WILEY | Hindawi

Research Article

A Deep Neural Network Model for the Detection and Classification of Emotions from Textual Content

Muhammad Zubair Asghar [ID],[1,2] **Adidah Lajis** [ID],[1] **Muhammad Mansoor Alam** [ID],[1,3] **Mohd Khairil Rahmat** [ID],[1] **Haidawati Mohamad Nasir** [ID],[1] **Hussain Ahmad,**[2] **Mabrook S. Al-Rakhami** [ID],[4] **Atif Al-Amri,**[4,5] **and Fahad R. Albogamy**[6]

[1]*Center for Research & Innovation, CoRI, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia*
[2]*Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan, Pakistan*
[3]*Faculty of Computing, Riphah International University, Islamabad, Pakistan*
[4]*Research Chair of Pervasive and Mobile Computing, Information Systems Department,*
 College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[5]*Software Engineering Department, College of Computer and Information Sciences, King Saud University,*
 Riyadh 11543, Saudi Arabia
[6]*Computer Sciences Program, Turabah University College, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Mabrook S. Al-Rakhami; malrakhami@ksu.edu.sa

Emotion-based sentimental analysis has recently received a lot of interest, with an emphasis on automated identification of user behavior, such as emotional expressions, based on online social media texts. However, the majority of the prior attempts are based on traditional procedures that are insufficient to provide promising outcomes. In this study, we categorize emotional sentiments by recognizing them in the text. For that purpose, we present a deep learning model, bidirectional long-term short-term memory (BiLSMT), for emotion recognition that takes into account five main emotions (Joy, Sadness, Fear, Shame, Guilt). We use our experimental assessments on the emotion dataset to accomplish the emotion categorization job. The datasets were evaluated and the findings revealed that, when compared to state-of-the-art methodologies, the proposed model can successfully categorize user emotions into several classifications. Finally, we assess the efficacy of our strategy using statistical analysis. This research's findings help firms to apply best practices in the selection, management, and optimization of policies, services, and product information.

1. Introduction

For a long time, computer-based emotion detection and classification has been a hot topic of research. Emotion detection can be done using a variety of mediums, including text, photographs, video, and audio [1, 2]. In recent years, social media sites such as Twitter, Facebook, and Instagram have undergone an unexpected global expansion. Twitter, for example, had over 200 million monthly active users by the fourth quarter of 2021 [3]. As long as analytical researchers working on social networking material can address the particular obstacles presented by such content, advances in computational linguistics and text analytics allow researchers to extract and evaluate textual emotions provided by users on social media via big data sources.

Computational linguistic experts have performed many studies to detect and identify emotions at various levels, including words, expressions, sentences, and analysis [4–7]. Many studies, on the other hand, focus on emotion-related bearing terms, with little attention paid to textual clues to emotions, which, if included, may improve the output of cognitive-based sentiment classification for social media data. Hence, the research and development of text-based emotion classification systems are motivated by studies of emotions expressed in text.

The proposed emotion-based sentiment analysis system for social media is based on previous work on emotion classification in the social media paradigm [5, 8]. Previous studies lacked the capability to represent emotion signals using advanced feature representations and do not exploit the capability of individual deep learning models for emotion classes. However, we present a social media-based cognitive-based sentiment analysis framework that focuses on detecting and classifying textual emotions using an advanced feature representation scheme and a fusion of deep learning models.

1.1. Research Motivation.

Over the last few years, emotion-based SA applications have become increasingly popular on the Internet for gauging the opinions, emotions, and sentiments of individuals on different issues and policies. However, it is often difficult to analyze text with existing emotion detection methods to detect emotions from social media content. Therefore, it is essential to extract and analyze social media content to automatically classify emotions.

The difference between social media and traditional blogs is that the former incorporates complex textual material. When opposed to text information alone, such material contains text and emotion cues, which are more suitable for expressing and conveying people's subtle thoughts, emotions, and personal characteristics [1]. However, emotion-based sentiment analysis, which is based on the identification of emotional cues, is still in its early stages.

There has been extensive work carried out in the area of text-based analysis of feeling [8], construction of lexicons [4], cognition and analysis of aspects of feeling [6], and analysis of visual feelings [1, 9]. However, further research is required in the area of cognitive-based social media analysis, with a focus on extracting and categorizing emotions from social media content.

1.2. Problem Statement.

In this work, we address the problem of emotion classification of English text using a deep learning technique. Considering a set of reviews = {$R1$, $R2$, $R3$, …, Rn} given as input, the aim is to develop a classifier for assigning an emotion label $Ei \epsilon$ {J-S, F-G, F-S} for the review ri, where J-S, F-G, and F-S show different emotion labels. It is our goal in this study to develop a powerful deep learning-based emotion recognition system that is capable of accurately classifying provided text reviews into the required emotional category.

1.3. Research Questions.

RQ1: How can textual content be identified and classified into different emotion groups using BILSTM, a deep learning technique?

RQ2: What is the proposed technique's efficiency in comparison with other machine learning and deep learning techniques?

RQ3: How can the efficacy of the proposed technique for emotion classification be estimated in comparison to baseline studies?

1.4. Proposed Contributions.

The following are the main contributions of the research work:

(1) To efficiently capture the semantics of words, an advanced function representation scheme called word embedding is used.

(2) For the emotion classification challenge, we present a deep neural network called Bi-LSTM (Bi-LSTM), which learns contextual information in both directions, forward and backward, using forward and backward LSTM.

(3) We tested different ML classifiers as well as deep learning models.

(4) In comparison to state-of-the-art techniques, our proposed technique yielded productive performance.

The goal is to improve the scientific literature by detecting emotional indicators in textual material on social networking sites. The proposed system will help companies identify and analyze their consumers' attitudes and emotions toward a program, policy, or product, which can help them make better decisions.

