# Enhanced Long Short-Term Memory (ELSTM) Model for Sentiment Analysis

Dimple Tiwari[1] and Bharti Nagpal[2]
[1]Ambedkar Institute of Advanced Communication Technologies and Research, Guru Gobind Singh Indraprastha University, India
[2]NSUT East Campus (Ambedkar Institute of Advanced Communication Technologies and Research), India

**Abstract:** *Sentiment analysis is used to embed an extensive collection of reviews and predicts people's opinion towards a particular topic, which is helpful for decision-makers. Machine learning and deep learning are standard techniques, which make the process of sentiment analysis simpler and popular. In this research, deep learning is used to analyze the sentiments of people. It has an ability to perform automatic feature extraction, which provides better performance, a more vibrant appearance, and more reliable results than conventional feature-based techniques. Traditional approaches were based on complicated manual feature extractions that were not able to provide reliable results. Therefore, the presented study aimed to improve the performance of the deep learning approach by combining automatic feature extraction with manual feature extraction techniques. The enhanced ELSTM model is proposed with hyper-parameter tuning in previous Long Short-Term Memory (LSTM) to get better results. Based on the results, a novel model of sentiment analysis and novel algorithm are proposed to set the benchmark in the field of textual classification and to describe the procedure of the developed model, respectively. The results of the ELSTM model are presented by training and testing accuracy curve. Finally, a comparative study confirms the best performance of the proposed ELSTM model*.

**Keyword:** *Deep learning, convolutional neural network, recurrent neural network, long short-term memory, term frequency-inverse document frequency, glove, natural language processing*.

*Received August 24, 2020; accept April 7, 2021*
*https://doi.org/10.34028/iajit/18/6/12*

## 1. Introduction

Sentiment Analysis is an application of Natural Language Processing (NLP) that focuses on extracting the emotions of a people from text, images or videos. The latest technological advancements and development in data analytics, online shopping, politics, economic, statistics, increased the demand of sentiment analysis. Sentiment analysis plays a vital role in various fields, including decision making [3]. It has also used for analyzing the behaviour of a customer [16]. In social media, analysis tweets are more popular for calculating the sentiments due to their variety and easiness [9]. Making the system that will be able to achieve decision-based on the user-provided information require storing that information [35].

In the Conventional Neural Networks (CNN), all the inputs and outputs are independent so, it was unable to memorize the previous information. Therefore, Recurrent Neural Network (RNN) was introduced to overcome this problem by adding various hidden layers. The first and most essential feature of RNN is an unknown state that acts as a memory to store all related information of the previously calculated sequences. Long Short-Term Memory (LSTM) is a variant of RNN and the most suitable model for modeling long-distance interactions in a text, which occurs at the time of analyzing the sentiments [49].

RNN processes the inputs in a sequence and accurately works with long-distance dependencies to extract the sentence-level sentiments [5]. Deep learning applies various robust algorithms to solve the problems of different research fields, such as the activation function, stock marketing and sentiment analysis. In recent years, deep learning for neural networks has achieved the most success in various domains. For example, CNNs and feed-forward architectures perform well for image and video processing, including speech recognition [30]. On the other hand, RNN and LSTM are useful in the field of NLP and speech recognition. Deep learning appears to overcome the dilemma of gradient descent and, due to its robust computing infrastructure, can contribute a large amount of data processing and advancements in academia. LSTM overcomes the weakness of the feed-forward model by acquiring complete contextual information and character level embedding [21]. Although, LSTM is an attention mechanism for capturing the most essential and attentive information [7] but it does not work appropriately on post-word information because it runs in a single direction. Bi-directional LSTM (Bi-LSTM) solves this problem by learning with prefix and suffix [14, 29]. Considering the expansion of social networking and blogging websites, this work focused on extracting users' opinions from text data using sentiment analysis, which is gaining popularity and interest from researchers. Herein, the LSTM model was applied for the implementation of sentiment analysis on Internet Movie Database (IMDB) movie reviews [4] and Grand Old Party (GOP) debate datasets, which are commonly used by researchers as a baseline. A Global vector (Glove) word embedding is used for feature vectorization, which directly captured the global corpus statistics and performs better than word2vec [36]. It contains a pre-trained word embedding [15].

### 1.1. Major Contribution

1. This research work proposes a novel ELSTM model with hyper-parameter tuning for sentiment analysis.
2. Advanced feature extraction techniques have been applied to get better sentiment results.
3. A novel sentiment analysis framework has been proposed to set the benchmark for the researcher.
4. A novel sentiment analysis algorithm has been proposed to illustrate the process of sentiment analysis.
5. Experiments have been conducted on the benchmark set IMDB movie reviews and GOP debate datasets. The results show that the proposed model outperformed on both datasets without the problem of over-fitting and under-fitting.
6. Finally, we compared the proposed ELSTM model with CNN and previous research work from 2016 to 2020, which shows the greater performance of the ELSTM model than other models.

The rest of this paper is arranged as follows. Section 2 presents a literature review of previous works and models related to the sentiment analysis using deep learning. Section 3 describes the methodology of the research work, which includes one novel framework and one novel algorithm of sentiment analysis using the deep learning approach. Section 4 discusses the experimental results of the proposed ELSTM model. Section 5 provides the comparative results and discussion related to the improved LSTM model in the field of textual classification. Finally, section 6 concludes the research and outlines future work.

## 2. Literature Review

This section presents a brief overview of previous works in the field of sentiment analysis using deep learning. Cambria [10] suggest that sentiment analysis and effective computing have an eminent potential to become a building block for other technologies and systems. Fersini *et al*. [18] proposed Bayesian Ensemble learning for obtaining the peoples' opinions with efficient and effective results. Li *et al*. [32] introduced a novel recursive deep model to predict the sentiment label distribution effectively, which can also depict the semantic sentiments of any sentence, phrase, and length.

