International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Sep-2023)

http://dx.doi.org/10.12785/ijcds/140171

A Hybrid Schema: LSTM-BiLSTM with Attention Mechanism to Predict Emotion in Twitter Data

Niraj Kumar[1] and Subhash Chandra Yadav[2]

[1]*Department of Computer Science and Engineering,Central University of Jharkhand, Central University of Ranchi, India*
[2]*Department of Computer Science and Engineering,Central University of Jharkhand, Central University of Ranchi, India*

Received 05 Jan. 2023, Revised 14 Jul. 2023, Accepted 10 Aug. 2023, Published 01 Sep. 2023

Abstract: On the World Wide Web, people may now express their views and opinions in a novel way on a wide variety of subjects, trends, and issues, which includes advertising, political polling, knowledge-based surveys, market prediction, feeling and business intelligence. The user-generated content is available on various platforms, such as internet forums, discussion groups, and blogs, which serve as a concrete and significant basis for decision-making. Analysis of Emotion deals with the issue of extracting feelings from internet-based text data and classifying the author's reactive mental reply as fear, anger, happiness etc. The underline research proposes a sequentially appended Deep Neural Network architecture to bridge the gap between previous approaches such as Maximum Entropy, Gradient Descent, Random Forest, Naïve Bayes, and SVM(Support Vector Machine) used in machine learning. The model uses a balanced dataset to achieve enhanced accuracy and scalability. In the proposed architecture the first layer is the LSTM layer, which is used to process and sustain data in sequence for a long time. In the second layer, Bi-LSTM is appended for processing the flow of information in forward (past-directed to future)and backward (future-directed to past) directions and attached with an attention mechanism for predicting the output. The proposed framework is evaluated by Utilising various matrices, including the confusion matrix, recall, precision, and F-measure. Consequently, and is compared with the balanced dataset after handling the imbalance issue of different classes in the dataset. The model outperformed the actual dataset, which only had an accuracy of 90.87%, and reached a high accuracy of 96.53% in the sampled dataset.

Keywords: Deep Learning, Deep Neural Network, Emotion Analysis, Sentiment Analysis, Text Classification

1. Introduction

Emotion analysis is the way to identify the feelings or understand the frame of mind of a person expressed in the form of a text message. By analyzing the spoke person's emotions, a necessary action plan can be created. In day-to-day life emotions of humans play a vital role [1]. Emotion can be understood as an intuition of a person which differs from thought and knowledge. The personal ability of a person exert influenced by emotion and provides the ability to work in different circumstances and to control the response [2]. Acceptance of emotion is well-used in many fields like e-learning, Law, advertising, medicine etc. Besides human interaction, an emotion prediction system is used for psycho-social interventions to identify criminal inspiration [3]. Many efforts have been made till now to cognize facial and vocal emotions despite this, emotion detection of textual data still needs to be focused[4]. Predicting human emotions in the document becomes extremely helpful in language modelling. There are three major classes of sentiment names neutral, positive and negative. This sentiment class is broken up into several categories known as emotions. The positive sentiment includes emotions like happiness, pleasure, boldness, joy etc. The negative sentiment class comprises emotion as anger, sadness, disgust etc. Similarly, neutral sentiments include emotions of silence and others. Emotion can be thought of as the subset of sentiment expressed. Mathematically it can be expressed as:

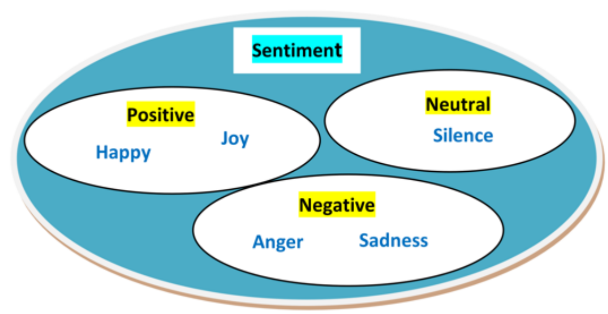

Figure 1. Association of Sentiment and Emotion.

Let us consider S to be the set of the sentiment and E to be the set of emotions S = { Positive, Negative, Neutral},

E-mail address: niraj.kumar@cuj.ac.in, dr.scyadav@cuj.ac.in

E= { happy, sadness, anger, joy etc.}
$E \subseteq S$
At first, the emotion analysis approach reads the text data of the tweets. The text data of the tweet is extracted from input data or the corpus or tweet. The text data is subjected to additional pre-processing in order to eliminate extraneous words and noisy data. From the pre-processed feature set, the model estimates similarity assessment for the different classes of emotion or sentiment. By only using the similarity value, the model performs emotion prediction by identifying end-user emotions, the predicted outcome is used to resolve several problems.
This paper come up with an improved hybrid deep learning model comprising of an LSTM layer stacked with Bi-LSTM which is attached to an attention mechanism to increase semantic understanding of text data for improving the classification accuracy.