The following is how the rest of the article is organized: (i) Section 2 discusses the current methods and frameworks related to emotion-based sentiment analysis, (ii) Section 3 presents the proposed framework for emotion-based sentiment analysis, (iii) experimental setup is introduced in this Section 4, and (iv) the proposed system is concluded with its drawbacks and potential future guidance in Sections 5 and 6.

2. Related Work

Due to the ambiguity of emotions and a large number of emotional terms, detecting and classifying emotions in sentiment analysis is a challenging activity. There have been several techniques and methods proposed for detecting and classifying emotions from social media textual content. This segment gives a summary of the current state of emotion detection and classification science.

The efficacy of different machine learning models is evaluated by [8]. Variant machine learning models such as SVM, NB, DT, LR, XGBoost, KNN, and backpropagation neural (BPN) classifier are tested using the state-of-the-art ISEAR emotion dataset to achieve this goal. According to the results, BPN had the highest accuracy of all the classifiers, with a score of 71.27 percent. The work has a few limitations: (i) it is limited to five emotion classes, (ii) it uses only the ISEAR dataset, (iii) conventional features are used, and (iv) classical machine learning models are used. Future directions may include utilizing various emotion combinations, (ii) utilizing other benchmark datasets, (iii) conducting additional deep learning experiments, and (iv) utilizing a word embedding feature representation approach. The authors of [10] are tasked with categorizing emotions in relation to Indonesian texts. The study used a variety of machine learning algorithms, including NB, SVM, KNN, and minimal optimization, to achieve this goal. Various text cleaning tasks (tokenization, stop word deletion, and case

conversion) are also introduced. During the tests, 10-fold cross-validation was used, and the findings show that the minimal optimization technology outperformed other techniques. For the extraction of variant tweets, we use emotion-word hashtags as well as the "Hashtag Emotion Corpus" data collection [11]. Furthermore, the study used an emotion-labeled tweet dataset to create an emotion dictionary of rich terms. Experiments show that the output of the SVM model is most closely linked to the primary emotion classes. The inclusion of emotion words with variant synonyms, on the other hand, can aid in improving the system's effectiveness. The identification of emotions in the human speech was performed by [12] in their research. Different speech recognition models are used, which are then followed by various speech features (peak to peak distance). In comparison to alternative methods, the dataset used in the experiments contains 30 distinct subjects, and the highest precision is achieved. An et al. [13] tackled the issue of music emotion classification based on lyrics in their research. They accomplish this by crawling the music lyrics with their tags on Baidu, a well-known network platform. Following that, a naive Bayes classifier that has been trained on four different datasets is used. According to the findings, the final classification accuracy on D-4 was 68 percent (Dataset). Using other algorithms, on the other hand, could improve the proposed system's effectiveness. Dini and Bittar [14] conducted research to classify emotions on Twitter. Two corpora have been created for this purpose: ETCC for emotion classification of tweets and ETCR for emotion relevance of tweets. This dataset is used to train and evaluate machine learning models, as well as to test a rule-based model. The results show that the symbolic technique outperformed the ML algorithm in determining tweet significance, but the ML algorithm is the best option for tweet emotion classification. The goal in the future is to create a hybrid model that combines both approaches, and marked data quality testing will be discussed as well. Kaewyong et al. [15] suggested a lexicon-based approach in their work on automated feedback inquiry for student feedback. To begin, data was gathered from over 1100 student responses about teaching faculty. Following that, preprocessing techniques are used, followed by using a sentiment lexicon to assign sentiment scores to opinion terms. In comparison with the other approaches, the findings show that the proposed solution has the best efficiency. Sen et al. [16] set out to determine a novel method for multitask learning of input word embedding by using supervised prompts for emotions and sentiment tasks. After that, they looked at using jointly qualified embedding to improve emotion detection. With a test accuracy of 57.46 percent, the findings show that KMeans (ES-SWE) outperformed the other approaches. In the future, it is expected to perform a configuration experiment in which word embedding is mutually trained, sensitive to emotion-topic, and used for emotion or topic downline recognition. Furthermore, their proposed approach can be used to solve other classification problems besides emotion and sentiment identification. The work can also be applied to n-gram embedding. Kollias et al. [17] investigated deep convolutional neural networks for facial expression and emotion detection (DCNNs). In comparison with the competing models, the experimental findings show that the proposed model outperforms them. However, accurate device estimation can be accomplished by developing a real-world application for human-computer communication. Poria et al. [18] use the convolutional learning approach to tackle the problem of extracting emotions from the media content, specifically text, audio, and video. In comparison to the baseline work, the inner layer of the network used an activation feature and achieved an accuracy of 96.55 percent for the MOUD dataset and 76.85 percent for the IEMOCAP dataset. To conduct microblog sentiment analysis, Severyn and Moschitti [19] proposed a deep learning-based Convolutional Neural Network (CNN). The input to his proposed model is seed words that have been correctly trained using a deep learning model. The key advantage of the framework is that it does not need support attributes to train the model on Twitter data records. The proposed model achieved the best results at the sentence and phrase stages, according to the results of the experiments. Gupta et al. [20] present a new profound learning approach to emotion identification in textual conversations, namely, sentiment and semanticized LSTM (SS-LSTM). Two datasets, ISEAR and SemEval2007 Affective Text, were used in the experiments. The experimental results show that the proposed method outperformed state-of-the-art machine learning as well as other deep learning approaches, with an average F1-score of 71.34 percent. However, the technique can be improved by using a context-sensitive framework to train a model. Cambria [21] uses effective computer methods to investigate the problem of the identification of feelings and emotional predictions. Text classification is carried out in the positive and negative sense module, and emotional clues are identified in the emotional prediction system. First, feelings are found in the proposed approach, and then a certain type of emotion is assigned to the unwanted feeling. To define and classify emotions from textual material, a number of research works are categorized as emotions, as seen in the literature review above.

2.1. Research Gap. To resolve the limitations of the aforementioned recent emotion-based sentiment classification techniques, it is necessary to detect and identify emotions from textual material. To close this void, we propose a comprehensive emotion-based sentiment system for emotion classification expressed in online social media. As a result, for emotion classification, we used automated feature engineering techniques followed by a deep neural network called Bi-LSTM.

