatic depression system using Twitter data from user profiles. They used an n-gram model, a bag of visual words, Linguistic Inquiry and Word Count (LIWC) dictionaries, and automatic image tagging for detection. Correlation-based models were used for feature selection and 91% accuracy was obtained. Ligthart et al. (2021) proposed a sentiment analysis model from a Twitter account using the k-nearest neighbour model. Using KNN, the opinions about Twitter service was viewed and PT PLN (Persero) was controlled by a text mining model. Initial data were pre-processed and KNN classified the data as positive, neutral or negative classes. The resultant accuracy was based on the data quality.

Meng et al. (2019) proposed a model using a firefly algorithm to find geographical data. A geotagging procedure followed and the tweet count was increased. Using this method, locality-aware keywords were extracted and grouped to detect the data. Detecting small news in the local space using this method overcomes data sparsity. Ren et al. (2020) proposed a novel concept-level sentiment analysis in the Persian language. It integrates DL and linguistic rules for polarity detection optimization. While triggering the pattern, the framework flows from words into concepts using symbolic dependency relations. While not triggering the pattern, deep neural network (DNN) models perform the classification. This approach improves the results by 10 to 15% more than traditional ML and DL approaches. Safa et al. (2021) developed DL and aspect-based opinion mining (OM) for recommendation systems. It consists of two-part aspect-based OM and rating prediction using multichannel deep CNN. For the overall rating, the tensor factorization is integrated with aspect-specific ratings. Shams et al. (2020) proposed a DL-based word embedding model to measure the similarities, polarity and reviews of learners attending online courses. They showed the usage of integrating word embedding with the DL model unlike standard ML approaches.

Emotion-based sentiment analysis is an essential research factor for various business and e-commerce applications. There are various research practices for emotion analysis using deep learning algorithms. However, it fails to achieve dynamic processing and analysis. Due to this problem, the learning accuracy of deep learning algorithms degrades. To overcome this research gap, this paper used convolutional Bi-LSTM with a frog leap optimizer to improve the learning capacity of the machine.

3 Proposed Solution

Figure 1 illustrates an overview of the proposed sentiment analysis system. The input raw data are pre-processed to remove punctuation, stop words, lowercase conversion, normalization, stemming and tokenization. The pre-processed data are used for feature selection using TF-IDF and used as input for word embeddings using Doc2Vec. Then, the proposed CBLSTM-FLA is used to classify positive or negative emotions. The major advantage of using Bi-LSTM with a convolutional layer is that it requires fewer labelled training data in the neural network. It considers contextual data and information on the order of words from the Twitter data for sentiment analysis.

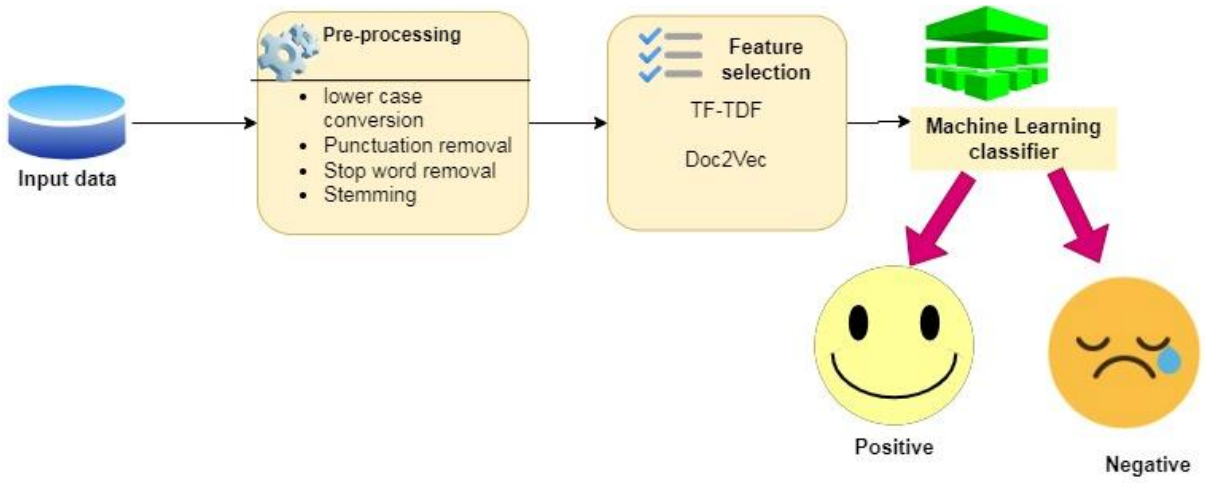

Figure 1. *Overview of proposed sentiment analysis system.*

3.1 Dataset collection

The four review datasets and two Twitter datasets (Al-Deen et al., 2021; Basiri et al., 2021) are described as follows.

Review datasets:

- Application (App): this dataset is collected from Android apps which consists of 752,937 Amazon product reviews and their related metadata.
- Kindle (Al-Deen et al., 2021): this Kindle store data set consists of 982,619 Amazon product reviews with its metadata.
- Electronics: this data set consists of 1,689,188 Amazon product reviews.
- CDs: this dataset consists of 1,097,592 product reviews for CDs and vinyls from Amazon.

Twitter datasets:

- Airline Twitter: this is the dataset for airline Twitter sentiment dataset which has 14,641 tweets related to significant issues in US airlines from Feb 2015.
- Sentiment140: It consists of 1,600,000 tweets categorized as positive and negative and it was created by Stanford University computer science students. The dataset's numerical data are listed in Table 1 (Al-Deen et al., 2021).