2. Related Work

In recent years many techniques for emotion/sentiment analysis in text data have emerged, including supervised learning methodology and unsupervised learning methodology. In supervised learning, methods train the classifier with knowledge-crafted features such as n-gram and BoW (Bag of Word) to predict output; some of the classifiers used are Naive Bayes, maximum entropy, SVM(Support Vector Machine) etc. In unsupervised learning methods, it uses sentiment lexicons, grammatical parsing and syntactic design to make cluster structure of sentiment space in the text. The performance of sentiment analysis depends on the knowledge-based feature extraction method applied to the text data.
The strong potential of self-feature learning has empowered deep neural networks with more effective in classifying the emotion of text data in comparison to the previous handcrafted feature extraction method. Resulting in various deep learning methods like LSTM networks and CNN have originated. Zhao et al.[5] propose a hierarchical LSTM structured model by considering features based on user and features based on content to generate sentence representation and document representation named LSTM-MF to analyze the sentiment of text data. Zhou et al.[6] presented a sentiment detection model for Chinese micro blogs using word2vec as an embedding technique with more than one Bi-LSTM layer stacked for features extraction of word vector and evaluated the performance on CBOW and skip-gram methods. Wang et al. [7] applied deep learning techniques such as LSTM with CNN to find the sentiment to opinions posted in the stock market Twits dataset and found that DL models are very effective in predicting the sentiment of financial data. Smadi et al.[8] proposed a model for sentiment analysis of hotel reviews received in the Arabic language they use two layers of DL methods LSTM and Bi-directional LSTM along with attention aspect to identify the sentiment polarity of the review. Wang et al. [9] proposed a DL framework for sentiment analysis of text in the Chinese language, architecture comprising LSTM –L2 with Nadam optimizer for assessing the accuracy of sentiment in text. Ombabi et al.[10]proposed an Arabic-language sentiment evaluation model. The suggested approach makes use of two levels of LSTM for maintaining dependencies and a CNN layer for feature extraction. The model is assisted by an embedding technique called FastText. Finally, it passes through the SVM classifier for generating the final classification output. Wang et al. [11]designed a framework by combining CNN and RNN structures. Local features are produced by CNN and dependencies of a long time are learned by RNN for analysis of text messages to predict the sentiment behind them. Shelke et al.[12] suggested LRA-DNN (Leaky Relu activated Deep Neural Network) model. The model completes the classification task in four stages: pre-processing the text input, extraction of useful features, ranking, and classification of the data following the ranking step, when applicable ranks are given to each retrieved feature. According to the findings, the suggested LRA-DNN has a maximum accuracy of 94.77%.. Bhuvaneshwari et al.[13] developed a model(BAC) by integrating CNN, Self Attention Mechanism, and Bi-LSTM to classify the review's subjectiveness. The proposed developed model has a 91% F1-measure value and a model accuracy of 89%. Rani et al.[14] classified the user tweets received from the two states of the US using a hybrid deep learning method. The model combined CNN and LSTM to identify the emotional attitude of the individual.94% of the time, the model correctly classifies the various emotions.

3. Methodology

This study presents an attention-build LSTM and Bi-LSTM-based neural network, which locates the attention space to find out the keyword for representing emotional words and repeats sequentially to address the problem. The Glove pre-trained word embedding technique extracted all the pertinent details from the textual data. After that, LSTM is used to capture dependencies for the long term and learn the feature context of the input text representation. Simultaneously to take out the information, analyze and predict emotion contained in the text data Bi-LSTM method is deployed. Attention technology is applied to pay worthy notice to an entire document. The Gaussian Dropout layer is linked with the continuation of the input stage in order to stop the training data from overfitting.The following are the paper's contributions:

* Extraction of significant features has been done by using GloVe embedding technique for emotion analysis.
* A hybrid deep learning model comprising of an Attention-based LSTM-Bi-LSTM framework is proposed. The model acquires the advantage of LSTM and Bi-LSTM to improve model accuracy.
* The attention technique is attached for semantic understanding and to concentrates on different important words, which improves feature expression for analysis.

* After addressing the issue of an unbalanced class distribution, a comparative experiment is conducted using the actual dataset versus the well-balanced dataset.

A. *Handling Class Imbalance Issue*

Typically a problem of class imbalance issue occurs in most classification problems when there is more instance of a particular class than the other classes. In such cases, standard classifiers perform poorly and are impressed by the instance of more classes. The problem of class imbalance issue is not taken into account by the standard learning algorithm. The model pays the same attention while classifying into the dominant category as well as the marginalized category. The sampling approach is used to overthrow the class imbalance issue either by eliminating some instance of data from the majority class is known as under-sampling or by adding some duplicate instance of data or by adding artificially created data in the minority class called oversampling[15]Here, the issue of class imbalance is dealt with using an oversampling approach.

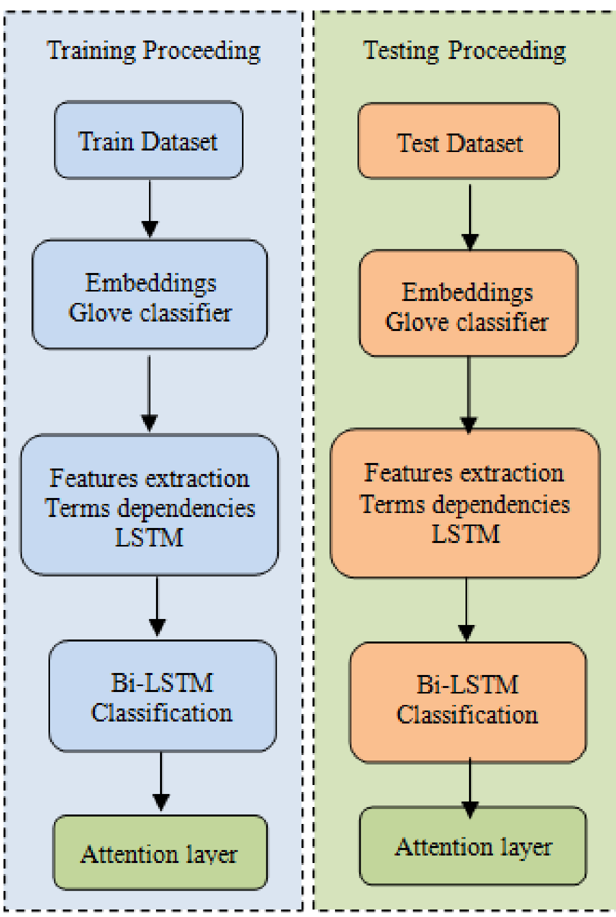

Figure 2. Emotion Analysis Overview framework pipeline

B. *Pre-Processing Text Data for Emotion Analysis*

A tweeter-based data contains a lot of noises, and missing values and are not in a usable format to be put directly in Deep learning architecture. Data pre-processing is a crucial step in preparing information to feed deep learning models; it also enhances the accuracy and efficiency of the model. Before providing input to the suggested architecture, the sub-sequence of NLP tasks is carried out mentioned in Table 1.

TABLE I. Task Performed during Pre-Processing Text Data

Task	Description
Expand contractions	remove all the shortened words with their full word
Stopwords	removed all text which has no significance in the analysis of emotion[16].
Mentions	removed word with @ and the name of the account
Punctuation	removed symbol used in between text such as”’’!()-[];:’”,¡¿./?@ etc[17].
Remove Emoji	all the emoji symbols are removed
Remove Url and HTML tag	all tags are eliminated from the tweet for better representation of the words.
Single Alphabet, Number and Blank space	it has no significant representation of the sentence hence removed.
Lower case conversion	to make symmetry in the sentence[18],[19].
Lemmatization	words are extracted to their correct lemma as per the dictionary data[20],[21].