Pennington *et al*. [36] constructed a model that employs the main advantage of count data and, at the same time, captures the considerable linear sub-structures. Kim [27] stated that the Simple little hyper-parameter in CNNs provides reliable results with high efficiency. Santos and Gatti [42] proposed a deep convolutional network on two data sets, a Stanford Sentiment Tree-Book (SSTB) and a Stanford Twitter Sentiment corpus (STS), which approach outperformed by 85.7% for SSTB and 86.4% for STS. Bao *et al*. [6] developed a novel deep learning framework that combines LSTM, Stacked Auto-encoders (SAEs), and Wavelet Transforms (WT) for stock price forecasting. Chong *et al*. [12] presented a stock market deep learning feature-based prediction model that includes a restricted Boltzmann machine, auto-encoder and principal component model with a Deep Neural Network (DNN). Heaton *et al*. [24] proposed a novel deep learning hierarchical decision model for financial classification and predictions and stated that deep learning is probable to improve sometimes-dramatically for predicting the performance in conventional applications. Kraus and Feuerriegel [28] reported a neural network encompassing more than 50,000 parameters that can provide accurate predictions. The new model enhances existing literature by including semantics, context-related information, and explicitly incorporating word order. Lee and Yoo [31] introduced a novel method for constructing portfolios using the LSTM network of a specific risk-based on predicted returns and found that it can be possible to build the required degree of collection of performance and risk by adjusting the threshold. Sohangir *et al*. [43], used the StockTwits dataset to build LSTM and CNN stock market-related opinion mining. It has stated that CNN outperforms others for predicting the sentiments of a people. Chopra [13], proposed a novel model of deep learning for domain adaptation, which shows the difference between the distributions of training data and testing data.

Poria *et al*. [37] proposed a novel feature extraction approach from short text dependent on the value of the activation of an internal layer of a CNN. Which was proven suitable for the heterogeneous dataset, and this model achieves 14% improved performance. Poria *et al*. [38], proposed an algorithm that assigns contextual polarity in the form of text and flows, which uses a dependency arc for achieving the final polarity label of each sentence. Poria *et al*. [39] proposed a novel rule-based approach for aspect-level opinion mining that utilizes sentence dependency trees and common-sense knowledge for extracting implicit and explicit aspects. This model outperformed on two data sets and provides high accuracy. Friedrichs *et al*. [19] proposed a Shallow-CNN model of stack ensemble on the annotated dataset, and random hyper-parameter was applied to get the best ensemble prediction model that achieves the average f-score of 0.693. Graves *et al*. [22] proposed collaboration of the deep bidirectional-LSTM and RNN using weight, noise, and end-to-end training, which outperformed on the TIMIT dataset for phoneme recognition. In addition, deep LSTM achieved a 17.7% error in the TIMIT dataset, which is an acceptable result.

Nandan *et al*. [34] proposed the Approximate Extreme Points Support Vector Machine (AEPSVM) for reducing the time taken to train a large dataset, which demonstrated much faster performance speed and comparatively fast classification time than all previously applied methods. Tahon and Devillers [45] used five realistic corpora; three were collected from the framework, including a French project on robotics (ROMIO), a popular (AIBO) corpus, and a (JEMO) game corpus. These were used in combination with a non-classified features-based method (Gaussian, Mixture Model, and Information gain with Bhattacharya distance) using multi corpora distance. The obtained results were promising with a small set of 24 voices cepstral coefficients, indicating that the approach yielded better results than that of OpenSmile standard. Xu *et al*. [51] proposed improved CNN to investigate and predict the Outage Probability (OP) performance of mobile IoT communication networks and stated that CNN achieves better prediction than Radial Basis Function (RBF). Chen *et al*. [11] proposed graph convolutional feature-based CNN model for stock trends prediction and demonstrated that Graph-Convolutional feature-based Convolutional Neural Network (GC-CNN) outperforms over multiple methods.

## 3. Proposed Methodology

This section presents the proposed methodology of the enhanced sentiment analysis. The main objective of the work is to improve the accuracy of Recurrent Neural Network in the field of text classification. More features were added into the LSTM to build the baseline by combining various handy, favourable, and state-of-the-art mechanisms of both machine learning and deep learning techniques. Then, a novel algorithm and framework are proposed for sentiment analysis, by which a new social media sentiment standard can be established, providing a new direction for researchers for the sentiment analysis technique. The framework also offers greater prominence over the sentiment analysis approach and sets the benchmark in the field of NLP. Figure 1 depicts the framework of the proposed enhanced LSTM methodology.

Further portion of this section illustrates the step-by-step procedure of the proposed framework for applying the sentiment analysis process. We also compare the proposed ELSTM model with CNN and previous researches on LSTM-based sentiment analysis. The results indicate that our enhanced model outperforms the compared models and sets the benchmark in the field of sentiment analysis.

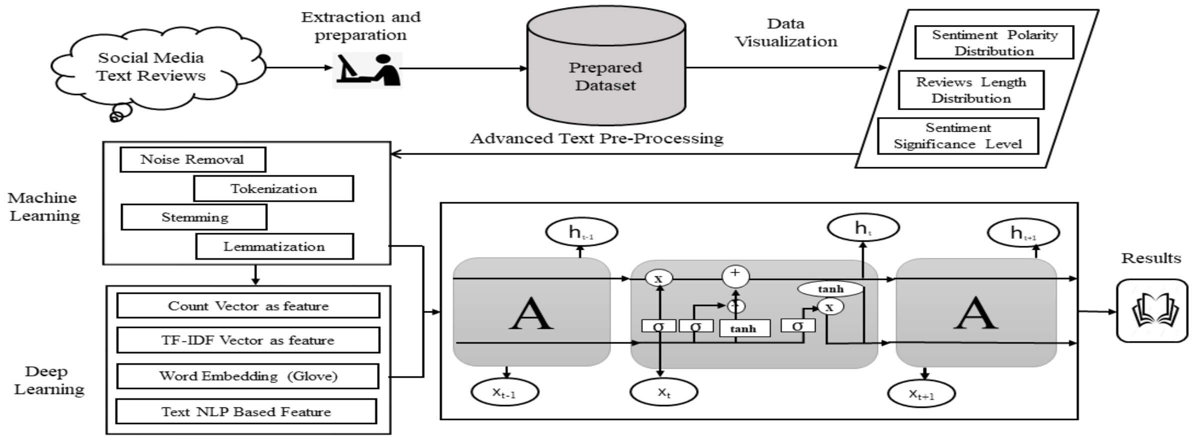



Figure 1. Proposed ELSTM framework for sentiment analysis.