Table 1. *Dataset description.*

Type of dataset	Dataset used	Total number of instances	Positive data	Negative data
Reviews	App	752,936	123,099	123,097
	Kindle	982,619	57,147	57,149
	Electronics	1,689,187	190,863	190,863
	CD	1,097,591	92,765	92,764
Tweets	Airline Tweets	14,642	2362	2364
	Sentiment140	1,600,000	800,000	800,000

3.2 Data pre-processing

The reviews and messages from the users are informal and have various styles based on their nationality, origin, gender and age. This results in noisy data with unwanted emoticons and symbols. These unnecessary data are eradicated with pre-processing methods such as eliminating stop words, punctuation elimination, converting to lowercase, tokenization, normalization and stemming. The obtained tweet data consist of hypertext markup language (HTML) tags and these tags are converted using an HTML parser. Multiple dots and spaces in the messages are rewritten with a single space and unwanted spaces at the end are removed. Repeatedly occurring words are called stop words; these include articles, prepositions, adverbs and conjunctions in English, and are removed. Punctuation is used for proper comprehension. It has different uses and shapes which helps understand sentence meaning. Punctuation mark removal reduces system processing time. Lowercase conversion is used frequently and consumers use mostly lowercase letters. In a single line, compared to uppercase letters, lowercase letters are familiar and most viewed.

Stemming can help reduce meaningless words in sentences. In a sentence, the initial part of each word is found with different suffixes. Image categorization is prominent for use on different images in tweets and reviews, which consist of smileys, hand gestures, and so on. These images are not to be removed since they are important in the analysis of sentiments. Therefore, two substitutions are used to differentiate the images into positive and negative emotions.

3.3 Feature selection using TF-IDF

TF-IDF is used in NLP to identify the most important words from text data. The TF converts words into string format for numerical processing by the DL algorithms. It finds the frequency of word occurrences in both classes. In our considered datasets the features are the words. The frequency of each word is computed using TF as in Equation (1).

$$TF(i,j) = \frac{termifrequencyindocumentj}{totalnumberofwordsinthedocumentj} \tag{1}$$

The frequency of the term in the document is represented as a table in the memory. IDF is used to find the most or least relevant word occurrences in the file. It helps identify words from documents. TF is used to get the highest-degree words and IDF is used to find the least-occurring words using logarithmic values as in Equation (2).

$$IDF(i) = log2(\frac{D}{D_i}) \tag{2}$$

Where D is total documents and D_i is the total documents with the term i. Finally, the TF and IDF product is computed as matrices with normalized weights. These are the TF-IDF results. The bag of words is denoted as text.

3.3.1 Word embedding using Doc2Vec

The Doc2Vec approach performs better for sentences with similar tasks. It is an improved version of Word2Vec. If the input has misspelled words, then it makes no choice of an ideal word. In this paper, it is used for document vectorization to convert words into vectors and uses these vectors to create a vector format of the entire document. In our study, the word vectors are created from Doc2Vec. The workflow of TF-IDF feature selection is shown in Figure 2.

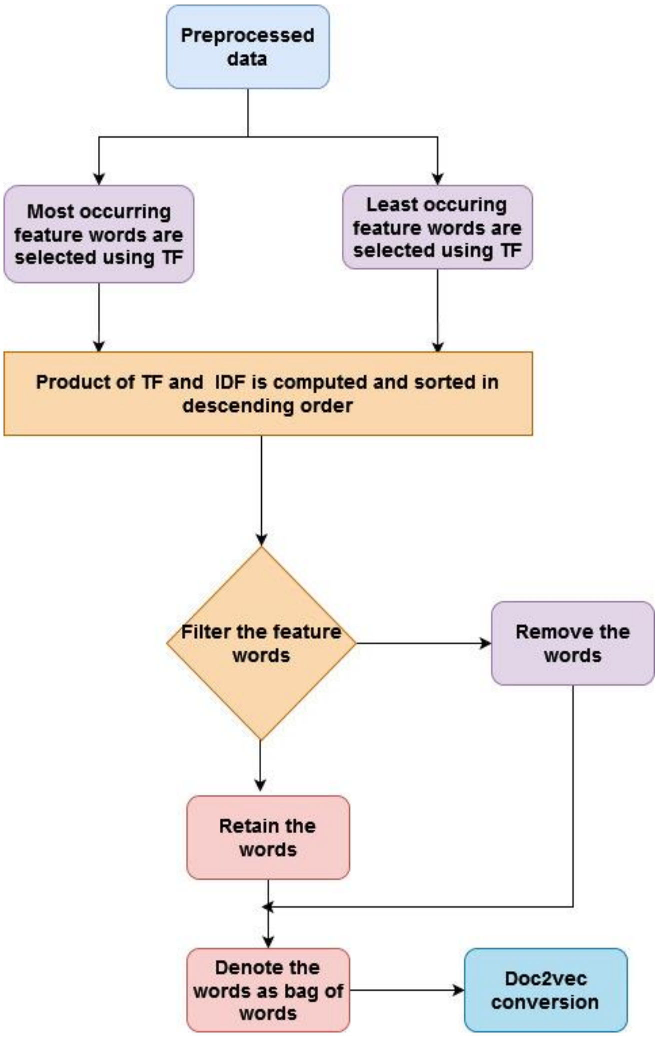

Figure 2. *Workflow of TF-IDF feature selection.*

3.4 Classification using proposed CBLSTM-FLA

The selected word vector features are provided as input to the classification phase using Convolutional Bi-LSTM with a frog leap approach (CBLSTM-FLA). The features from the TF-IDF are classified using a hybrid classifier called convolutional Bi-LSTM and the hyperparameters are optimized with a meta-heuristic algorithm called FLA to improve the classifier performance.

3.4.1 Convolutional neural network model

The CNN model consists of artificial neurons with bias, weights and activation functions for task performance. These neurons are mediators between the input and output layers. The CNN model has five layers, namely input, output, pooling, convolution and fully connected layer. The selected feature vectors in words are given as input to the input layer and then forwarded to the convolution layer for feature learning through convolution operations. The rectified linear unit (ReLU) activation function is used to make null values instead of negative values. The convolution output is fed as input to the pooling layer which reduces the spatial volume of input data. Max pooling is used for this purpose.

3.4.2 Bi-LSTM

Bi-LSTM is the convolutional recurrent neural network that is used to overcome the limitation of conventional CNN (Singh et al., 2016). The neural states of this network are categorized into two forward and backward states. The results operate forward and backward as two distinct factors in RNN. Both networks are connected via the same output layer to produce the results. In this case, future evaluation and sequential past situations are possible before producing the results. Compared to traditional LSTM, two different LSTMs are constructed in Bi-LSTM architecture.