3. Proposed Methodology

The current section presents the proposed method for emotion classification's comprehensive architecture.

3.1. Acquiring Data. We have obtained a benchmark data set called "ISEAR" [8] containing 5474 records. Each input review is tagged with a separate class of three emotion class

sets such as J-S, F-S, and F-G. This work employs the Python language [1], and the library used is Keras, which is based on the TensorFlow deep learning framework [1]. Figure 1 provides a description of the dataset.

3.1.1. Training Set.
The model has been trained using the training dataset, with 80 percent of the data used for model training [1]. Figure 2 shows an exemplary review of training examples.

3.1.2. Validation Set.
In the training phase, the model is typically accurate, but in the testing phase, the model's performance declines. As a result, in order to overcome the model's performance error in terms of underfitting and overfitting, the validation set must be used [22]. Keras has two methods for determining the optimum model parameters: manual data validation and automatic data validation [3]. We are currently employing manual data validation in our current project.

3.1.3. Testing Set.
Based on new/unseen cases, the testing set is used to evaluate the performance of the model. It is used once the model has been trained properly using both the train and validation sets. The test set is responsible for the model's ultimate prediction [23]. Table 1 contains test set samples for emotions classification.

3.2. Main Modules of the Proposed System.
The proposed method consists of three main modules: I Embedding Layer-based Word Representation, (ii) Bi-LSTM-based Forward and Backward Context Information Saving, and (iii) Sigmoid Layer-based Classification. The first module's goal is to obtain a numeric representation of the terms, which will be fed into the second module, which will produce an encoded representation of features. Bi-LSTM is used to create this encoded representation, which keeps track of both the forward and backward contextual details of the word within a series. Finally, a sigmoid activation mechanism is used to perform classification in the final module (see Figure 3). The following is a breakdown of each module:

3.2.1. Words Representation Exploiting Embedding Layer.
The emotion dataset is represented as a set of more than one user review and an individual review E "*I felt very happy when I won the soccer pool*" involving a series of r words, i.e., w1, w2, w3, and wr. A single term wi "felt" represents an embedding vector wi-Rn that includes real values [0.6, 0.9, 0.2]. An embedding matrix is made up of the embedding vectors of each term. During this project, the kera embedding layer was used. The embedding matrix is a two-dimensional matrix denoted by $D R (rn)$, where r denotes the length of the input review and n denotes the embedding dimension. After that, the next layer receives an embedding matrix D, also known as a sentence/input matrix.

Task	Dataset with title	Number of Review Text	Labels
Emotion Classification	ISEAR	5474	Joy (1094), Fear (1095), Sadness (1096), Shame (1096), Guilt (1093)

FIGURE 1: The dataset's description.

Review Text ID	Review Text	Label
1.	When I was involved in a traffic accident.	F-G
2.	When I lost the person who meant the most to me.	J-S
3.	When I did not speak the truth.	F-S
4.	When I caused problems for somebody because he could not keep the appointed time and this led to various consequences.	F-G
6.	When I got a letter offering me the Summer job that I had applied for.	J-S
7.	When I was going home alone one night in Paris and a man came up behind me and asked me if I was not afraid to be out alone so late at night.	F-G
9.	When my friends did not ask me to go to a New Year's party with them.	J-S
10.	When my uncle and my neighbour came home under police escort.	F-G

FIGURE 2: Train set examples for emotion classification.

3.2.2. Bi-LSTM Layer.
The Bi-LSTM layer is responsible for learning long dependencies. It aids in the saving of the two preceding and succeeding contexts in the form of an encoded user analysis. A unidirectional LSTM, on the other hand, saves only the information from the previous context, leaving out the information from the subsequent context. As a result, Bi-LSTM gathers even more information for encoded review processing. To learn past and future meaning knowledge of tokens (words), Bi-LSTM employs forward and backward LSTM [2]. *Forward LSTM.* During forward LSTM, the processing of the sequence is done from the left toward right through the concatenation of binary inputs. The first one is present (current) input "x_1," while the second one is prior input/hidden_state "h_{t-1}." Forward LSTM generates a certain outcome "h" regarding a provided sequence of input: $x_1, x_2, x_3, \ldots, x_{y-1}$. *Backward LSTM.* For backward LSTM, the processing of the sequence is done right toward the left through the concatenation of two inputs, the first one is the current input "x_1," and the second one is the next input/future hidden state "h_{t+1}." Forward LSTM generates a certain outcome "h" regarding a provided sequence of input: $x_{z+1}, \ldots, x_3, \underset{\leftarrow}{x_2}, x_1$.

The integration of forward $\overrightarrow{h}$ and backward $\overleftarrow{h}$ context representations is done to generate different sentence matrices $H = [h_1, h_2, h_3, \ldots, h_y]$, while $H \varepsilon \mathbb{R}^{y \times m}$. The integration of forward and backward output is conducted through the computation of an element-wise sum (see (1)).

TABLE 1: Test set examples for emotion classification.