### 3.1. Dataset Visualization

The proposed model was trained and tested on an IMDB movie reviews [48] and GOP debate datasets. Table 1 provides the full information regarding both the datasets. After obtaining the necessary information about the IMDB dataset, extra features of the reviews, such as text length and its distribution that describe the advanced information of text stored in the dataset were identified before analyzing the sentiment of reviews.

Table 1. IMDB and GOP dataset information.

| Properties | IMDB | GOP |
|---|---|---|
| Range-Index | 50000 | 21,458 |
| Data-Columns | 2 | 2 |
| Review | 50000 | 21,458 |
| Sentiments | 50000 (0 or 1) | 21,458 (0,1) |
| D-Types | object (2) | object (2) |
| Memory-Usage | 781.4+ KB | 251.5+KB |
| Polarity Distribution | 25,000-Negative (0) and 25,000-Positive (1) | 4,472-Positive and 16,986-Negative |

Figure 2 depicts the review length distribution of IMDB movie reviews dataset. Figure 3 depicts the review length distribution of GOP debate tweets dataset. Figure 4 describes the sentiments significance level of IMDB movie reviews dataset. Figure 5 describes the sentiment significance level of GOP debate tweets dataset.

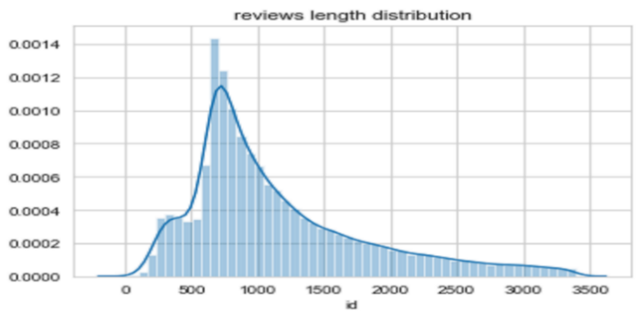



Figure 2. The IMDB dataset reviews length distribution.

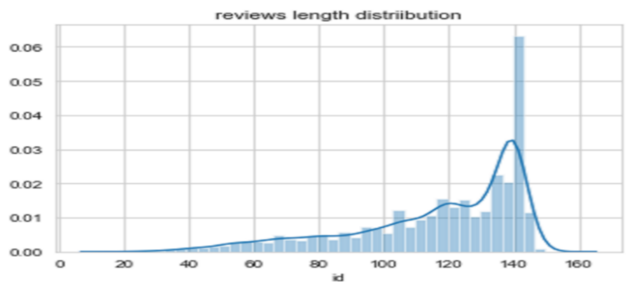



Figure 3. The GOP dataset reviews length distribution.

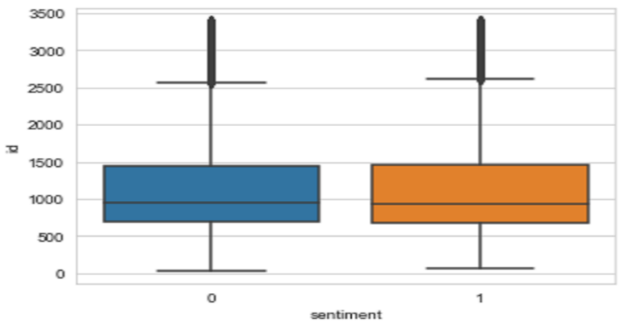



Figure 4. IMDB dataset sentiment significance level

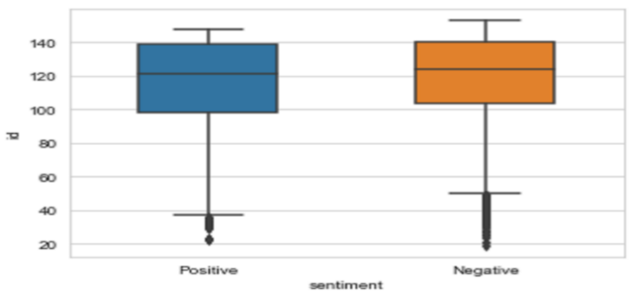



Figure 5. GOP dataset Sentiment significance level.

## 3.2. Data Preprocessing

To improve the performance of the proposed model, the combination of both machine learning and deep learning text preprocessing were applied. Preprocessing of dataset is an essential step, which discard the conflicts during learning phase of the model [8].

### 3.2.1. Noise Removal

Initially, noise removal was performed by removing punctuations, stop-words, numbers, urls, whitespaces and convert-to-lower [47]. Various punctuations, like ((), [], *, &, ^, %, $, #, @, /, \,), whitespaces, numbers, et al., that commonly occur in the text reviews were also removed before the further implementation of the advanced preprocessing steps.

### 3.2.2. Tokenization

In the process of tokenization, sensitive data of our dataset were replaced with unique and identifiable symbols, which break the sentence into various parts like phrase, words, or keywords that are called tokens. This essential information of text was acquired without breaking the security of the data. After that, subsequent processing has started with tokens [46].

### 3.2.3. Stemming

In the Stemming process, the stemmed words or root words, which have semantic similarity and same roots are segregated. To extract the necessary information from selected un-structured text, Stemming was applied to choose the same root words and eliminate the same words [23].