C. *Word Embedding Technique*

Word embedding is the technique to map the word's vocabulary in its real vectors. It captures the syntactic and semantic relationship of words in a large dataset. The primary source for a model is the word occurrences statistics of a corpus for learning the representation of words from available information. Here to build a new model, the word representation method GloVe is employed, because the statistic generated by Glove vector's (Global corpus) is instantly captured by the model[22].

For mathematical representation. Let us consider word occurrence count matrix X and $X_l m$ represent the number of counts of occurrence of word m in the reference of word'l'. consider $X_i = \sum_n X_{ln}$ is a count of the number of any word that seems in the context of 'l' word. model takes the form,

$$F\left((w_l - w_m), w_n^T\right) = \left(\frac{P_l n}{P_m n}\right) \tag{1}$$

In equation(1) l,m, and n are three words. The relationship among these words is determined by the ratio of the probabilities of the co-occurrence of words concerning n. Here

$\frac{P_{l}n}{P_{m}n}$ is the word vector(probabilities of l w.r.t n/probabilities of m w.r.t n), $w \in \mathbb{R}^{D}$ represent word Vector and $w^{T} \in \mathbb{R}^{D}$ represent context word vector.In the above equation, the ratio of probabilities is determined by the corpus itself and function F may be based on some anonymous parameters.

D. Sustaining Data for Long Time

A solution which can address the problem of vanishing error is LSTM, based on method gradient base[23].LSTM network can bridge minimal time lags, It uses constant error carousels (CECs), to create an ongoing error flow inside the specific cell. Multiple gate units learn to provide control access to different cells. Let us consider that we have a single unit v connected to a single connection with v itself. At a time t internal backward flow error of v follows equation 2.The standard architectonics of LSTM is shown in Figure 3.

$$Q_{v}(t) = F'_{v}(z_{v}(t))\, w_{[v,v]} Q_{v}(t+1) \tag{2}$$

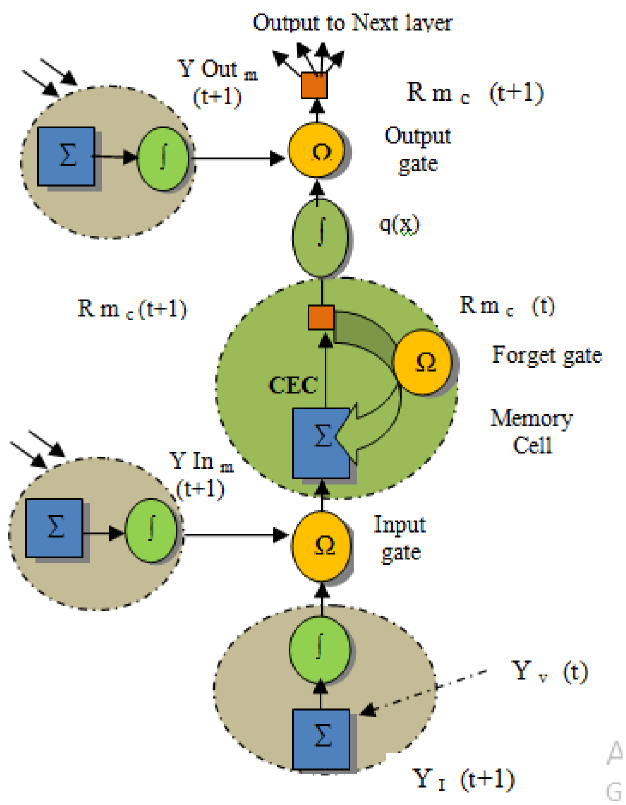

Figure 3. Memory Blocks of Standard LSTM

functionF_{v} is the identity function,$w_{[v,v]}$ called CEC's preserves the error and is the LSTM central features to achieve short-term memory to storage for the long period. LSTM extend CEC accompanied by the input layer of the network and the stored information memory cells are connected by input gates and output gates. A forget gate is connected to CEC, it resets the information in the memory cell that is no longer needed for the operation, creating the sophisticated LSTM block referred to as a memory cell.
In the Figure 3, It has the minimum of a single cell (CEC) block that is repetitive and fully self-connected, its weight set to'1' initially.sm_{c} symbolises the status of the CEC cell. The reading and writing approach is controlled through the gate that receives input(YIn)and the final product as output (YOut). By multiplying the input gate result by the g(x) squashed input, we get the internal cell's state. Following that, it is added to the duration of the time scale $Rm_{c}(t)$ the time scale $Rm_{c}(t+1)$ for achieving the present state. Then by multiplying the cell value with the activation mechanism of the resultant gate, the output of the cell is ultimately anticipated.
The input gate YIn_{m} activation function In_{m} is calculated as

$$YIn_{m}(t+1) = FIn_{m}\,(ZIn_{m}(t+1)) \tag{3}$$

The Out_{m}, output gate's activation mechanism operates as follows.

$$YOut_{m}(t+1) = FOut_{m}\,(ZOut_{m}(t+1)) \tag{4}$$

A non-linear squash function is used to scale the results of both gates $FIn_{m} = Fout_{m}$= F,to keep the value in the range[0,1], defined by

$$F(v) = \frac{1}{1+e^{-}v} \tag{5}$$

Activation function of forget gate Yc is computed as

$$Yc_{m}(t+1) = Fc_{m}\,(Zc_{m}(t+1){+}bc_{m}) \tag{6}$$

Where F is squash function with range of [0,1] and bc_{m} is bias of the forget gate.

E. Bi-LSTM for Emotion Analysis

In the proposed architecture Bi-LSTM has been used for emotion analysis of English tweets and learning to express the document in matrix representation. The sequence of information is processed in both directions in forward direction and backward directions in the different timestamps. The vector in the lower dimension obtained from pre-trained Glove word embedding passes through Bi-LSTM to compute input sequentially and create a document's hidden vector. After that, the vector is connected to the sigmoid function to perform emotions classification. The basic architecture as shown in Figure 5 To learn the context of the features Bi-LSTM Network is enforced to the output of the attention score and considers forward-to-backwards and backwards-to-forward features parallel and concurrently the concatenation of the hidden state of both LSTM to represent each position. Forward to backward LSTM is represented as equation 7 and Backward to forward LSTM is represented by equation 8.Figure 4 depicts the data procedure diagram for the model.