Review text ID	Review text	Label
5.	When I realise that I value material possessions more than the well-being of my family, I feel a pang of guilt wash over me. I'm feeling quite self-obsessed at the moment.	F-G
8.	When my child was born.	J-S
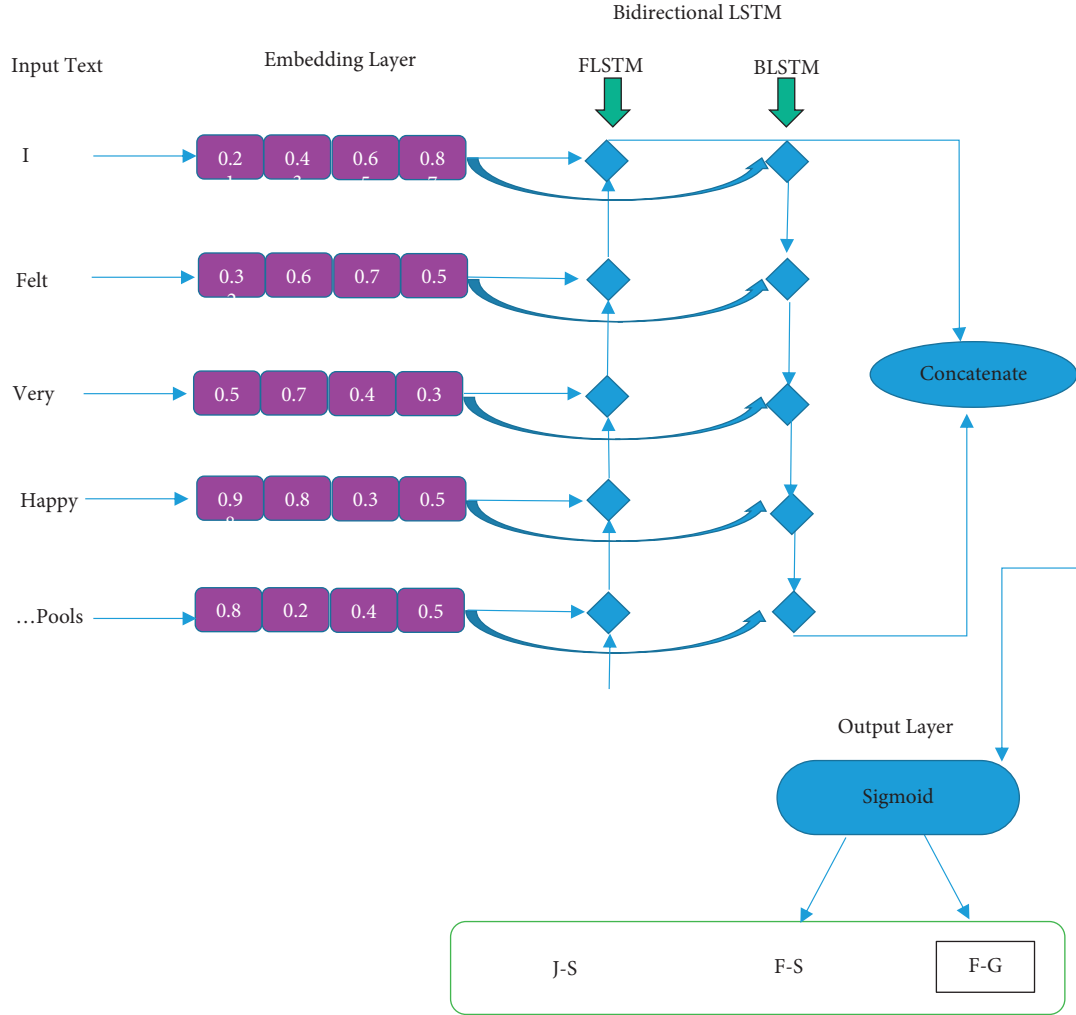
FIGURE 3: The Bi-LSTM model architecture concerning emotion classification.

$$h = \overrightarrow{h} \oplus \overleftarrow{h}, \tag{1}$$

Lastly, the new representation of a sentence matrix (1) is fed into a classification layer regarding final classification work.

The equations ((2)–(7)) used for forward LSTM [24] are given as follows:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right), \tag{2}$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right), \tag{3}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right), \tag{4}$$

$$c\sim_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right), \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c\sim_t, \tag{6}$$

$$h_t = o_t \odot \tanh\left(c_t\right). \tag{7}$$

The equations ((8)–(13)) used for backward LSTM [24] are given as follows:

$$f_t = \sigma\left(W_f x_t + U_f h_{t+1} + b_f\right), \tag{8}$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t+1} + b_i\right), \tag{9}$$

$$o_t = \sigma(W_o x_t + U_o h_{t+1} + b_o), \tag{10}$$

$$c\sim_t = \tanh(W_c x_t + U_c h_{t+1} + b_c), \tag{11}$$

$$c_t = f_t \odot c_{t+1} + i_t \odot c\sim_t, \tag{12}$$

$$h_t = o_t \odot \tanh(c_t), \tag{13}$$

where n represents the size of the input, m represents the size of the cell state, x_t depicts the vector of input with $n \times 1$ size, f_t depicts the vector of forget gate with $m \times 1$ size, i_t depicts the vector of input gate with $m \times 1$ size, o_t depicts the vector of output gate with $m \times 1$ size, h_t depicts the vector of output with $m \times 1$ size, and c_t depicts a vector of cell state with $m \times 1$ size.

The output gate weight matrices are represented as W_c, W_o, W_i, and W_f with size $m \times n$. Weight matrices of the output gate are represented as U_c, U_o, U_i, and U_f with size $m \times m$. The bias vector is represented by b_c, b_o, b_i, and b_f with size $m \times 1$. The tangent function is denoted by tanh, and the logistics sigmoid function is denoted by σ.

Every gate in Bi-LSTM performs its own function. The function of the forget gate f_t is to delete useless information. The input gate i_t takes a decision about storing which information. Lastly, the output gate o_t calculates the final output h_t.

The notations used during forward and backward LSTM are listed in Table 2.

3.2.3. Feature Classification Using Sigmoid Layer.

This layer performs the classification of input features (final representation) obtained from the previous module. We add a dense layer with two neurons that have a sigmoid function for this purpose. The sigmoid activation function performed a nonlinear operation, and its task was to calculate the probability of various emotion classes. It converts the weighted sum into a number between 0 and 1. Therefore, after the output layer passes the review text "I felt very happy when I won the football pool," it is tagged with one of the three binary classes "J-S," "F-S," or "F-G."