### 3.2.4. Lemmatization

Lemmatization was implemented to improve the preprocessing task done by Stemming. This process follows the morphological analysis rule to find the stem or root of the words using a dictionary to search for a link to its lemma in our dataset. Identifying the correct lemma of a particular word is a complicated task, since the word could have an inflected form with more than one lexeme, and every lexeme can have various lemmas [2].

### 3.2.5. Count Vector as Features

The count vector presents a matrix of the dataset, which displays every row in the form of a corpus, and the columns are represented as a term from the corpus. It merely works by combining both tokenized collections of a text document and known words vocabulary blocks. It has a "CountVectorizer" class with two primary functions, "Fit" and "Transform," that complete the task of count vector feature selection. The CountVectorizer counts the occurrence of each word in whole corpus.

### 3.2.6. TF-IDF Vectors as Features

Term Frequency-Inverse Document Frequency (TF-IDF) provides numerical statistic and calculates the importance of a particular word in a corpus, collection, or document. It is repeatedly used as a weighted factor in the search procedure of text mining and information retrieval. TF-IDF proportionally increases the value of a word at every occurrence in a document and filters all the stop words in a document [20]. It also lays a foundation for subsequent text by converting original text into a feature matrix of TF-IDF. The formula for TF-IDF is as follows [40]:

$$w_d = f_{w,d} * \log(|\mathrm{D}| / \mathrm{f}_{\mathrm{w,D}}) \quad (1)$$

Where *fw*, *d* represents the number of times *w* appears in *d*; |D| is the size of the corpus; and *fw*, *D* is the number of sentences in which *w* appears in *D*. Here, a few different situations can occur for each word related to the values of |D|, *fw*, *D*, and *fw*, *d*, which are the most remarkable to examine. The TF-IDF score combines two terms first: TF that computes the normalized term frequency and Inverse Document Frequency (IDF). It calculates the logarithmic number of sentences in the corpus, which is divided by the number of sentences where the specific term appears:

$$tf_{i,j} = \frac{n_{i,\mathrm{j}}}{\sum_k n_{i,j}} \quad (2)$$

Term Frequency (TF): *ni*, *j* represents the number of times *t* appears in a document, and ∑*k* represents the total number of terms in the text.

$$idf(\mathrm{w}) = \log(\frac{N}{df_t}) \quad (3)$$

IDF: *N* represents the number of documents, and *dft* is the number of documents containing term the t. TF-IDF vectors can originate at different token input levels, i.e., character, word, or N-grams-levels, for achieving a proposed task that TF-IDF has implemented at every level. Word-level TF-IDF represents the scores of TF-IDF of every term in different documents.

#### 3.2.7. Word Embedding

Word embedding was performed to determine the vocabulary of our document, which captures the lexicon and syntactical and semantic similarity with other words. Word embedding is a vector representation of words that focuses on ensuring the word has the same meaning. Various techniques like word2vec and Glove are available for word embedding of text data. Since Glove is the latest word embedding technique, an extension of word2vec, and more potent than Word2Vec, it was applied in this work. GloVe combines the functionality of both LSA and word2vec methods. The performance of GloVe vectorization has increase, as the number of datasets increases in training set [26].

#### 3.2.8. Text NLP-Based Features

In this step, numerous extra text-based features were created to increase the efficiency of the text classification method. Some relevant features, such as character count, word count, average word density, the total number of punctuations, number of the title word, number of upper counts in the document, verb count, noun count, adverb count, adjective count, and pronoun count, were extracted from the dataset.

### 3.3. Train RNN-LSTM for Sentiment Analysis

RNNs are a generalized version of Feed-Forward network due to their internal memory. RNN is a remarkably effective network which outperforms in NLP. To process the sequence of inputs, the RNN uses the internal state to perform the task of connected handwriting recognition. RNN is a remarkable effective network that to outperforms in natural language processing [17]. The basic concept of RNN is presented by the formula below:

$$h_t = f(\mathrm{h}_{t-1}, \mathrm{x}_t) \quad (4)$$

Where *ht* is the current hidden state; *f* is the function; *ht-1* is the previous hidden state, and *xt* presents the current input.

However, RNNs have some issues regarding gradient vanishing and exploding problems. Therefore, in this proposed methodology, the Long Short-Term Memory model was applied for sentiment analysis. This model is a modified version of an RNN, which improves the capability to remember previous data from the memory and uses back-propagation to train the model. It also performs better than RNN in learning long-term dependency. Herein, the LSTM model was trained on the IMDB dataset for the sentiment analysis process. Three major gates, including the forget, input, and output gates, were applied together to remove the issue related to the gradient vanishing cell of LSTM. These gates regulate the data flow-in and flow-out at the appropriate step. Table 2 presents working gates of the LSTM.

Table 2. LSTM gates.

| S-Number | Gate |
|---|---|
| 1. | $f_i = \sigma(\mathrm{W_f}[\mathrm{x_i}, \mathrm{h_{i-1}}] + \mathrm{b_f})$ |
| 2. | $I_i = \sigma(\mathrm{W_I}[\mathrm{x_i}, \mathrm{h_{i-1}}] + \mathrm{b_I})$ |
| 3. | $C_i = f_i * \mathrm{C}_{i-1} + I_i * C^{\sim}{}_i$ |
| 4. | $C^{\sim}{}_i = \tanh(\mathrm{W_c}[\mathrm{x_i}, \mathrm{h_{i-1}}] + \mathrm{b_c})$ |
| 5. | $o_i = \sigma(\mathrm{W_o}[\mathrm{x_i}, \mathrm{h_{i-1}}] + \mathrm{b_o})$ |
| 6. | $h_i = o_i * \tanh(\mathrm{C}_i)$ |

Where fi, Ii presents the weight; Oi represents the forget gate, input gate and output gate; Wf, Wi, Wo, bf, bi and bo indicate the matrix and bias scalar for every gate; hi is the hidden state; and Ci represents the cell [33]. In this work, a single LSTM was used, which typically encodes the sequence only from one direction.