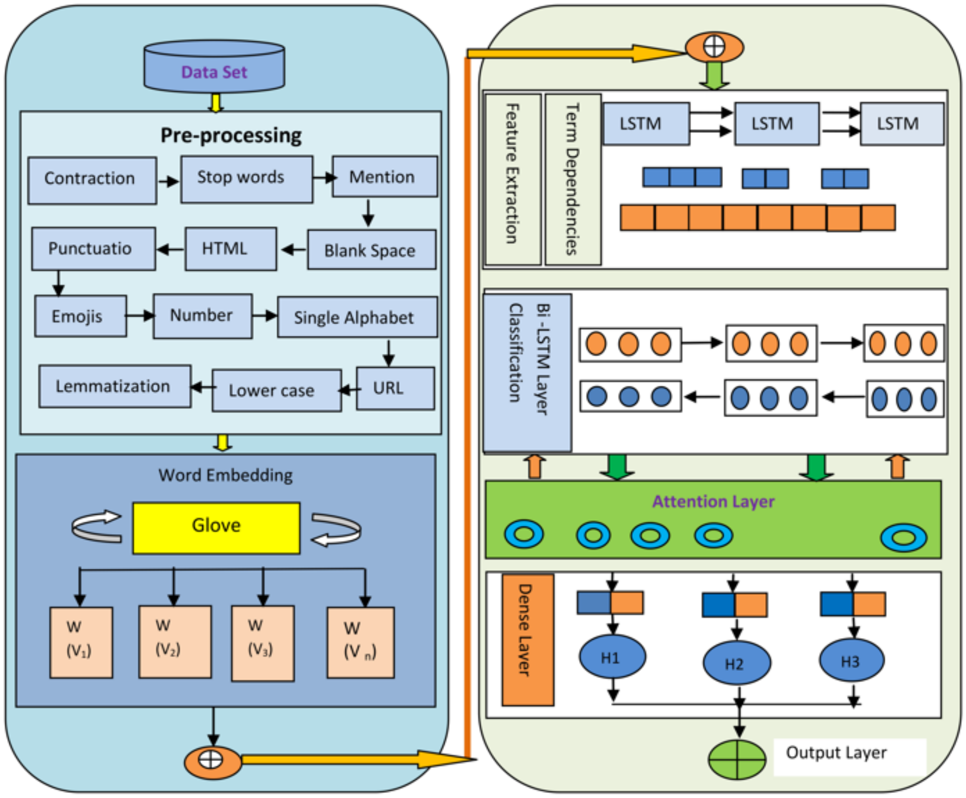

Figure 4. Emotion Analysis Proposed Model Flow Diagram

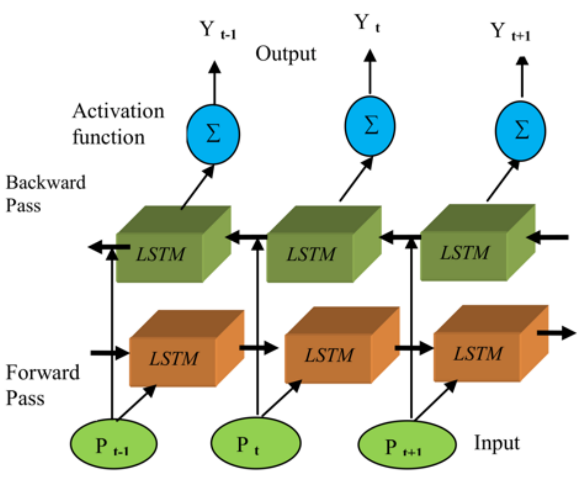

Figure 5. Information Flow Diagram of Standard Bi-LSTM

$$h_{tFLSTM} = F_{LSTM}(C_{t-1}, h_{t-1}, A_i) \tag{7}$$

$$h_{tBLSTM} = B_{LSTM}(C_{t-1}, h_{t-1}, A_i) \tag{8}$$

Here c_t represents the hidden state of the network,h_t represents the memory cell of the network and h_{t-1} represents the previous state of the network and c_{t-1} previous memory cell of the BiLSTM network.A_i is attention score passes as the input vector in the BiLSTM network.

F. *Attention Built Deep Layers*

Attention is a process of concentrating on selective things and ignoring the rest of the other things. The attention mechanism in deep NN is used to perform the same action to focus on only selective things and ignore the others.

In the below-given figure, Bi-LSTM is used to generate a chain of annotation $(t_1, t_2, ..., t_{nx})$ for all the input sentences. All the $t_1, t_2...$ vectors used in the process are the concatenation of the hidden state of the F$_{LSTM}$encoder and B$_{LSTM}$ encoder. Let us say the input sentence contains T_x annotation words. The context vector(c_i) considering the output word(y_i) is computed by adding all the weight of the annotations. It is represented by the below equation.