The probability of each of the emotion classes is calculated using a softmax activation function. The net input for classifying the final emotion representation (equation (14)) can be approximated as follows:

$$\mathbf{uj} = \sum_{i}^{l} \mathbf{wixi} + \mathbf{b}, \tag{14}$$

where "$\mathbf{w}$" denotes a weight vector, "$\mathbf{x}$" denotes a vector of inputs, and "b" denotes a bias factor.

The phases of the Bi-LSTM system for emotion categorization are depicted in Algorithm 1.

4. Results and Discussion

The experiments in this work are implemented in Python using the Keras library (All things Keras, n.d.) which is based on the TensorFlow deep learning framework [25]. The machines utilized for testing are Intel Core i7 with a 64 bit OS and 8 GB internal storage. The original dataset is divided into three sets: training, testing, and validation.

4.1. Answer to RQ1.

To answer RQ1, "*How can textual content be identified and classified into different emotion groups using BILSTM, a deep learning technique?*," Different experiments are carried out with respect to different emotion cues such as Joy (J), Sadness (S), Fear (F), Shame (S), and Guilt (S) for the output assessment of the proposed deep learning model (G). We created three binary classifiers for this experiment: Joy-Sadness, Fear-Shame, and Fear-Guilt.

4.1.1. Experiment #1: Joy-Sadness.

Experiment #1 was performed to determine the efficiency of the proposed BILSTM model with respect to Joy-Sadness emotion signals, as shown in Table 3. According to the experimental findings, the BILSTM model performed better for the emotion clue "joy" with an F1-score of 0.89 and recall of 0.91, while both emotion clues "joy" and "sadness" performed best in terms of precision (0.88). The overall precision is 0.88 percent.

4.1.2. Experiment #2: Fear-Shame.

Experiment #2 was performed to determine the efficacy of the proposed BILSTM model regarding Fear-Shame emotion clues, as shown in Table 4. The BILSTM model had the best performance for the "Shame" emotion hint, with precision (0.89), recall (0.91), F1-score (0.89), and overall accuracy of 0.86 percent, according to the results.

4.1.3. Experiment #3: Fear-Guilt.

Experiment #3 was performed to determine the efficacy of the proposed BILSTM model regarding Fear-Guilt emotion signals, and the findings are summarized in Table 5. According to the findings, the BILSTM model performed well for both "Fear" and "Shame" emotion clues, with accuracy (0.89), recall (0.89), and F1-score (0.89). The overall precision is 0.89 percent.

4.2. Answer to RQ2.

To find the answer to RQ2, "*What is the efficiency of the proposed technique in relation to other machine learning and deep learning techniques?*," we conducted experiments to compare the efficiency of a word embedding scheme trained using the BILSTM method to machine learning classifiers that use traditional feature representation schemes such as the TF-IDF and Count-Vectorizer. The assessment results are presented in Tables 6 and 7.

4.2.1. BILSTM and Machine Learning Approaches Attempting to Exploit Traditional Features in Comparative Study.

The proposed BILSTM model's output is compared with that of various machine learning approaches using classical feature representation schemes, as shown by the following experiments:

TABLE 2: Notations used during forward and backward LSTM.

Notation	Depiction
W_c, W_o, W_i, W_f	Weight matrices of the input gate
U_c, U_o, U_i, U_f	Weight matrices of the output gate
b_c, b_o, b_i, b_f	Biases
h_{t-1}	Prior cell output
h_t	Hidden state
h	The output obtained from element-wise sum regarding prior cell output and subsequent cell output
h_{t+1}	subsequent(future) cell output
c_t	Current state cell memory
i_t	Gate of input
f_t	Gate of forget
o_t	Gate of output
x_t	Current input
σ	Sigmoid operation
tanh	Hyperbolic tangent function
H	Sentence matrix obtained through Bi-LSTM

Data: Emotion dataset "M", Train Set "TAS", Test Set "TSS"
Result: Review Text Label: "J-S", "F-S", "F-G"
Start
// Review Text Encoding toward machine understandable word vectors (real valued)
while each review text $R \ \varepsilon \ M$ **do**
   while each word $T \ \varepsilon \ M$ **do**
(1)   Word(token) indices allocation
   End while
End while
Initializing Hyperparameter
(2) embed_dim = 100, 128,300, max_features = 2000, epochs = 7, batch_size = 32, train set = 90%, test size = 10%
//Deep Learning model training
while each review text $R \ \varepsilon \ M_{TAS}$ **do**
(3)   Generate all word embedding vectors in $R = [r_1, r_2, r_3, \ldots, r_n]$
(4)   Implement Bi-LSTM operation exploiting equations (1)–(13)
End while
// Allocating a label to Review Text final depiction
while each Review Text $R \ \varepsilon \ M_{TSS}$ **do**
(5)   Trained(learned) model is built
(6)   Employ a softmax classifier using Eq. 14, for the classification of output obtained from the Bi-LSTM into "J-S", "F-S", "F-G"
End while
End

ALGORITHM 1: Emotion classification through Bi-LSTM neural network.

TABLE 3: Proposed BILSTM model performance evaluation results.

Emotion cues	F1-score	Recall	Precision	Accuracy (%)
Joy	0.89	0.91	0.88	0.88
Sadness	0.86	0.83	0.88	

TABLE 4: Proposed BILSTM model performance evaluation results.

Emotion cues	F1-score	Recall	Precision	Accuracy (%)
Fear	0.84	0.85	0.83	0.86
Shame	0.88	0.86	0.89	

TABLE 5: Proposed BILSTM model performance evaluation results.

Emotion cues	F1-score	Recall	Precision	Accuracy (%)
Fear	0.89	0.89	0.89	0.89
Guilt	0.89	0.89	0.89	

TABLE 6: Comparative results of the proposed method with the baseline studies for emotion classification.

	Models	Emotion classes	Metrics results			
			Avg. acc (%)	Avg. pre (%)	Avg. rec (%)	Avg. F1-score (%)
Baseline Study: Neural Network Technique	Back propagation neural Classifier (BPN) [8]	Joy (J), Sadness (S), Fear (F), Shame (S), Guilt (G)	71.27	0.67	0.66	0.67
Proposed Study: Deep Learning Technique	BISLTM	Joy (J), Sadness (S), Fear (F), Shame (S), Guilt (G)	87.66	87.66	87.66	87.66

TABLE 7: J-S emotion class significance.