### 3.4. Experimental Setup

This section presents the experimental setup of the proposed ELSTM model for sentiment analysis, which was validated on an IMDB and GOP debate datasets. The experiments were conducted to evaluate the effectiveness of the proposed model. Where datasets were split into two parts, 80% of data were selected for the training purpose and 20% were used for the testing set. After that, 6-folds cross-validation has applied to verify the accuracy and loss of the model on various iterations. Basically, cross-validation is used to check bias and variance, and for assuring the correct patterns of the dataset. The Algorithm 1, shows the subsequent steps of the proposed model. This algorithm presents the procedure of the whole experiment that has included in this research work, two-way sentiment classification (positive, negative) has performed to assess the model. The main focus of this research work is to Enhance Long Short-Term Memory (ELSTM) with hyperparameter tunning and by including machine learning and deep learning preprocessing steps.

*Algorithm 1: Enhanced LSTM model training for text sentiment analysis*

*Input* $X = \{x_1, x_2, \ldots\ldots\ldots\ldots, x_n\}$ *a set of reviews as on sentence basis*<br>
$Y = \{y_1, y_2, \ldots\ldots\ldots\ldots, y_n\}$ *sentiment labels of X either positive or negative (positive=1 and negative=0)*<br>
*Pre-Processing Noise removal and word normalization (listOfReviews)*<br>
*finalReviews =initialize list of reviews*<br>
*for review in listOfReviews do*<br>
*tempreview ← NoiseRemoval(review)*<br>
*tempwords ← Tokenization(tempreview)*<br>
*stemwords ← Stemming(tepmwords)*<br>
*normalizewords ← Lemmatization(stemwords)*<br>
*cleanReviews ← initializelist*<br>
*for keyword in cleanReviews do*<br>
*countwords ← CountVectors(keyword)*<br>
*TFIDFwords ← TF-IDFFeatureVector(countwords)*<br>
*embeddedwords ← WordEmbedding (TFIDFwords)*<br>
*textfeaturewords ← TextNLPFeatures(embeddedwords)*<br>
*topicmodels ← TopicModelling(textfeaturewords)*<br>
*cleanSet ← topicmodels*<br>
*finalReviews ← cleanReviews + cleanSet*<br>
*return finalReviews*<br>
*X=finalReviews*<br>
*TRAIN and TEST RNN-LSTM with input X and Y*<br>
*Fine tune LSTM with input X' and Y' to get ELSTM*<br>
*Get accuracy of* $TRAIN_{XY}$ *and* $TEST_{XY}$<br>
*return ELSTM*

### 3.5. Hyper-Parameter Settings

Glove word embedding and RNN-LSTM were used for sentiment classification, and various hyperparameters were set for acquiring the best results. Hyperparameter tuning was applied to solve the deep-learning problem optimally. Table 3 presents the information regarding the selected hyperparameters for the proposed model.

Table 3. Hyperparameter setting of the proposed model.

| **Hyperparameter** | **Value** |
|---|---|
| Class | Positive=1, Negative=0 |
| vocab-size | 92547 |
| Word-Embedding | Glove |
| **Layers Added** | |
| embedding-9 | Param# 9254700 |
| Lstm-13 | Param# 117248 |
| Dense-9 | Param# 129 |
| Neurons | 138 |
| Epochs | 6 |
| Verbose | 2 |
| Validation-split | 0.2 |

Various methods namely Grid search, Stochastic search, Heuristic search and Bayesian optimization are available for hyperparameter setting. Here, we used the stochastic search method to tune the hyperparameters of LSTM, which is capable to tune various hyperparameters simultaneously. Some critical values of hyperparameters tuned the LSTM model and improve the results.

## 4. Experimental Result

We have applied the proposed ELSTM model on two popular text datasets namely IMDB movie reviews and GOP debate for comparing the authenticity and accuracy of the model. The conducted experiments yielded state-of-the-art results for sentiment analysis on both the datasets. The accuracy of this model was calculated based on the standard accuracy procedure [41], as follows:

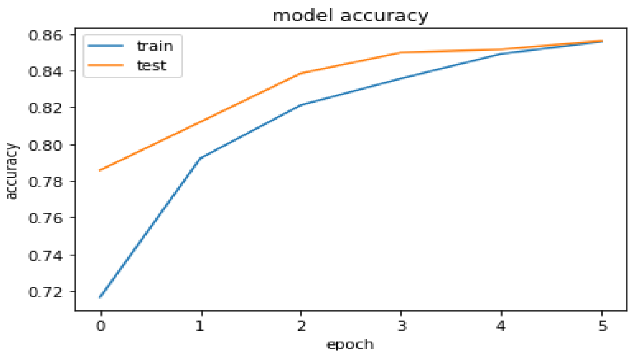
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Where the True Positive (TP) indicates the correctly predicted items; True Negative (TN) is the rightly rejected items; False Positive (FP) is the incorrectly predicted items; and False Negative (FN) is the improperly dismissed items. Accuracy was calculated as the ratio of correctly predicted elements to the total anticipated items. Table 4 depicts the training and testing accuracy score of proposed ELSTM on both the datasets.

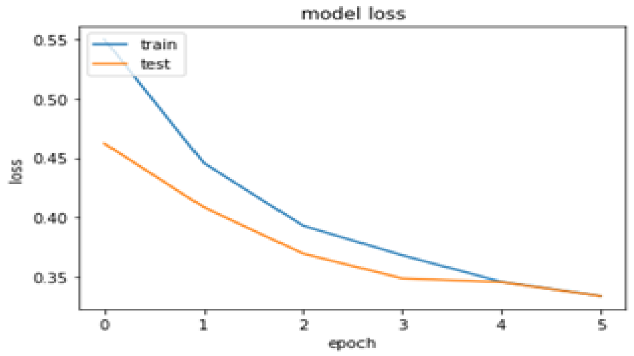
Table 4. The ELSTM accuracy score.

| **ELSTM** | **IMDB** | **GOP-Debate** |
|---|---|---|
| **Training-Accuracy** | 86.76% | 91.01% |
| **Testing-Accuracy** | 85.09% | 85.00% |

The results show that the proposed model archives a good accuracy score on both the experimented datasets. This is also proved that the model is neither over-fitted nor under-fitted as both the training and testing accuracies are very similar, further suggesting that ELSTM was trained appropriately and provides generalized results.