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \tag{9}$$

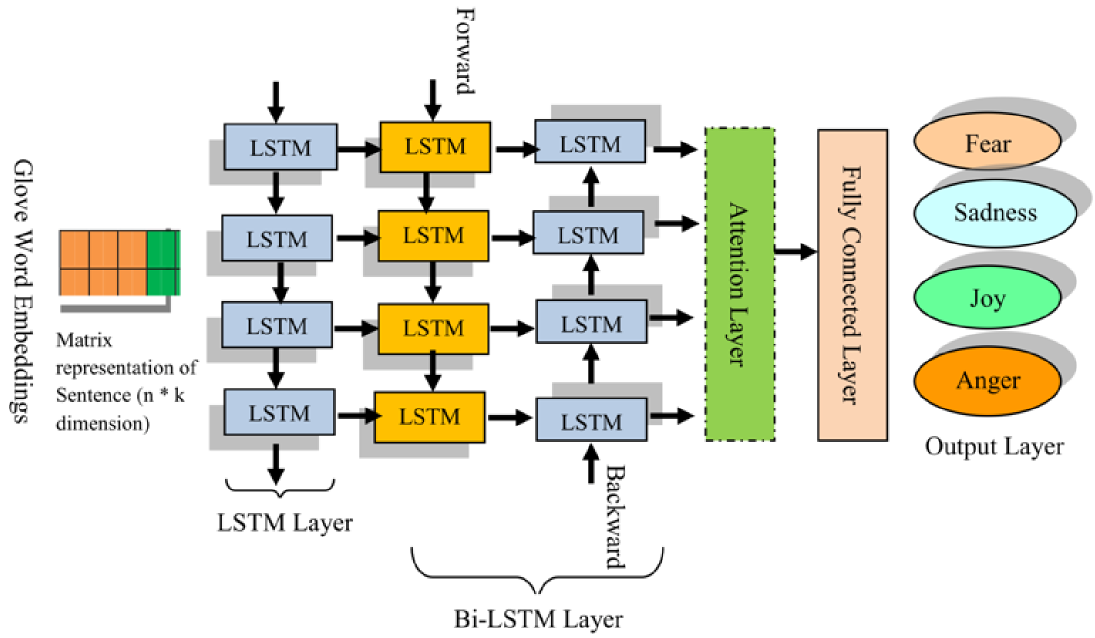

Figure 6. Proposed Model Architecture

The computation of weight $a_i j$ is done by softmax function represented by the below equation:

$$a_{ij} = \frac{exp(p_{ij})}{\sum_{k=1}^{T_x} exp(p_{ik})} \tag{10}$$

and

$$p_{ij} = f(s_{i-1}, h_j) \tag{11}$$

p_{ij} denote the output score of the neural network given by the function (f),which aims to record the sequence between input from j to out-put at i.The attention mechanism is shown in Figure 7. In attention, the BiLSTM softmax layers r are used for generating the conditional probability over the whole class space for emotion classification and to avoid over-fitting dropout layer is used. The effectiveness of the classification is frequently assessed using the Cross entropy loss function. Here adam optimizer function is used to reduce the network loss function and fine-tune the parameter so that the model performs the same as the backpropagation model. Cross entropy as a loss function has reduced the chance of gradient disappearance in the time of processing of SGD(stochastic gradient-descent).

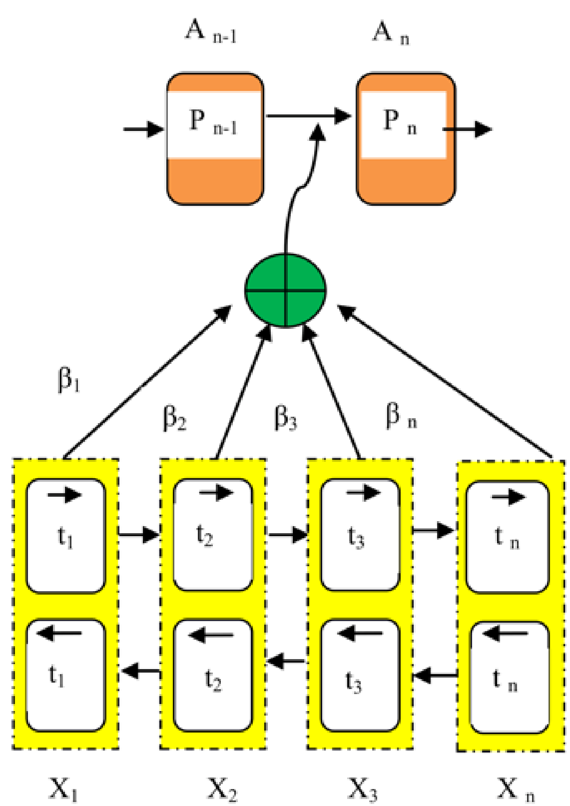

Figure 7. Attention Mechanism

The below equation represents the loss function:

$$L_{total} = \frac{-1}{Num} \sum_{sp} [y \ \ ln \ \ o + (1-y) * ln(1-o)] \tag{12}$$

G. Output Layer

The dense layer fully connected is used for transforming the bi-LSTM network to high dimensional sentiment characterization for prediction of tweet sentiment polarity. The output of this is obtained by the below given equation.

$$h_i = Softmax\left(W_i h_p + b_i\right) \tag{13}$$

In the above equation w_i, and b_i are the parameter which was learned in training the model,h_i denotes features obtained and the hp denote the feature map acknowledged by the bidirectional LSTM network. Emotion classification was performed by the output layer.

4. Result and Discussion

This section presents the experiment and briefly explains the outcome. The model is tested on the English tweet dataset, by considering the accurate analysis of the proposed method.

A. Dataset

Recognizing emotion is intrinsic to the NLP project, one can build a different robust model with the help of a dataset and conduct emotion analysis. Dataset used in the experiment is WASSA-2017 Shared Task on Emotion Intensity[24]based on the emotion intensity detected of a tweet felt by the spokesperson.

Task- For a tweet T given an emotion E, determine the degree of emotion E experienced by the user, which ranges from 0 to 1. The highest score of 1 denotes experiencing the most quantity of emotion E (or having an intellectual state at most prone to experiencing emotion E).The lowest possible score of 0 denotes, the undershot degree of emotion E (or a mental state that is as far away from emotion E as feasible). The tweet T and the associated feeling E are referred to as an illustration.here the absolute scores serve solely to illustrate that cases with higher scores correspond to a higher degree of emotion E than examples with lower scores. They have no vested significance. By using Best Worst scaling technique(BWS) dataset was created having 6375 tweets which are annotated in four types of emotion states: fear, anger, sadness and joy intensities.The structure of the dataset is shown in Table II and the comparison of the dataset after sampling is given in Table III. The algorithm of proposed model is shown in Table IV.