	Proposed BILSTM correct classification	Proposed BILSTM misclassification	Total
KNN correct classification	165	11	176
KNN misclassification	26	17	43
Total	191	28	219

When we ran a McNemar-type of the test, we found that the chi-squared statistic was 5.5.3, and a 1 represents the degree of freedom, which means that the 2-tailed p value is 0.021. Consequently, the null hypothesis is deemed to be false, and an alternative hypothesis is adopted.

ML Driven. We estimate the output of different classical feature representation strategies for machine learning classifiers, such as Countvectorizer and TF-IDF, in these experiments. The method used in Countvectorizer is called count-of-words, and the text is converted into a feature vector using the TF-IDF scheme. Many classifiers are employed by Countvectorizer, TF, and TF x IDF feature encoding techniques, with KNN achieving the best accuracy score of 80.82 percent and XGBoost achieving the worst accuracy score of 71.23 percent.

Proposed BILSTM Model. We used the BISLTM model over a word embedding feature representation scheme to perform an emotion classification task in this experiment. The main advantage of using a word embedding over the traditional BOW system is that the BOW model's output degraded as the vocabulary grew larger, while deep neural network models yielded more successful results.

Table 8 shows that the BILSTM (proposed method) exploits a word embedding scheme to generate more efficient findings than traditional feature representation methods such as Countvectorizer and TF-IDF.

4.2.2. Comparison of Proposed BILSTM with Variants of Deep Learning Models.
Different studies are carried out to assess the efficacy of the proposed BILSTM model in comparison to various DL models.

(1) DL Driven. Individual CNN, Individual LSTM, Individual BILSTM, and Individual RNN with an advanced feature representation scheme were used to test various deep learning models (word embedding). The Individual GRU model achieves the highest accuracy of 82 percent among all DL classifiers, while the Individual RNN model achieves the lowest accuracy of 66 percent.

This segment uses two datasets to compare our approach with different deep learning approaches such as Individual BILSTM, Individual LSTM, Individual CNN, and Individual RNN. The results of the experiments in Tables 8 and 9 are summarized as follows.

(2) Comparing Proposed BILSTM with Individual CNN. In the initial experiments for the emotion classification task, the proposed BILSTM model is compared to the individual CNN. In terms of accuracy, precision, recall, and F1-score, the proposed model outperformed the individual CNN model (accuracy = 88 percent, precision = 88 percent, recall = 88 percent, and F1-score = 88 percent). The decrease in CNN performance is due to the fact that the text classification task requires the preservation of sequential data, which a single CNN cannot help with. In addition, a large dataset must be given to the CNN model in order for it to enhance its accuracy.

(3) Comparing Proposed BILSTM with Individual LSTM. In the second experiment, an Individual LSTM is compared to the proposed model in an emotion classification task. The downside of the unidirectional LSTM is that it only retains previous information and not subsequent information. Maintaining information on both sides of words (previous and subsequent) helps in a greater understanding of sentence context. As a result, when compared to the proposed deep neural network model, an Individual LSTM layer will reduce the performance by up to 10% for J-S, 2% for F-G, and 1% for F-S, as shown in Table 8.

(4) Comparing Proposed BILSTM with Individual GRU. For the classification of emotions, a performance comparison between the proposed BILSTM model and an Individual GRU is performed in the third experiment. The main disadvantage of GRU is that it does not perform feature extraction and instead focuses on preserving contextual details. In comparison with the proposed BILSTM model, GRU's output dropped by up to 6% for J-S, 3% for F-G, and 4% for F-S, as shown in Table 8.

(5) Comparing Proposed BILSTM with Individual RNN. A comparison of the proposed model and the RNN model is carried out in the final experiment. The RNN's output is

Table 8: Evaluation results of variant deep learning and the proposed technique.

Classifier/model	Accuracy (%)		Precision (%)		Recall (%)		F1-score (%)	
	J-S F-S	F-G	J-S F-S	F-G	J-S F-S	F-G	J-S F-S	F-G
Individual CNN [26]	74 / 76	75	74 / 76	75	74 / 76	75	74 / 76	75
Individual LSTM [20]	78 / 85	87	78 / 85	87	78 / 85	87	78 / 85	87
Individual GRU	82 / 82	86	82 / 82	86	82 / 82	86	82 / 82	86
Individual RNN	66 / 80	78	74 / 83	83	73 / 83	83	72 / 83	83
Proposed (BILSTM)	88 / 86	89	88 / 86	89	88 / 86	89	88 / 86	89

Table 9: Evaluation results of machine learning and the proposed technique.

Classifier/model	Accuracy (%)		Precision (%)		Recall (%)		F1-score (%)	
	J-S F-S	F-G	J-S F-S	F-G	J-S F-S	F-G	J-S F-S	F-G
SVM [8]	79 / 67.73	81.74	79 / 75	82	79 / 68	82	79 / 82	67
KNN	80.82 / 79.55	79	81 / 79	79	81 / 79	79	81 / 79	79
LR	78.08 / 81.36	82.19	78 / 81	82	78 / 82	82	78 / 81	82
RF	73.52 / 72.27	67.12	73 / 72	67	73 / 72	67	73 / 72	67
MNB [27]	74.89 / 84.09	85.84	77 / 84	86	76 / 84	86	75 / 84	86
DT [27]	76.71 / 76.82	73.52	77 / 76	74	77 / 76	73	77 / 76	73
XGBoost	71.23 / 73.18	72.15	80 / 73	72	71 / 73	72	67 / 73	72
Proposed (BILSTM)	88 / 86	89	88 / 86	89	88 / 86	89	88 / 86	89

degraded because it is unable to manipulate additional broad sequences, demonstrating the RNN's inability to monitor long-range relationships. It is necessary to retain the details for a long time in order to maintain the context, which an Individual RNN cannot do because it just keeps track of short-term memory sequences.