3.4.3 Proposed CBLSTM

The proposed model comprises twelve layers, namely three convolution layers, three max pooling, two Bi-LSTM layers, one flattening, two dense layers and an output layer, as stated in Table 2.

Table 2. *Description of CBLSTM.*

Layers	Output neuron	Kernel size	Activation function
Convolution layer 1	1212*4	35	ReLU
Max pooling 1	604*4	-	-
Convolution layer 2	572*6	25	ReLU
Max Pooling 2	284*6	-	-
Convolution layer 3	272*3	10	ReLU
Max pooling 3	136*3	-	-
Bi-LSTM 1	136*8	Dropout = 0.15	ReLU
Bi-LSTM 2	136*16	Dropout = 0.15	ReLU
Flattening layer	2214	-	-
Dense layer 1	280	-	ReLU
Dense layer 2	80	-	ReLU
Output layer	4	-	Softmax

By using the convolution operation, CNN automatically learns the features from the best fit of input features. The neuron weight remains the same in the feature map, which results in parallel network learning that can reduce the learning time. The convolution operations with kernel sizes 35, 25 and 10 are in the first, third and fifth layers based on Equation (3).

$$Y_n = \sum_{i=0}^{N-1} X_i f_{n-i} \tag{3}$$

Where X is the input feature, Y is the output vector, f is filters and N is several elements. Followed by convolution layers in the process, the max pooling layer helps reduce the feature map size using the ReLU activation function. The temporal data are learned through the Bi-LSTM layer. For sequential analysis, convolution and max pooling features are segmented into sequential parts that are provided as input to Bi-LSTM for processing. The output of Bi-LSTM is fed to a fully connected layer for producing the output using the activation function mentioned in Equation (4).

$$f_t = \alpha\left(w\left[X_{t,}h_{t-1}, c_{t-1}\right]\right) + b_f \tag{4}$$

Where X is the input, b is bias, w is weight, α is the activation function, h is previous state outcome and c is previous LSTM memory. The CBLSTM architecture is shown in Figure 3.

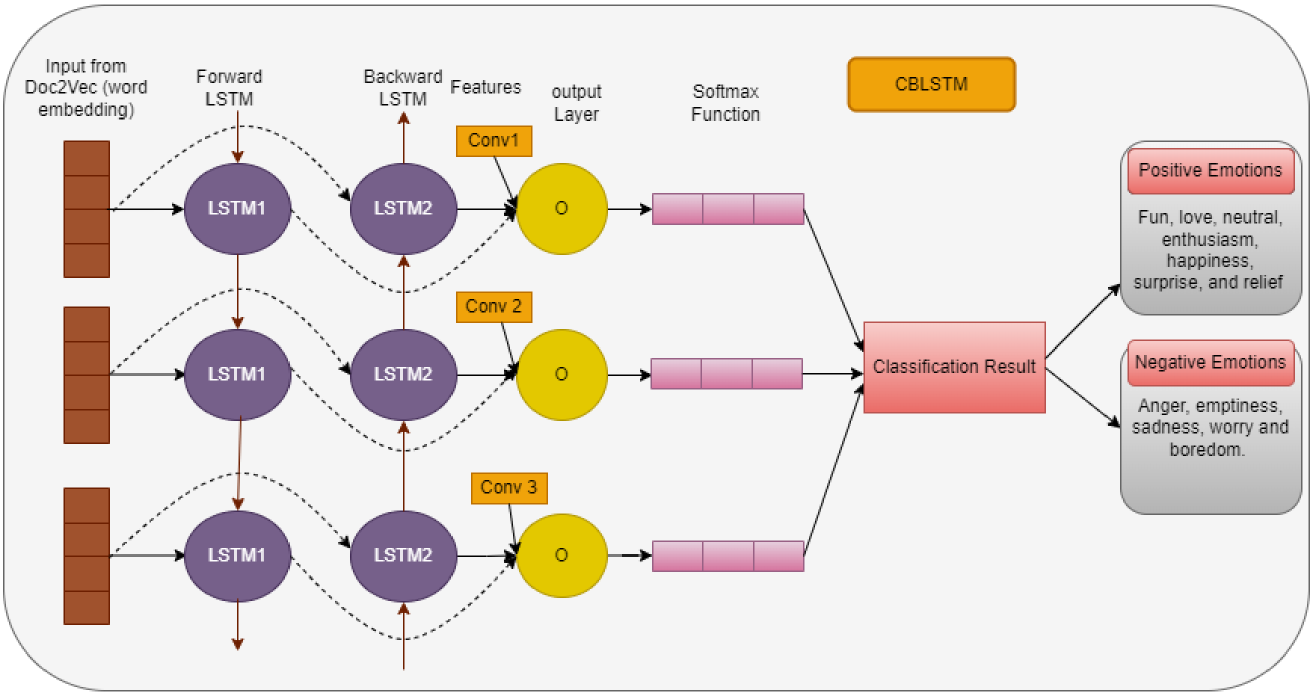

Figure 3. *CBLSTM network architecture.*

The final layer utilizes a softmax activation function that classifies the output into two classes such as the user's positive or negative state. The CBLSTM is trained iteratively. Finally, all the records are used to compute the statistical analysis of the proposed model. The training parameters of this network are listed in Table 3. The best values of these hyper-parameters increase the detection process accuracy. For that, the MH algorithm called FLA is used to optimize the hyper-parameters and selects the best possible values for these parameters.

Table 3. *Hyperparameters of CBLSTM.*

Parameters	Values
Learning rate	0.0001
Regularization	0.01
Batch size	54
Loss parameter	Cross entropy
Number of iterations	100
Dropout rate	0.15
Number of epochs	30
Activation function	ReLU, Softmax

3.4.4 FLA-based hyperparameter optimization

The frog leap algorithm (Sitaula et al., 2021) was developed by Eusuff and Lansey. It is a population-based meta-heuristic (MH) algorithm inspired by frogs' social behaviours which are executed iteratively. Frogs naturally inhabit wetlands and they live on land as well as in water. Frogs find food using small moves. To perform this, frogs form groups in the wetland which are called memeplexes. The information about each frog is known as a meme. In a memeplex, there will be a search operation local, which aims to update the worst frog position towards the best place using the global or local best frog position (Su et al., 2018).