TABLE II. Dataset Structure

ID	Tweet	Emotion
1009	How is it suppose to work if you do that? Wtf dude? Thanks for pissing me off.	Anger
20249	Look at this # massiah of # youngleader# Pakistan # massiah of # terrorism	Fear
30250	I'm due for a big change! I've prayed on it, I think I deserve it # positivity	Joy
40483	Awareness of time is awareness of time lost. # awareness # time	Sadness

TABLE III. Comparison of Dataset Structure after Sampling Minority Classes

Emotion	Actual Dataset shape Counter	Re-sampled Dataset shape Counter
Anger	1701	2252
Fear	2252	2252
Joy	1616	2252
Sadness	806	2252

TABLE IV. Algorithm of Proposed Model

Algorithm for AT-LSTM-BiLSTM Model
1:Create a word embedding matrix using pre-trained GloVe word vectors represented with Eq.1;
2:Employed the LSTM layer for feature selection and sustaining data for the long term using Eq.2, Eq.3, Eq.4, Eq.5 and Eq.6;
3: Utilized BiLSTM for obtaining antecedent context features $h_t f$ and the supervene contextual features $h_t b$ from the predicted feature sequence represented by Eq.7 and Eq.8;
4: Employ attention layer to acquire a representation of the future content along with historical context from succeeding and preceding contextual features represented by Equation7 and 8;
5: Combine both the representations of future context and historical context to obtain extensive context representation;
6: Feed the extensive context characterization to the soft-max classifier for getting the labels of the classes;
7: Update the model parameters by using a loss function with the Adam optimizer presented in equation 12;

B. Experiment Structure

The inputs given to the model suggested, are initially embedded using the Glove embedding technique with 100 dimensions and additional parameters throughout the training phase. Gaussian dropout (0.3) is used at the input layer to avoid overfitting in the network. The LSTM layer eliminates superfluous information while preserving important information, which provides concise information to the next layer. The feature vector is extracted from the LSTM layer and sent to the bi-LSTM layer to create the context vector and classify the data. The output is then applied to the attention mechanism to generate the worry attention score. The bi-LSTM uses softmax regularisation and has a density of 128 sizes with 256 batch sizes. For emotion classification, It uses a softmax function.

TABLE V. Deep Learning Methods used in Proposed Model

Deep Learning Methods	Description
Gaussian dropout(0.3)	Avoid overfitting
bi-LSTM,Dense layer	Batch size-256,128
Softmax function	for classification
Binary cross entropy, Optimizer	for loss function, Adam

TABLE VI. Initializing Parameter of Proposed Model

Initializing Parameter	Description
dictionary Word Size	10000
Sequence Max length	24
Validation set size	1000
Epochs to start train	15
GLOVE word embedding dimension	100

Finally, we employ Adam's optimizer and binary-cross entropy as the loss function for training the model. The model was executed on window-10, processor core i5 and Ram of 16 GB. The operating system 64-bit has been used to run the dedicated user interface and Jupyter Notebook using a Python 3.7 version environment. Pandas and Num-py were used to pre-process the dataset. All the relevant histograms, graphical calculations and drawings were created using Python Software.

C. Result Analysis

The proposed classification algorithm with feature selection was implemented and tested with an unlabeled Twitter dataset for the evulsion of four types of emotion class intensities (Happy, Fear, Joy and Sadness). The model predicted very effectively with the preceding investigation. It is necessary to evaluate the model developed before putting it into practice to address a real-world issue. The performance measure of the developed model is the way to assess the performance for which context it was developed. The performance measure includes Accuracy; Precision; Recall; and F1 Score. Accuracy (Accy) is the proportion of the count of text data that were properly categorised and the total count of text data. Recall (Re) measures the ratio of correctly identified text data among all text data within the same class. Precision (Py) is the proportion of correctly identified text data among all acknowledged text data in this class. F1-measure indicates the symmetric average of the Precision and the recall. All these matrices are used to measure the performances of any deep learning-based models. Moreover, all these matrices utilized a confusion matrix to compute. The confusion matrix presented in

TABLE VII. Performance Evaluation Confusion Matrix

Actual Class	Predicted class(Classified Positive)	Predicted Class(Classified Negative)
Actual Positive	TP	FN
Actual Negative	FP	TN

Table VII is used to calculate other matrices. True positive denotes the number, which is properly labelled in their respective classes(TPj), and true negative represents the number, which is not properly classified to their respective class(TNj).On the other hand, a false positive indicates the number, which is improperly classified to their class(FPj) and the false negative represents the number, which is not identified in a particular class(FNj).

$$Accy = \frac{TP_j + TN_j}{TP_j + TN_j + FP_j + FN_j} \tag{14}$$

$$Re = \frac{TP_j}{TP_j + FN_j} \tag{15}$$

$$Py = \frac{TP_j}{TP_j + FP_j} \tag{16}$$

$$F1Score = \frac{2 * P_j * R_j}{P_j + R_j} \tag{17}$$

where j denotes dataset size

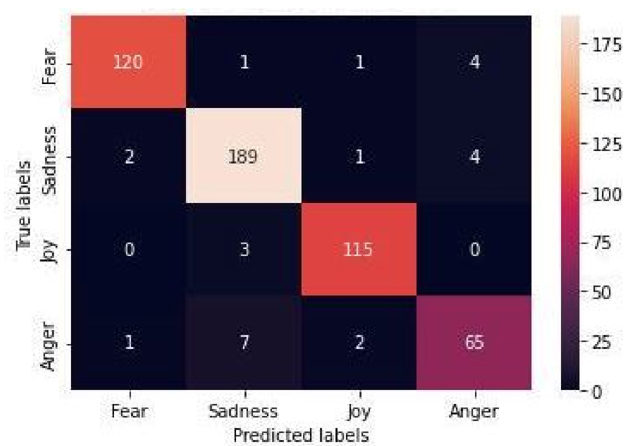

Figure 8. Confusion Matrix Actual Dataset

TABLE VIII. Performance Evolution of Proposed Model with Actual dataset

Emotion	Pre	Rec	F1
Fear	0.98	0.95	0.96
Sadness	0.94	0.96	0.95
Joy	0.97	0.97	0.97
Anger	0.89	0.87	0.88