In comparison to the BILSTM, RNN resulted in a performance decline of up to 22% for J-S, 11% for F-G, and 6% for F-S, as shown in Table 8.

4.2.3. Why the Proposed BILSTM Model Is Better?

As we employed "Bidirectional" LSTM (BILSTM) on the supplied datasets, the findings outperform other deep learning approaches. The purpose of the BILSTM model is to retain background data from both sides of an expression, i.e., the left and right sides, within a phrase. After receiving data from the embedding layer, the BILSTM generates an enhanced encoding of the data that takes into consideration both the current and earlier input information. As a consequence, it is obvious that the BILSTM

deep learning model can successfully collect current and past background information through time and produce predictions. The suggested deep learning model performed well on the dataset in categorizing the input text into various emotions such as Pleasure, Sadness, Fear, Guilt, and Shame.

4.3. Answer to RQ.3.

To answer RQ3, *"How to estimate the efficiency of the proposed technique regarding emotion classification with regard to baseline studies?"*, we ran an experiment to see how well the proposed BISLTM model performed in comparison to the baseline sample. The experimental results are mentioned in Table 6.

4.3.1. Proposed (BILSTM) Compared with Baseline (BPN).

In this experiment, we compared the output of the proposed BILSTM approach with that of [8], who used a back-propagation neural classifier to classify emotions. With an accuracy of 87.66 percent, the proposed method outperformed the current state-of-the-art analysis, as shown in

TABLE 10: F-G emotion class significance.

	Proposed BILSTM correct classification	Proposed BILSTM misclassification	Total
MNB correct classification	150	10	160
MNB misclassification	19	40	59
Total	169	50	219

With one degree of freedom, we observed that the 2-tailed p value was 0.137 with a chi-square statistic of 2.22. As a result, the null hypothesis is discarded in favor of an alternative hypothesis.

TABLE 11: F-S emotion class significance.

	Proposed BILSTM correct classification	Proposed BILSTM misclassification	Total
MNB correct classification	145	20	165
MNB misclassification	29	25	54
Total	174	45	219

A chi-square statistic $a = 1.3$ gives us a p value of 0.253 for the two-tailed McNemar's test, and a degree of freedom of 1. An alternate hypothesis is therefore accepted, and the null hypothesis is rejected.

Table 6. The following are the reasons for our model's improved performance: The reason for our improved model output is that we used a "bidirectional" LSTM model that is effective at maintaining both the left and right contextual details of the series. Furthermore, BILSTM is good at storing knowledge over a long period of time. As a result, keeping information for a long time is very useful for text classification and prediction tasks. BPN, on the other hand, is not suitable for classification issues [1]. In addition, the BPN classifier has issues with delay convergence, neural network weights declining over local optima, and network insensibility [28].

4.3.2. Significance Test. We also ran experiments to see whether the proposed BILSTM classifier, which uses the word embedding function, is statistically different from KNN for J-S and MNB for F-G and F-S, to perform a significance test for the "emotion classification task." The findings suggest that our proposed BILSTM model's efficacy in the emotion classification task for J-S, F-G, and F-S was greatly enhanced by innovative features (word encoding), as seen in Tables 7, 10, and 11.

5. Conclusion

We further analyzed why and how users react in different emotional states. To carry out the research task, we proposed applying a Bi-LSTM technique. The study will include the following modules: (i) take the data, (ii) prepare the data, and (iii) apply the deep learning algorithm. For the emotion classifier task, we used a deep learning model, namely, Bi-LSTM. Bi-LSTM performs two tasks at once; i.e., it can remember both the forward and backward sequence of the previous time sequence [8]. After using numerous other programs to encode the text, the decoded text is manually classified as J, F, and G. Experiments to apply different machine learning and deep learning algorithms to emotion datasets were also conducted. The results show that the proposed Bi-LSTM model produced improved results in terms of improved accuracy (87.66%), precision (87.66%),

recall (87.66%), and F1-score (87.66%) with respect to the compared studies.

The following are some potential constraints on the proposed work:

(1) We perform the emotional classification of text content

(2) The research is limited to random word embedding, with no use of word representation models such as Glove, FastText, or word2vec

(3) Emotions are not classified using other configurations of deep learning techniques

(4) The current study's content is solely focused on the ISEAR and Twitter emotion datasets

(5) In the work, limited emotion clues are exploited

(6) The current research does not mention emotion intensities such as strong negative, strong positive, weak negative, and weak positive, which must be discussed to make the system more effective

(7) The research is limited to texts written in English

(8) AUC, density, and error rate can be used to better estimate the classifier's performance; however, the proposed study topic is confined to performance metrics like accuracy, preciseness, recall, and F1-score

(9) A modest number of machine learning models are currently being used for testing, but that can be extended

6. Future Directions

The following is a list of probable future options for the research work:

(1) Photographs and videos may be used to expand the work

(2) Other previously trained schemes such as Glove, word2vec, and FastText can be used for word embedding layer in future work

(3) For emotional classification, we will investigate the various combinations of deep neural networks

(4) Exploring additional emotional classification tasks in different data sets

(5) One potential plan in the future is to increase the performance of the research project by extending the different emotion clues

(6) The research will be expanded to include other languages to test the efficacy of the proposed model in other languages

(7) The problem of long computation times can be addressed using GPU, which allows for large-scale dataset experiments

(8) Using a combination of deep neural network systems to solve the problem of emotion classification would be more efficient. As a result, we will use other neural networks in the future

(9) In the future, we will put more emphasis on using ensemble approaches to improve device efficiency

(10) We will focus on adding more base models and looking for other criteria that might further improve the overall accuracy of the proposed work

(11) In future studies, we will look at combining our proposed approach with other NLP strategies such as part of speech and marking to achieve better results for NLP problems

Data Availability

Two publically available datasets (ISEAR and SemEval2007) were used to support the findings of this study. The two datasets were cited and included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. Khattak, M. Z. Asghar, M. Ali, and U. Batool, "An efficient deep learning technique for facial emotion recognition," *Multimedia Tools and Applications*, pp. 1–35, 2021.