The FLA process starts with parameter initialization. The first population of frogs is randomly generate in the decision space. Each frog fitness value is computed. Population D is split based on fitness value where the first member is the best. The filtered population is divided into memeplexes using Equation (5) and each memeplex has a memetic process. After this, the memeplexes are grouped and shuffled. The new iteration process starts with the population resorted until it meets the termination condition. The workflow of the FLA process is illustrated in Figure 4 (Wei et al., 2020).

$$Y_l = [(D_i)_l | D_i = D\big(l + m * (i-1)\big) where\ i = 1,2, \dots n\ and\ l = 1,2,3, \dots m \tag{5}$$

Where m denotes the number of memeplexes and n is the number of frogs in each memeplex. Using Equation (5), the members are equally distributed. For instance, if m=2, the first memeplex receives the first member and the second memeplex receives the second member.

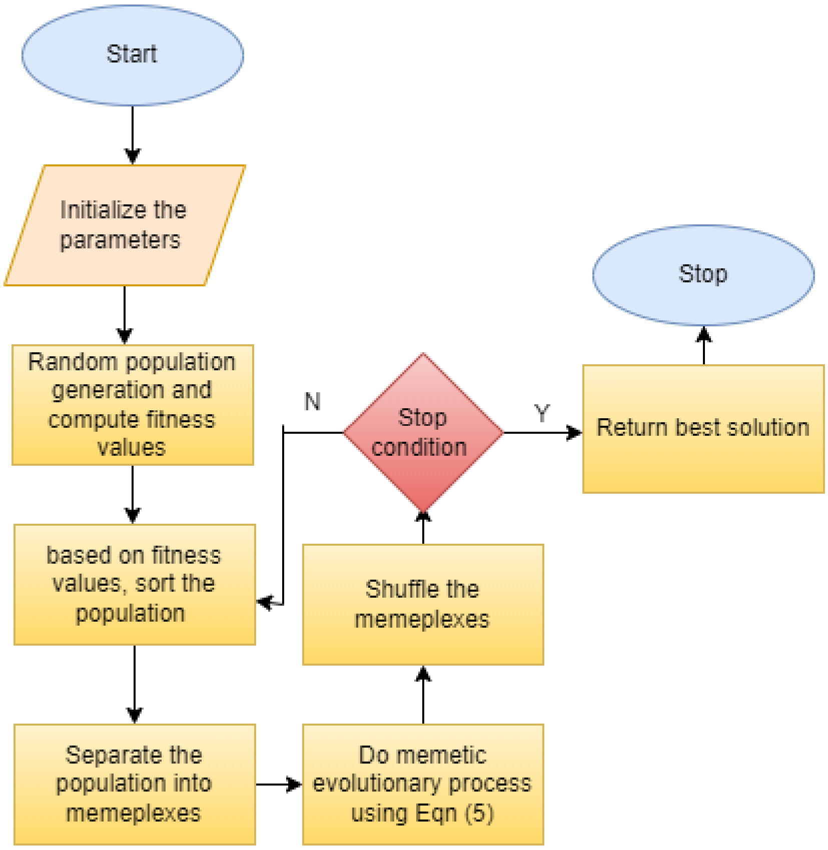

Figure 4. *Workflow of FLA.*

4 Results and Discussion

This section discusses the efficiency of the proposed CBLSTM-FLA model for sentiment analysis using emotions in user reviews.

4.1 Simulation parameters

The proposed model is implemented using TensorFlow with Keras in Python 3.7.1. A Ubuntu 16.04 machine with a four-core i7-7700K central processing unit (CPU) is used. The tokenizer approach employs 100,000 words to build the input comment matrix C. We fixed the padding values to 45 and 100, accordingly, assuming the first 45 and 100 words of remarks in the tweet and review datasets, respectively. The pre-trained, publicly accessible GloVe and Word2Vec models were used in the investigation as the embedding layer weights. It used the "Gigaword 5 + Wikipedia 2014" version of GloVe, which has a vocabulary size of 400,000 words and six billion tokens. The embedding length of 300 was employed for the embedding layer. Table 4 shows the simulation parameters for this model.

Table 4. *Simulation parameters.*

Parameters used	Values
Dimensions (d)	100
Bi-LSTM hidden units	120
Comments for reviews in words	150
Tweet comments in words	80
Optimizer	FLA
Kernel-size	4
Convolution layer	64

The proposed model effectiveness is evaluated using evaluation metrics, namely accuracy, recall, precision and F1-score.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} X100 \tag{6}$$

$$Precision = \frac{TP}{TP+FP} X100 \tag{7}$$

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

$$F1-score = \frac{2*Precision*Recall}{Precision+Recall} \tag{9}$$

The proposed CBLSTM-FLA-based sentiment analysis is compared with conventional approaches such as LSTM-RNN (Shilpa et al., 2021), LSTM-CNN (Yadav et al., 2020) and ensemble CNN (Sitaula et al., 2021).

4.2 Evaluation of review datasets

In the long reviews of the four datasets such as Kindle, App, CDs and Electronics, positive and negative emotions are classified and the results are shown in Tables 5 to 8.