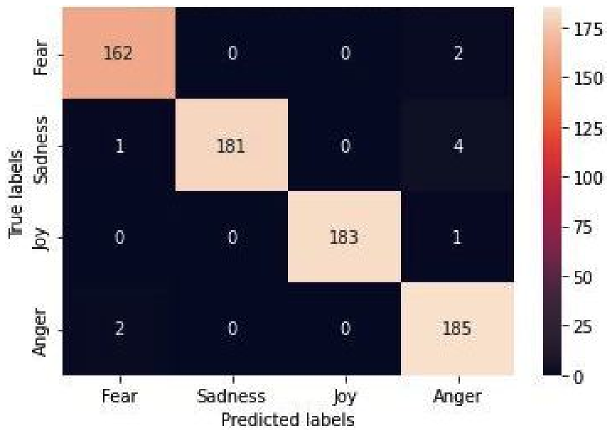

Figure 9. Confusion Matrix Sampled Dataset

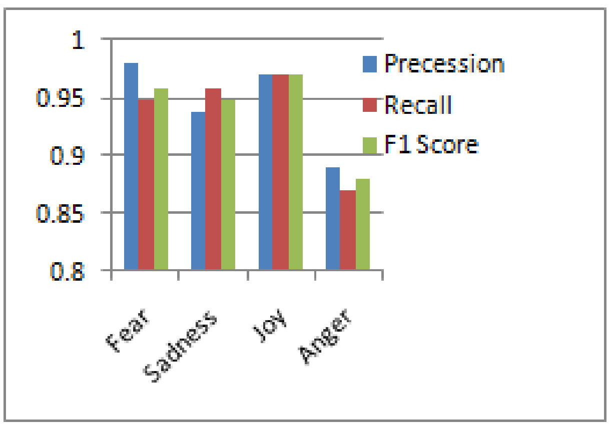

Figure 10. Precision; Recall; F1 Score of Actual Dataset

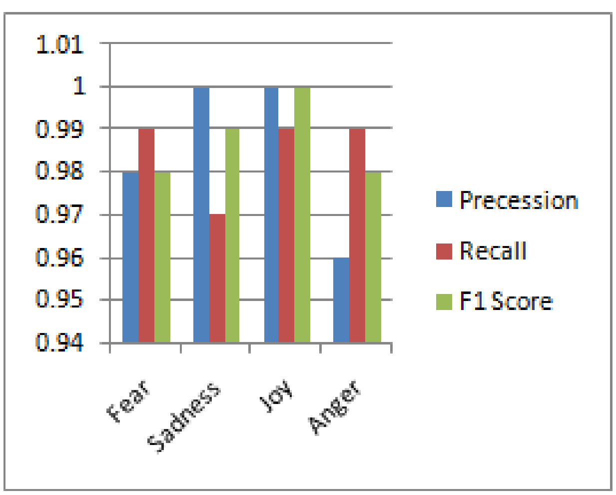

Figure 11. Precision; Recall; F1 Score of Sampled Dataset

TABLE IX. Performance Evolution of Proposed Model with Sampled Dataset

Emotion	Pre	Rec	F1
Fear	0.98	0.99	0.98
Sadness	1.00	0.97	0.99
Joy	1.00	0.99	1.00
Anger	0.96	0.99	0.98

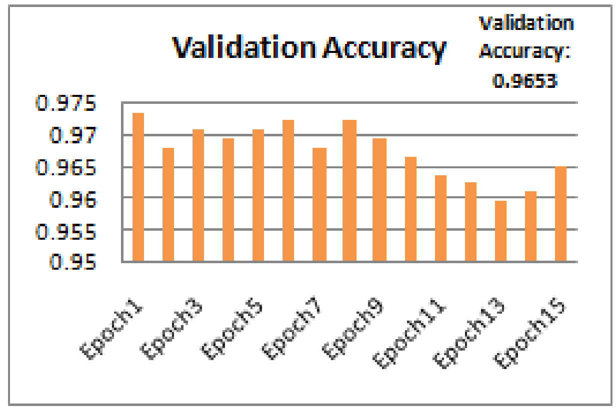

Figure 13. Validation Accuracy of Sampled Dataset

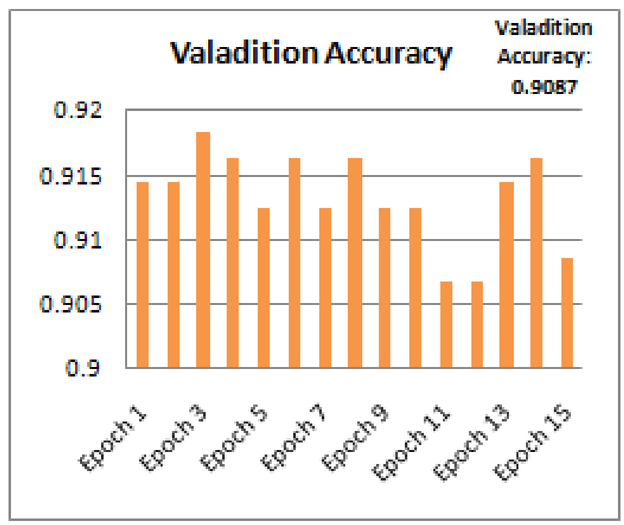

Figure 12. Validation Accuracy of Actual Dataset

Table X demonstrates the accuracy comparison of the proposed AT-LSTM –Bi LSTM model with a actual dataset and the Balance dataset (after addressing the imbalance issue of the class label). The proposed model performs better in predicting emotion labels by both the dataset but when the dataset class label is balanced then the model accuracy is high in comparison to the actual dataset. The accuracy achieved while training with the actual dataset is 90.87% and the accuracy increased to 96.53% when the model is trained with the balanced dataset.

TABLE X. Accuracy Comparison of Proposed Model with Other Experimented Similar Model using Same Dataset

Model	Dataset	Emb.Tech.	Accuracy
AT-LSTM-BiLSTM	Actual Dataset	GloVE	**90.87** %
AT-LSTM-BiLSTM	Sampled Dataset	GloVE	**96.53** %
AT-Nest-BiLSTM	Actual Dataset	GloVE	86.55%
AT-Nest-BiLSTM	Sampled Dataset	GloVE	92.29%
AT-BiLSTM	Actual Dataset	Keras	83.20%
AT-BiLSTM	Sampled Dataset	Keras	91.40%

TABLE XI. Proposed Model Accuracy Compared with other State-of-the-art Models

Study	Feature Representation and Classification	Dataset	Accuracy
Shelke et al.(2022)[12]	LRA-DNN model	Social media	94.77%
Bhuvaneshwari et al. (2022)[13]	Bi-LSTM with CNN	Product Review	89%
Umer et al.(2021)[25]	CNN-LSTM	Twitter	82%
Dang et al.(2020)[26]	CNN	Twitter	82.36%
Mozhdehi et al.(2023)[27]	BERT	Twitter	86%
Proposed Model	LSTM, Bi-LSTM in 2L with AT	Twitter	96.53%

Table XI demonstrates the noticeable accuracy of 96.53% achieved from the proposed hybrid AT-LSTM-BLSTM architecture on Twitter Data for predicting Anger, Fear, Joy, and Sadness emotions. The overall comparative outcome shows that the model presented performs better in comparison to another state-of-the-art framework.