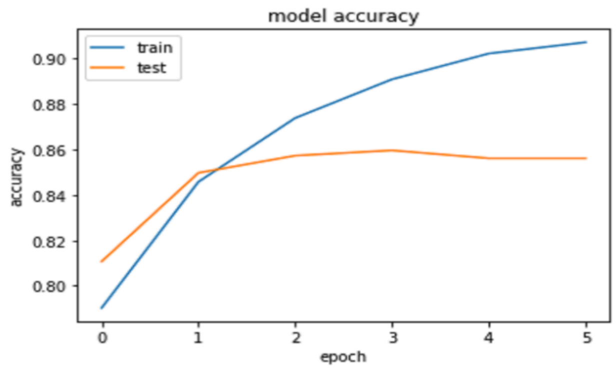



a) Accuracy on IMDB reviews dataset.

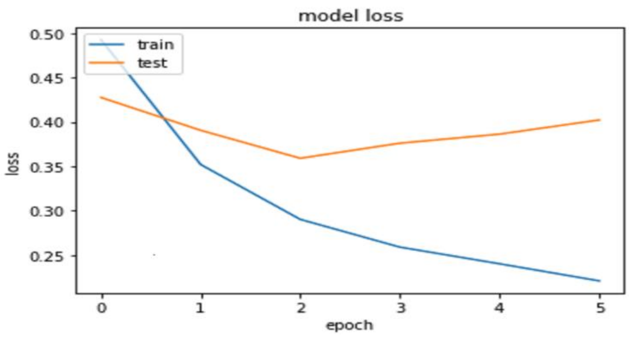



b) Loss on IMDB reviews dataset.

Figure 6. ELSTM model accuracy and loss on IMDB reviews dataset.




a) Accuracy on GOP debate dataset.




b) Loss on GOP debate dataset.

Figure 7. ELSTM model accuracy and loss on GOP debate dataset.

Figure 6-a) and Figure 7-a) presents the training and testing accuracies of the ELSTM model on IMDB movie reviews and GOP debate dataset, in which the x-axis represents the accuracy, and y-axes presents the six epochs, showing the total iteration of the model. It can be seen that at each iteration, the performance of the model improved step-by-step for both training and testing. Figure 6-b) and Figure 7-b) displays the loss of the proposed model on IMDB movie reviews and GOP debate dataset, whereby the model's accuracy increased at each epoch as model loss decreased. The margin between the training and testing accuracy reveals the best generalization of the model.

## 5. Comparative Results and Discussion

In this section, we compare our proposed model accuracy with CNN and diverse researcher's work, who use the same deep-learning model for the implementation of sentiment analysis. Table 5 presents the comparative results of the proposed ELSTM with various proposed models of LSTM over a year, who worked on text classification. Researchers modify LSTM according to their innovations; few add different layers in LSTM or few includes feature extraction techniques. Still, our proposed model ELSTM outperforms others and provides state-of-the-art results as obtains 85.09% testing accuracy score, which is more than all previously proposed LSTM models.

Table 5. Comparative results.

| Author | Title | Accuracy |
|---|---|---|
| Huang *et al.* [25] | Modelling Rich Context for Sentiment Classification with LSTM | 64.1% |
| Sosa [44] | CNN-LSTM and LSTM-CNN | 69.7%, 75.2% |
| Al-smadi *et al.* [1] | conditional random field classifier (Bi-LSTM-CRF) and Aspect-Opinion Target Expression (OTE) | 69.2%, 82.7% |
| Wen *et al.* [50] | Memristive LSTM (MLSTM) | 84.3% |
| **Proposed** | **ELSTM** | **85.09%** |

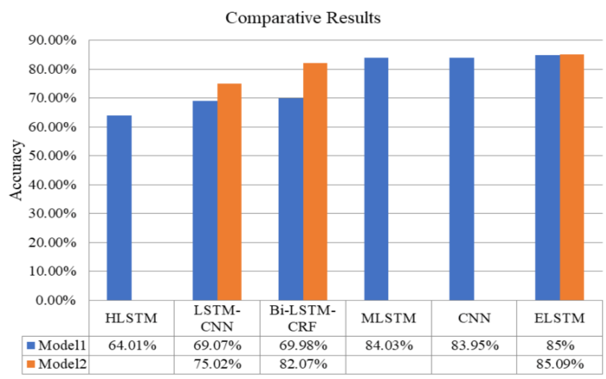



Figure 8. Comparative results.

We have also applied the CNN model on the IMDB dataset that generates 83.95% accuracy score. The accuracy score of the CNN model is less than that of the ELSTM model, which shows the greater performance of the proposed ELSTM model. Figure 8 presents the comparative results of the proposed ELSTM model and the other models. The presented bar chart shows that the accuracy of the LSTM model has increased over the year as researchers introduced innovations, where our proposed ELSTM achieves the higher accuracy score. As time efficiency, our proposed ELSTM model completes 6 cross-validations within 354 seconds and 909 milliseconds. Whereas, CNN completes execution within 376 seconds and 949 milliseconds. The ELSTM model is 22 seconds and 40 milliseconds faster than CNN, which shows our proposed model is more reliable and time-efficient than the previous approach. It is apparent that RNN-LSTM performed best proven for text classification problems among the various deep neural networks.