Table 5. *Performance of CBLSTM-FLA on Kindle dataset.*

Method	Accuracy	Precision	Recall	F1-score
LSTM with RNN	0.9102	0.9032	0.8902	0.8843
LSTM with CNN	0.9211	0.9018	0.8553	0.9018
Ensemble CNN	0.9304	0.9103	0.9113	0.9281
Proposed CBLSTM-FLA	0.9881	0.9728	0.9629	0.9792

Table 6. *Performance of CBLSTM-FLA on APP dataset.*

Method	Accuracy	Precision	Recall	F1-score
LSTM with RNN	0.9098	0.9032	0.8902	0.8843
LSTM with CNN	0.9182	0.9018	0.8553	0.9018
Ensemble CNN	0.9261	0.9178	0.9195	0.9201
Proposed CBLSTM-FLA	0.9782	0.9671	0.9511	0.9672

Table 7. *Performance of CBLSTM-FLA on CD dataset.*

Method	Accuracy	Precision	Recall	F1-score
LSTM with RNN	0.9112	0.9101	0.9114	0.8993
LSTM with CNN	0.9273	0.9064	0.9071	0.9125
Ensemble CNN	0.9289	0.9171	0.9192	0.9103
Proposed CBLSTM-FLA	0.9789	0.9616	0.9664	0.9772

Table 8. *Performance of CBLSTM-FLA on electronics dataset.*

Method	Accuracy	Precision	Recall	F1-score
LSTM with RNN	0.9378	0.9101	0.9114	0.8993
LSTM with CNN	0.9337	0.9172	0.9082	0.9125
Ensemble CNN	0.9298	0.9189	0.9213	0.9282
Proposed CBLSTM-FLA	0.9816	0.9717	0.9775	0.9872

An evaluation of Tables 5 to 8 demonstrates that the proposed model secured improved performance compared to other sentiment analysis approaches such as LSTM-RNN, LSTM-CNN and ensemble CNN. The average accuracy, precision, recall and F1-score values of all the four long review datasets is illustrated in Figure 5. As an average, the proposed model achieved the accuracy, precision, recall and F1-score values of 0.98, 0.968, 0.964 and 0.977 respectively. Meanwhile, the other approaches achieved worse performance than the proposed model in detecting emotions for sentiment analysis. Their results are as follows: LSTM-RNN (0.917, 0.906, 0.9008, 0.891), LSTM-CNN (0.92, 0.906, 0.88, 0.907) and ensemble CNN (0.928, 0.916, 0.917, 0.921). The ROC comparison of these models is shown in Figure 6; it proves the efficiency of the proposed model with an ROC of 0.95.

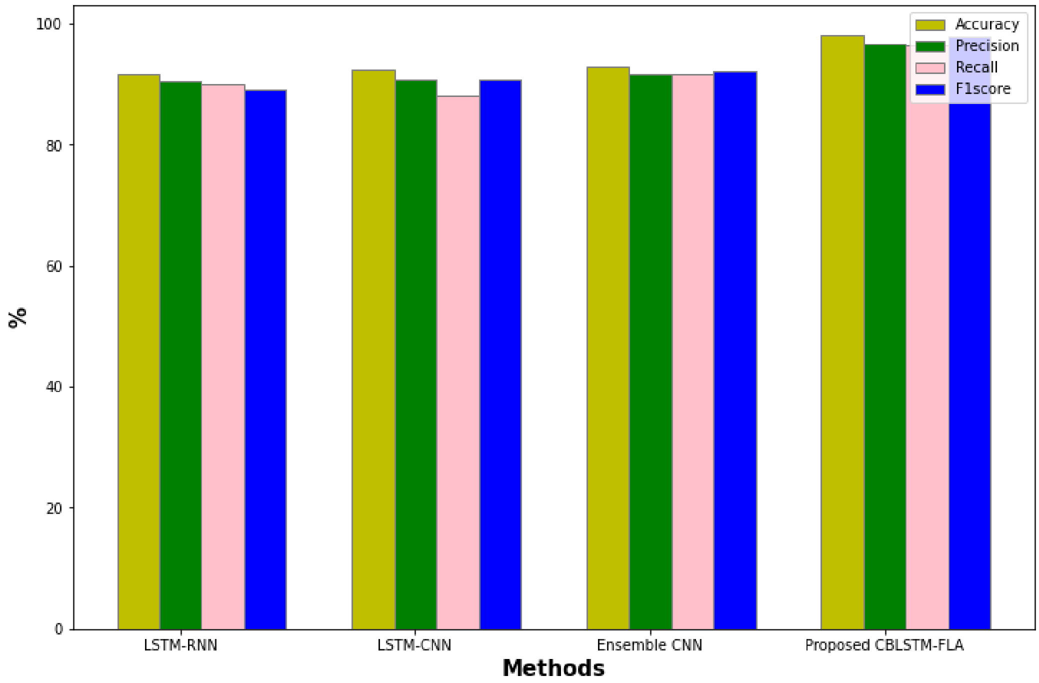

Figure 5. *Average performance of proposed model on review datasets.*

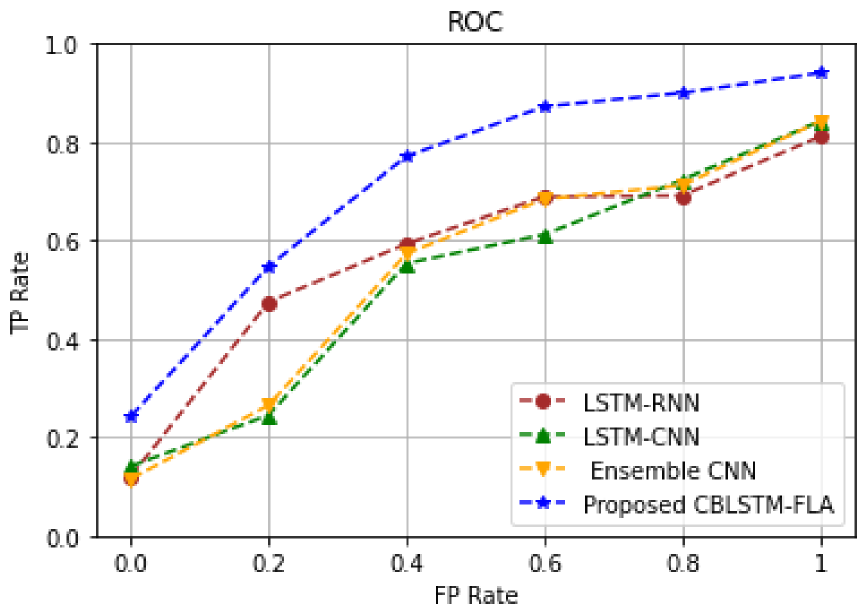

Figure 6. *ROC comparison of proposed models on review datasets.*

4.3 Evaluation for Twitter datasets

The short reviews in the Twitter datasets such as Airline Twitter and Sentiment140 are evaluated by evaluation metrics using the proposed CBLSTM-FLA model. The evaluated results are shown in Table 9 for the Twitter dataset and in Table 10 for the Sentiment140 dataset. Compared to the conventional sentiment analysis systems, the proposed model secured improved accuracy, recall precision and F1-score values, which proves the efficiency of the proposed model.