[2] M. Z. Asghar, F. Subhan, H. Ahmad et al., "Senti-eSystem: a sentiment-based eSystem -using hybridized fuzzy and deep neural network for measuring customer satisfaction," *Software: Practice and Experience*, vol. 51, no. 3, pp. 571–594, 2021.

[3] A. Khattak, A. Khan, H. Ullah et al., "An efficient supervised machine learning technique for forecasting stock market trends," in *Information and Knowledge in Internet of Things*, pp. 143–162, Springer, Cham, New York, NY, USA, 2022.

[4] Y. Wang, F. Subhan, S. Shamshirband, M. Z. Asghar, I. Ullah, and A. Habib, "Fuzzy-based sentiment analysis system for analyzing student feedback and satisfaction," *Computers, Materials & Continua*, vol. 62, no. 2, pp. 631–655, 2020.

[5] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.

[6] D. Xue, L. Wu, Z. Hong, S. Guo, and L. Gao, "Deep learning-based personality recognition from text posts of online social networks," *Applied Intelligence*, vol. 48, no. 11, pp. 4232–4246, 2018.

[7] U. Gupta, A. Chatterjee, R. Srikanth, and P. Agrawal, "A sentiment-and-semantics-based approach for emotion detection in textual conversations," 2017, https://arxiv.org/abs/1707.06996.

[8] M. Z. Asghar, F. Subhan, M. Imran, F. M. Kundi, A. Khan, S. Shamshirband, A. Mosavi, P. Csiba, and A. R. Varkonyi-Koczy, Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content," *Computers, Materials & Continua*, vol. 63, no. 3, pp. 1093–1118, 2020.

[9] H. Ahmad, M. U. Asghar, M. Z. Asghar, A. Khan, and A. Mosavi, "A Hybrid Deep Learning Technique for Personality Trait Classification from Text," *IEEE Access*, vol. 9, 2021.

[10] N. A. S. Winarsih and C. Supriyanto, "Evaluation of classification methods for Indonesian text emotion detection," in *Proceedings of the 2016 International Seminar on Application for Technology of Information and Communication (ISemantic)*, pp. 130–133, IEEE, Indonesia, Asia, August 2016.

[11] S. M. Mohammad and S. Kiritchenko, "Using hashtags to capture fine emotion categories from tweets," *Computational Intelligence*, vol. 31, no. 2, pp. 301–326, 2015.

[12] A. Davletcharova, S. Sugathan, B. Abraham, and A. P. James, "Detection and analysis of emotion from speech signals," *Procedia Computer Science*, vol. 58, pp. 91–96, 2015.

[13] Y. An, S. Sun, and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," in *Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pp. 635–638, IEEE, Wuhan, China, May 2017.

[14] L. Dini and A. Bittar, "Emotion analysis on twitter: the hidden challenge," in *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pp. 3953–3958, LREC, Portorož, Slovenia, May 2016.

[15] P. Kaewyong, A. Sukprasert, N. Salim, and A. Phang, "The possibility of students' comments automatic interpret using lexicon based sentiment analysis to teacher evaluation," in *Proceedings of the 3rd International Conference on Artificial Intelligence and Computer Science*, vol. AICS2015, pp. 179–189, George Malaysia, September 2015.

[16] A. Sen, M. Sinha, S. Mannarswamy, and S. Roy, "Multi-task representation learning for enhanced emotion categorization in short text," in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 324–336, Springer, Jeju, Republic of Korea, April 2017.

[17] D. Kollias, A. Tagaris, and A. Stafylopatis, "On line emotion detection using retrainable deep neural networks," in *Proceedings of the Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, pp. 1–8, IEEE, Athens Greece, December 2016.

[18] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL based multimodal emotion recognition and sentiment analysis," in *Proceedings of the Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pp. 439–448, IEEE, Barcelona, Spain, December 2016.

[19] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval 2015*, pp. 959–962, ACM, New York, NY, USA, August 2015.

[20] S. Gupta, R. Singh, and V. Singla, "Emoticon and text sarcasm detection in sentiment analysis," in *Proceedings of the First International Conference on Sustainable Technologies for Computational Intelligence*, pp. 1–10, Springer, Jaipur, Rajasthan, India, March 2019.

[21] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[22] H. M. Khan, F. M. Khan, A. Khan, M. Z. Asghar, and D. M. Alghazzawi, "Anomalous behavior detection framework using HTM-based semantic folding technique," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 5585238, 2021.

[23] D. Alghazzawi, O. Bamasag, H. Ullah, and M. Z. Asghar, "Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection," *Applied Sciences*, vol. 11, no. 24, Article ID 11634, 2021.

[24] J. Asghar, S. Akbar, M. Z. Asghar, B. Ahmad, M. S. Al-Rakhami, and A. Gumaei, "2021, Detection and classification of psychopathic personality trait from social media text using deep learning model," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 5512241, 10 pages, 2021.

[25] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "2015, DECAF: MEG-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222.

[26] M. Baali and N. Ghneim, "Emotion analysis of Arabic tweets using deep learning approach," *Journal of Big Data*, vol. 6, no. 1, pp. 1–12, 2019.

[27] M. Osama and S. R. El-Beltagy, "A transfer learning approach for emotion intensity prediction in microblog text," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 512–522, Springer, Cairo, Egypt, October 2019.

[28] N. Prasad, R. Singh, and S. P. Lal, "Comparison of back propagation and resilient propagation algorithm for spam classification," in *Proceedings of the 2013 5th International Conference on Computational Intelligence, Modelling and Simulation*, pp. 29–34, IEEE, Seoul, Republic of Korea, September 2013.