5. Conclusion

The architecture of the classifier and feature extraction play crucial roles in the text classification problem. In many text classification and real-life issues, the LSTM network has shown promising performance, still, it is difficult to design the best architecture to improve the accuracy. To address the existing problem, the hybrid emotion analysis framework's design architecture, development procedure, and evaluation are presented. The model was trained with two forms of the same dataset, by fixing the class imbalance problem of the dataset (Balance dataset) and using the actual (as-is) dataset. The experiment results show that AT-LSTM-Bi LSTM can acknowledge more semantics and enhance the performance than other LSTM-based architecture by obtaining an accuracy of 90.87%. In our analysis of the Twitter dataset after addressing the class imbalance issue, the model accuracy has increased from 90.87% to 96.53%. We conclude that the emotion analysis framework achieved the highest accuracy when the model was trained with a balanced dataset. The precision of the model has increased by 5.66% compared to the actual dataset. In future work, we aim to extend this emotion prediction model to predict emotion in a multi-domain dataset, which includes text data, emotion-based image data and audio-based data.

References

[1] A. O. R. Rodríguez, M. A. Riaño, P. A. G. García, C. E. M. Marín, R. G. Crespo, and X. Wu, "Emotional characterization of children through a learning environment using learning analytics and ar-sandbox," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 5353–5367, 2020.

[2] M. S. Hossain and G. Muhammad, "Emotion recognition using deep learning approach from audio–visual emotional big data," *Information Fusion*, vol. 49, pp. 69–78, 2019.

[3] T. Chen, S. Ju, X. Yuan, M. Elhoseny, F. Ren, M. Fan, and Z. Chen, "Emotion recognition using empirical mode decomposition and approximation entropy," *Computers & Electrical Engineering*, vol. 72, pp. 383–392, 2018.

[4] K. Shrivastava, S. Kumar, and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," *Multimedia tools and applications*, vol. 78, pp. 29 607–29 639, 2019.

[5] S. Shi, M. Zhao, J. Guan, Y. Li, and H. Huang, "A hierarchical lstm model with multiple features for sentiment analysis of sina weibo texts," in *2017 International Conference on Asian Language Processing (IALP)*. IEEE, 2017, pp. 379–382.

[6] J. Zhou, Y. Lu, H.-N. Dai, H. Wang, and H. Xiao, "Sentiment analysis of chinese microblog based on stacked bidirectional lstm," *IEEE Access*, vol. 7, pp. 38 856–38 866, 2019.

[7] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big data: Deep learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, pp. 1–25, 2018.

[8] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "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, pp. 2163–2175, 2019.

[9] J. Wang and Z. Cao, "Chinese text sentiment analysis using lstm network based on l2 and nadam," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*. IEEE, 2017, pp. 1891–1895.

[10] A. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning cnn–lstm framework for arabic sentiment analysis using textual information shared in social networks," *Social Network Analysis and Mining*, vol. 10, pp. 1–13, 2020.

[11] X. Wang, W. Jiang, and Z. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, 2016, pp. 2428–2437.

[12] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare, G. Yogapriya, and P. Pandey, "An efficient way of text-based emotion analysis from social media using lra-dnn," *Neuroscience Informatics*, p. 100048, 2022.

[13] P. Bhuvaneshwari, A. N. Rao, Y. H. Robinson, and M. Thippeswamy, "Sentiment analysis for user reviews using bi-lstm self-attention based cnn model," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 12 405–12 419, 2022.

[14] S. Rani, A. K. Bashir, A. Alhudhaif, D. Koundal, E. S. Gunduz *et al.*, "An efficient cnn-lstm model for sentiment detection in# blacklivesmatter," *Expert Systems with Applications*, vol. 193, p. 116256, 2022.

[15] R. Laza, R. Pavón, M. Reboiro-Jato, and F. Fdez-Riverola, "Evaluating the effect of unbalanced data in biomedical document classification," *Journal of integrative bioinformatics*, vol. 8, no. 3, pp. 105–117, 2011.

[16] H. Saif, M. Fernández, Y. He, and H. Alani, "On stopwords, filtering and data sparsity for sentiment analysis of twitter," 2014.

[17] B. Pahwa, S. Taruna, and N. Kasliwal, "Sentiment analysis-strategy for text pre-processing," *International Journal of Computer Applications*, vol. 180, no. 34, pp. 15–18, 2018.

[18] M. K. Dalal and M. A. Zaveri, "Automatic text classification: a technical review," *International Journal of Computer Applications*, vol. 28, no. 2, pp. 37–40, 2011.

[19] V. Gupta, G. S. Lehal *et al.*, "A survey of text mining techniques and applications," *Journal of emerging technologies in web intelligence*, vol. 1, no. 1, pp. 60–76, 2009.

[20] J. Plisson, N. Lavrac, D. Mladenic *et al.*, "A rule based approach to word lemmatization," in *Proceedings of IS*, vol. 3, 2004, pp. 83–86.

[21] G. Sampson, *The'Language Instinct'Debate: Revised Edition*. A&C Black, 2005.

[22] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] S. M. Mohammad and F. Bravo-Marquez, "Wassa-2017 shared task on emotion intensity," *arXiv preprint arXiv:1708.03700*, 2017.

[25] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model," *Computational Intelligence*, vol. 37, no. 1, pp. 409–434, 2021.

[26] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, p. 483, 2020.

[27] M. Hadikhah Mozhdehi and A. Eftekhari Moghadam, "Textual emotion detection utilizing a transfer learning approach," *The Journal of Supercomputing*, pp. 1–15, 2023.

Niraj Kumar Research Scholar, Dept. of Computer Science and Engineering, Central University of Jharkhand, Ranchi, India-835222

Prof. Subash Chandra Yadav Dept. of Computer Science and Engineering, Central University of Jharkhand, Ranchi, India-835222