Table 9. *Performance of CBLSTM-FLA on airline Twitter dataset.*

Method	Accuracy	Precision	Recall	F1-score
LSTM with RNN	0.927	0.919	0.908	0.901
LSTM with CNN	0.918	0.902	0.890	0.903
Ensemble CNN	0.894	0.887	0.867	0.872
Proposed CBLSTM-FLA	0.978	0.962	0.976	0.958

Table 10. *Performance of CBLSTM-FLA on Sentiment140 dataset.*

Method	Accuracy	Precision	Recall	F1-score
LSTM with RNN	0.916	0.905	0.912	0.903
LSTM with CNN	0.934	0.927	0.928	0.918
Ensemble CNN	0.927	0.923	0.921	0.908
Proposed CBLSTM-FLA	0.972	0.964	0.971	0.962

The average performance of the proposed and existing methods for the Twitter dataset is shown in Figure 7. It shows that the proposed model is superior with the values of 0.975, 0.963, 0.973 and 0.96 for accuracy, precision, recall and F1-score, respectively. As for the other approaches, the LSTM-RNN secured 0.921, 0.912, 0.91 and 0.902 respectively, the LSTM-CNN yielded 0.926, 0.914, 0.909 and 0.9105 respectively, and the ensemble CNN achieved 0.9105, 0.905, 0.894 and 0.89 respectively. An ROC comparison of the proposed and existing approaches is shown in Figure 8, which proves that the proposed model achieved a better receiver operating characteristic (ROC) value of 0.9 compared to the other approaches.

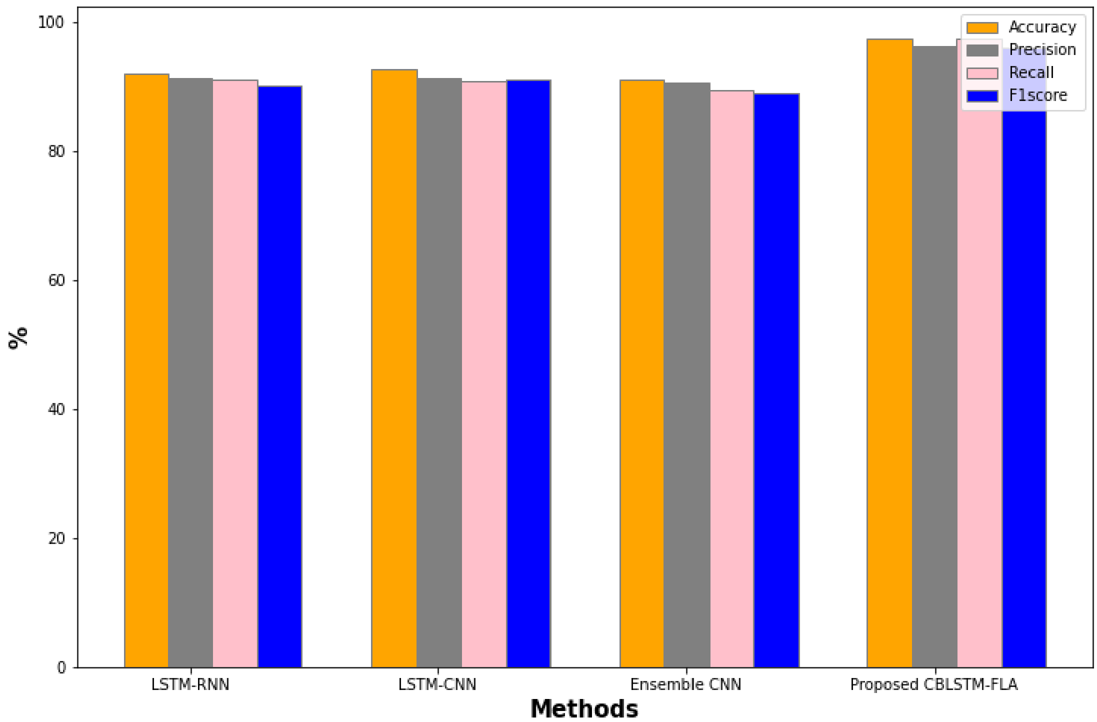

Figure 7. *Average performance of proposed model on Twitter datasets.*

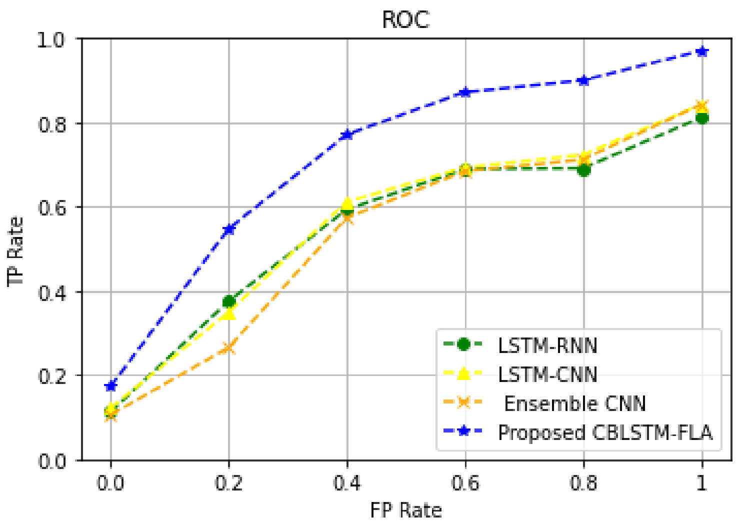

Figure 8. *ROC comparison of proposed model on Twitter datasets.*

An evaluation in terms of average error comparison of existing and proposed sentiment analysis models for the review and Twitter datasets is shown in Figures 9 and 10. For both the datasets, the proposed model achieved an average error rate of 0.101 for the review datasets and 0.09 for the Twitter datasets.