## 6. Conclusions and Future Work

This work demonstrates the feasibility of the deep-learning approach to extract the opinion of people. The researcher frequently chooses the LSTM neural network for text classification, which shows it is an appropriate model for text classification. This paper proposes an Enhanced LSTM (ELSTM) model for sentiment analysis. LSTM model is trained with hyperparameter setting on experimented datasets to improve the efficiency of the ELSTM model. The results obtained by the proposed ELSTM model shows that little hyperparameter tuning and a combination of machine-learning and deep-learning preprocessing techniques enhance the capability of the LSTM approach. Moreover, it is observed that feature extraction from text data using deep learning techniques has capable to boost the effectiveness of the model and provides reliable results. IMDB and GOP debate dataset has been used for presenting the performance of the ELSTM model. It has stated that ELSTM outperforms CNN and various researchers' models. Furthermore, for the benefit of future researchers, one sentiment analysis model and one algorithm have been proposed, which set a benchmark for researchers to follow the sentiment classification steps.

For future work, we have planned to implement another deep neural network for text classification. We intend to extend the domain of this model and try to optimize the performance of that model on multiple text datasets.

## References


[1] AI-Smadi M., Talafha B., Al-Ayyoub M., and Jararweh Y., "Using Long Short-Term Memory Deep Neural Networks for Aspect-Based Sentiment Analysis of Arabic Reviews," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 8, pp. 2163-2175, 2019.

[2] Aker A., Petrak J., and Sabbah F., "An Extensible Multilingual Open Source Lemmatizer," *in Proceedings of the International Conference Recent Advances in Natural Language Processing*, Varna, pp. 40-45, 2017.

[3] An H. and Moon N., "Design of Recommendation System for Tourist Spot Using Sentiment Analysis based on CNN-LSTM," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-11, 2019.

[4] Araque O., Corcuera-Platas I., Sánchez-Rada J., and Iglesias C., "Enhancing Deep Learning Sentiment Analysis with Ensemble Techniques in Social Applications," *Expert Systems with Applications*, vol. 77, pp. 236-246, 2017.

[5] Arras L., Montavon G., Müller K., and Samek W., "Explaining Recurrent Neural Network Predictions in Sentiment Analysis," *arXiv preprint arXiv: 1706.07206*, 2017.

[6] Bao W., Yue J., and Rao Y., "A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and Long-Short Term Memory," *PloS One*, vol. 12, no. 7, pp. e0180944, 2017.

[7] Baziotis C., Pelekis N., and Doulkeridis C., "Datastories at Semeval-2017 Task 4: Deep Lstm with Attention for Message-Level and Topic-Based Sentiment Analysis," *in Proceedings of the 11th International Workshop on Semantic Evaluation*, Vancouver, pp. 747-754, 2017.

[8] Bhati B., Rai C., Balamurugan B., and Al-Turjman F., "An Intrusion Detection Scheme Based on the Ensemble of Discriminant Classifiers," *Computers and Electrical Engineering*, vol. 86, pp. 106742, 2020.

[9] Bilgin M. and Köktaş H., "Sentiment Analysis with Term Weighting and Word Vectors," *The International Arab Journal of Information Technology*, vol. 16, no. 5, pp. 953-959, 2019.

[10] Cambria E., "Affective Computing and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 31, no. 2 102-107, 2016.

[11] Chen W., Jiang M., Zhang W., and Chen Z., "A Novel Graph Convolutional Feature Based Convolutional Neural Network for Stock Trend Prediction," *Information Sciences*, vol. 556 pp. 67-94, 2021.

[12] Chong E., Han C., and Park F., "Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies," *Expert Systems with Applications*, vol. 83, pp. 187-205, 2017.

[13] Chopra S., Balakrishnan S., and Gopalan R., "Dlid: Deep Learning for Domain Adaptation by Interpolating between Domains," *in the proceedings of the ICML, Workshop on Representation Learning*, Atlanta, 2013.

[14] Cliché M., "Bb-Twtr at Semeval-2017 Task 4: Twitter Sentiment Analysis with Cnns and Lstms," *arXiv preprint arXiv: 1704.06125*, 2017.

[15] De Clercq O., Lefever E., Jacobs G., Carpels T., and Hoste V., "Towards an Integrated Pipeline for Aspect-Based Sentiment Analysis in Various Domains," *in Proceedings of the 8th Workshop on Computational Approaches to Subjectivity,*

*Sentiment and Social Media Analysis*, Copenhagen, pp. 136-142, 2017.

[16] Dhandayudam P. and Krishnamurthi I., "Rough Set Approach for Characterizing Customer Behaviour," *Arabian Journal for Science and Engineering*, vol. 39, no. 6, pp. 4565-4576, 2014.

[17] Dyer C., Kuncoro A., Ballesteros M., and Smith N., "Recurrent Neural Network Grammars," *arXiv preprint arXiv: 1602.07776*, 2016.

[18] Fersini E., Messina E., and Pozzi F., "Sentiment Analysis: Bayesian Ensemble Learning," *Decision Support Systems*, vol. 68, pp. 26-38, 2014.

[19] Friedrichs J., Mahata D., and Gupta S., "Infynlp at SMM4H Task 2: Stacked Ensemble of Shallow Convolutional Neural Networks for Identifying Personal Medication Intake from Twitter," *arXiv preprint arXiv: 1803.07718*, 2018.

[20] Gabbard S., Yang J., and Liu J., "Quora Insincere Question Classification," *Baskin Engineering, University of California, Santa Cruz*, pp. 1-6, 2018.

[21] Graves A., "Generating Sequences with Recurrent Neural Networks," *arXiv preprint arXiv: 1308.0850*, 2013.

[22] Graves A., Mohamed A., and Hinton G., "Speech Recognition with Deep Recurrent Neural Networks," *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, pp. 6645-6649, 2013.

[23] Gupta R. and Jivani A., "Analyzing The Stemming Paradigm," *in Proceedings of International Conference on Information and Communication Technology for Intelligent Systems*, Ahmedabad, pp. 333-342, 2017.

[24] Heaton J., Polson N., and Witte J., "Deep Learning for Finance: Deep Portfolios," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 1, pp. 3-12, 2017.