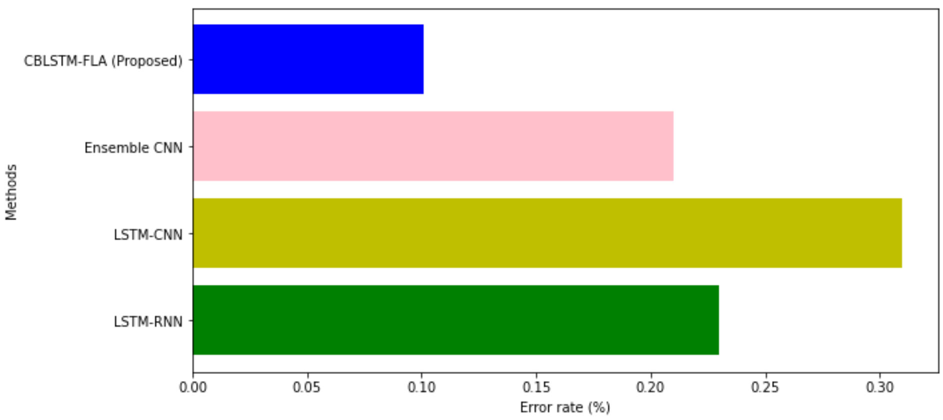

Figure 9. *Average error rate on review datasets.*

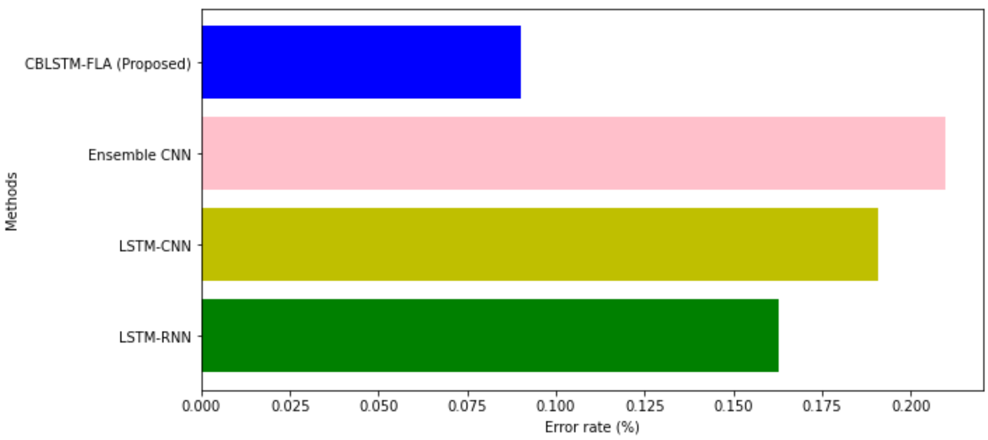

Figure 10. *Average error rate on Twitter datasets.*

It can be observed that the error rate of CNN-based models is higher compared to the LSTM-RNN and the proposed model due to its convolution process. In this paper, the complexity of the convolution process is reduced with the implementation of the MH algorithm since the hyperparameters are optimized and the best possible values are chosen using the FLA model. Hence, mathematical evaluation and experimentation outcomes show the efficiency of the proposed model. Compared to the other sentiment analysis models, the proposed CBLSTM-FLA secured improved accuracy, precision, recall, and F1-score with a reduced error rate. It classifies the user emotions efficiently and analyses user sentiments effectively.

5 Conclusion

In this paper, sentiment analysis of long review datasets and short Twitter datasets are analysed using deep learning and MH algorithms. The considered reviews and tweets are a mixture of different emotions and words. We developed a DL model based on MH for sentiment analysis. Initially, the input raw data were pre-processed to remove special symbols, stop words, numerals, and so on. From the pre-processed data, the features were selected using the TF-IDF approach. The selected features were fed as input to Doc2Vec for word embeddings. Here, we implemented convolutional Bi-LSTM with FLA to classify the sentiments into positive or negative emotions. The positive emotions were classified as fun, love, neutral, enthusiasm, happiness, surprise and relief, and the negative emotions were classified as anger, emptiness, sadness, worry and boredom. The proposed emotion-based sentiment analysis using CBLSTM-FLA was implemented using review and Twitter datasets.

The effectiveness of this model was compared with conventional models such as LSTM-RNN, LSTM-CNN and ensemble CNN. The results show a better performance of the proposed model in analysing sentiments based on emotions. Compared to other approaches, the proposed model secured improved average accuracy of 98.1% for the review datasets and 97.5% for the Twitter datasets. In terms of error rate, the proposed model achieved a reduced average error rate of 0.101% for the review datasets and 0.09% for the Twitter datasets. Hence, our proposed model is efficient and effective in classifying the review and Twitter emotional messages. In the future, sentiment analysis with personality analysis needs investigation to personalize the system more effectively.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: S.Y: Conceptualization, Methodology, Software, Visualization, Investigation. N.G: Data curation, Writing – Original draft preparation, Supervision, Software, Validation, Writing – Reviewing and Editing.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the use of the external and publicly available datasets.

Data availability: The datasets used in this article were also used in previous research, see (Al-Deen et al., 2021; Basiri et al., 2021). The datasets are also available from https://kaggle.com/datasets (use dataset name as keyword).

References

Al-Deen, H. S. S., Zeng, Z., Al-Sabri, R., & Hekmat, A. (2021). An Improved Model for Analyzing Textual Sentiment Based on a Deep Neural Network Using Multi-Head Attention Mechanism. *Applied System Innovation*, 4(4), Article no. 85. https://doi.org/10.3390/asi4040085

Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Generation Computer Systems*, 115, 279–294. https://doi.org/10.1016/j.future.2020.08.005

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge Based Systems*, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134

Damarta, R., Hidayat, A. M., & Abdullah, A. G. (2021). The application of k-nearest neighbors classifier for sentiment analysis of PT PLN (Persero) twitter account service quality. *Journal of Physics*, 1722(1), 012002. https://doi.org/10.1088/1742-6596/1722/1/012002

Dang, N. C., Moreno, M. N., & De La Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, 9(3), 483. https://doi.org/10.3390/electronics9030483