[25] Huang M., Cao Y., and Dong C., "Modeling Rich Contexts for Sentiment Classification with Lstm," *arXiv preprint arXiv: 1605.01478*, 2016.

[26] Kim J., Tur G., Celikyilmaz A., Cao B., and Wang Y., "Intent Detection using Semantically Enriched Word Embeddings," *in Proceedings of IEEE Spoken Language Technology Workshop*, San Diego, pp. 414-419, 2016.

[27] Kim Y., "Convolutional Neural Networks for Sentence Classification," *arXiv preprint arXiv: 1408.5882*, 2014.

[28] Kraus M. and Feuerriegel S., "Decision Support from Financial Disclosures with Deep Neural Networks and Transfer Learning," *Decision Support Systems*, vol. 104, pp. 38-48, 2017.

[29] Lample G., Ballesteros M., Subramanian S., Kawakami K., and Dyer C., "Neural Architectures for Named Entity Recognition,' *arXiv preprint arXiv: 1603.01360*, 2016.

[30] LeCun Y., Bengio Y., and Hinton G., "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.

[31] Lee S. and Yoo S., "A New Method for Portfolio Construction using A Deep Predictive Model," *in Proceedings of the 7th International Conference on Emerging Databases*, pp. 260-266, 2018.

[32] Li C., Xu B., Wu G., He S., Tian G., and Zhou Y., "Parallel Recursive Deep Model for Sentiment Analysis, *in Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Ho Chi Minh, pp. 15-26, 2015.

[33] Ma Y., Peng H., Khan T., Cambria E, and Hussain A., "Sentic LSTM: A Hybrid Network for Targeted Aspect-Based Sentiment Analysis," *Cognitive Computation*, vol. 10, no. 4, pp. 639-650, 2018.

[34] Nandan M., Khargonekar P., and Talathi S., "Fast SVM Training Using Approximate Extreme Points," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 59-98, 2014.

[35] Ortigosa A., Martín J., and Carro R., "Sentiment Analysis in Facebook and its Application to E-Learning," *Computers in Human Behavior*, vol. 31, pp. 527-541, 2014.

[36] Pennington J., Socher R., and Manning C., "Glove: Global Vectors for Word Representation," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Doha, pp. 1532-1543, 2014.

[37] Poria S., Cambria E., and Gelbukh A., "Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, pp. 2539-2544, 2015.

[38] Poria S., Cambria E., Gelbukh A., Bisio F., and Hussain A., "Sentiment Data Flow Analysis By Means Of Dynamic Linguistic Patterns," *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, pp. 26-36, 2015.

[39] Poria S., Cambria E., Ku L., Gui C., and Gelbukh A., "A Rule-Based Approach to Aspect Extraction From Product Reviews," *in Proceedings of the Second Workshop on Natural Language Processing for Social Media*, Dublin, pp. 28-37, 2014.

[40] Ramos J., "Using Tf-Idf To Determine Word Relevance in Document Queries," *in Proceedings of the 1st Instructional Conference on Machine Learning*, pp. 29-48, 2003.

[41] Sangeetha K. and Prabha D., "Sentiment Analysis of Student Feedback Using Multi-Head Attention Fusion Model of Word and Context Embedding for LSTM," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4117-4126, 2021.

[42] Santos C. and Gatti M., "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts," *in Proceedings of COLING, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, pp. 69-78. 2014.
[43] Sohangir S., Wang D., Pomeranets A., and Khoshgoftaar T., "Big Data: Deep Learning for Financial Sentiment Analysis," *Journal of Big Data*, vol. 5, no. 1, pp. 1-25, 2018.
[44] Sosa P., "Twitter Sentiment Analysis Using Combined Lstm-Cnn Models," *Eprint Arxiv*, pp. 1-9, 2017.
[45] Tahon M. and Devillers L., "Towards a Small Set of Robust Acoustic Features for Emotion Recognition: Challenges," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 1, pp. 16-28, 2016.
[46] Tiwari D. and Singh N., "Ensemble Approach for Twitter Sentiment Analysis," *International Journal of Information Technology and Computer Science*, vol. 11, no. 8, pp. 20-26, 2019.
[47] Tiwari D. and Singh N., "Sentiment Analysis of Digital India using Lexicon Approach," *in Proceedings of 6th International Conference on Computing for Sustainable Global Development*, New Delhi, pp. 1189-1193, 2019.
[48] Wang S. and Manning C., "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," *in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, pp. 90-94, 2012.
[49] Wang X., Jiang W., and Luo Z., "Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts," *in Proceedings of COLING, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, pp. 2428-2437, 2016.
[50] Wen S., Wei H., Yang Y., Guo Z., Zeng Z., Huang T., and Chen Y., "Memristive LSTM Network for Sentiment Analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 3, pp. 1794-1804, 2019.
[51] Xu L., Wang J., Li X., Cai F., Tao Y., and Gulliver T., "Performance Analysis and Prediction for Mobile Internet of Things (IoT) Networks: A CNN Approach," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13355-13366, 2021.

**Dimple Tiwari** is a PHD scholar in Computer Science and Engineering Department at Ambedkar Institute of Advanced Communication Technologies and Research–AIACTR, Delhi, India. She is also a Microsoft certified in .Net framework. Her areas of interest include Sentiment Analysis, Artificial Intelligence, Information Security, Internet of things (IoT), Big Data. She has published various research papers in reputed International Journals/Conferences and contributed Book Chapters.

**Bharti Nagpal** is currently working as Assistant Professor in Computer Science and Engineering Department at Ambedkar Institute of Advanced Communications Technologies and Research–AIACTR, Delhi, India. She has 21 years of teaching experience. Her areas of interest include Sentiment Analysis, Artificial Intelligence, Information Security, Data mining and Data Warehouse, Internet of things (IoT), Big Data. She has published various research papers in reputed International Journals/Conferences and contributed Book Chapters.