Dashtipour, K., Gogate, M., Li, J., Jiang, F., Kong, B., & Hussain, A. (2020). A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing*, 380, 1–10. https://doi.org/10.1016/j.neucom.2019.10.009

Datta, S., & Chakrabarti, S. (2021). Aspect based sentiment analysis for demonetization tweets by optimized recurrent neural network using fire fly-oriented multi-verse optimizer. *Sadhana – Academy Proceedings in Engineering Sciences*, 46(2), Article 79. https://doi.org/10.1007/s12046-021-01608-1

Da'u, A., Salim, N., Rabiu, I., & Osman, A. (2020). Recommendation system exploiting aspect-based opinion mining with deep learning method. *Information Sciences*, 512, 1279–1292. https://doi.org/10.1016/j.ins.2019.10.038

De Haro-García, A., García, G. C., & García-Pedrajas, N. (2020). Ensembles of feature selectors for dealing with class-imbalanced datasets: A proposal and comparative study. *Information Sciences*, 540, 89–116. https://doi.org/10.1016/j.ins.2020.05.077

Dessì, D., Dragoni, M., Fenu, G., Marras, M., & Recupero, D. R. (2020). Deep Learning Adaptation with Word Embeddings for Sentiment Analysis on Online Course Reviews. In *Algorithms for intelligent systems* (pp. 57–83). Springer Nature. https://doi.org/10.1007/978-981-15-1216-2_3

Eusuff, M., & Lansey, K. (2003). Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm. *Journal of Water Resources Planning and Management*, 129, 210–225. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:3(210)

Hameed, Z., & Garcia-Zapirain, B. (2020). Sentiment Classification Using a Single-Layered BiLSTM Model. *IEEE Access*, 8, 73992–74001. https://doi.org/10.1109/access.2020.2988550

Imran, M., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. *IEEE Access*, 8, 181074–181090. https://doi.org/10.1109/access.2020.3027350

Jadidoleslam, M., Bijami, E., Amiri, N., Ebrahimi, A., & Askari, J. (2012). Application of Shuffled Frog Leaping Algorithm to Long Term Generation Expansion Planning. *International Journal of Computer and Electrical Engineering*, 4(2), 115–120. https://doi.org/10.7763/ijcee.2012.v4.461

Jain, D., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. *Applied Soft Computing*, 91, 106198. https://doi.org/10.1016/j.asoc.2020.106198

Karakoyun, M., & Babalik, A. (2015). Data Clustering with Shuffled Leaping Frog Algorithm (SFLA) for Classification. In *International Conference on Intelligent Computing, Electronics Systems and Information Technology*. https://doi.org/10.15242/iae.iae0815009

Kastrati, Z., Dalipi, F., Imran, M., Nuci, K. P., & Wani, M. A. (2021). Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. *Applied Sciences*, 11(9), Article 3986. https://doi.org/10.3390/app11093986

Ligthart, A., Rodriguez, D., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review, 54(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

Meng, W., Wei, Y., Liu, P., Zhu, Z., & Yin, H. (2019). Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM. *IEEE Access*, 7, 167240–167249. https://doi.org/10.1109/access.2019.2952888

Pimpalkar, A., & Raj, J. R. (2022). MBiLSTMGloVe: Embedding GloVe knowledge into the corpus using multi-layer BiLSTM deep learning model for social media sentiment analysis. *Expert Systems With Applications*, 203, 117581. https://doi.org/10.1016/j.eswa.2022.117581

Ren, F., Feng, L., Xiao, D., Cai, M., & Cheng, S. (2020). DNet: A lightweight and efficient model for aspect based sentiment analysis. *Expert Systems With Applications*, 151, 113393. https://doi.org/10.1016/j.eswa.2020.113393

Safa, R., Bayat, P., & Moghtader, L. (2021). Automatic detection of depression symptoms in twitter using multimodal analysis. *The Journal of Supercomputing*, 78(4), 4709–4744. https://doi.org/10.1007/s11227-021-04040-8

Shams, M., Khoshavi, N., & Baraani-Dastjerdi, A. (2020). LISA: Language-Independent Method for Aspect-Based Sentiment Analysis. *IEEE Access*, 8, 31034–31044. https://doi.org/10.1109/access.2020.2973587

Shilpa, P., Shereen, R., Jacob, S., & Vinod, P. (2021). Sentiment Analysis Using Deep Learning. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks* (ICICV). IEEE. https://doi.org/10.1109/icicv50876.2021.9388382

Singh, J., Singh, G., & Singh, R. (2016). A review of sentiment analysis techniques for opinionated web text. *CSI Transactions on ICT*, 4(2–4), 241–247. https://doi.org/10.1007/s40012-016-0107-y

Sitaula, C., Basnet, A., Mainali, A., & Shahi, T. B. (2021). Deep Learning-Based Methods for Sentiment Analysis on Nepali COVID-19-Related Tweets. *Computational Intelligence and Neuroscience*, 2021, Article ID 2158184. https://doi.org/10.1155/2021/2158184

Su, M., Wu, C., Huang, K., & Hong, Q. (2018). LSTM-based Text Emotion Recognition Using Semantic and Emotional Word Vectors. In *2018 First Asian Conference on Affective Computing and Intelligent Interaction*. IEEE. https://doi.org/10.1109/aciiasia.2018.8470378

Wei, H., Sankaranarayanan, J., & Samet, H. (2020). Enhancing local live tweet stream to detect news. *Geoinformatica*, 24(2), 411–441. https://doi.org/10.1007/s10707-019-00392-9

Yadav, N., Kudale, O., Gupta, S., Rao, A., & Shitole, A. (2020). Twitter Sentiment Analysis Using Machine Learning For Product Evaluation. In *2020 IEEE International Conference on Inventive Computation Technologies* (pp. 181–185). IEEE. https://doi.org/10.1109/icict48043.2020.9112381

Editorial record: The article has been peer-reviewed. First submission received on 5 November 2022. Revisions received on 28 November 2022, 27 December 2022 and 6 January 2023. Accepted for publication on 9 January 2023. The editor in charge of coordinating the peer-review of this manuscript and approving it for publication was Zdenek Smutny.

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951