IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 9, 2020, accepted March 29, 2020, date of publication April 14, 2020, date of current version May 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2987842

Classification of Poetry Text Into the Emotional States Using Deep Learning Technique

SHAKEEL AHMAD[1], MUHAMMAD ZUBAIR ASGHAR[2], FAHAD MAZAED ALOTAIBI[3], AND SHERAFZAL KHAN[3]

[1]Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2]Institute of Computing and Information Technology, Gomal University, Dikhan 29220, Pakistan
[3]Computer Science Department, Abdulwali Khan University, Mardan 23200, Pakistan

Corresponding author: Shakeel Ahmad (shakeelkau.fcitr@gmail.com)

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant No. G:49-830-1441. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

ABSTRACT The classification of emotional states from poetry or formal text has received less attention by the experts of computational intelligence in recent times as compared to informal textual content like SMS, email, chat, and online user reviews. In this study, an emotional state classification system for poetry text is proposed using the latest and cutting edge technology of Artificial Intelligence, called Deep Learning. For this purpose, an attention-based C-BiLSTM model is implemented on the poetry corpus. The proposed approach classifies the text of poetry into different emotional states, like love, joy, hope, sadness, anger, etc. Different experiments are conducted to evaluate the efficiency of the proposed system as compared to other state-of-art methods as well as machine learning and deep learning methods. Experimental results depict that the proposed model outperformed the baselines studies with 88% accuracy. Furthermore, the analysis of the statistical experiment also validates the performance of the proposed approach.

INDEX TERMS Deep learning, emotion recognition, poetry text, attention-based C-BiLSTM, formal text, emotional states.

I. INTRODUCTION

The classification of opinions, sentiments and emotional states has gained the attention of experts from different fields like natural language processing, computational linguistics and computational intelligence [1]. There are two types of writings that can be analyzed by machine: formal and informal [2]. The formal textual content pertains to poetry, novels, essays, novel, plays, and official/legal documentation, whereas the informal textual content is about SMS, chat, and social media posts [1], [3].

Due to complex nature of the formal text (poetry), detection and classification of emotional states is a challenging task. For instance, the verse "And the sunlight clasps the earth, And the moonbeams kiss the sea:", taken from the poem "Love Philosophy" (Shelley) conveys a love emotion. The manual strategy for detecting emotional states expressed by the poet in the poetry text is difficult and time-consuming [3].

In recent times, machine learning techniques have been applied successfully for extracting and analyzing emotional states and themes from poetic text [1]–[3]. However, small datasets labeled with a limited number of emotional states are the major limitations of such studies [1], [3]. The existing studies on the detection of emotional states from poetry text have used traditional machine learning techniques [1], [4] with limited datasets tagged with a small number of emotion classes. One of the study [3] conducted on emotion classification from poetry text has used one machine learning classifier, namely Support Vector Machine (SVM) and a BiLSTM classifier, for classifying poetry text into two emotion classes. This gap can be bridged by investigating Attention-based C-BiLSTM model, which can take advantage of both the Convolutional Neural Network (CNN), Bidirectional Long Short Term Memory (BiLSTM), as well as, we also exploited the Attention mechanism of deep learning. Furthermore, we exploit 13 emotion classes, which is an extension in baseline work, for a more accurate classification of emotional states from poetry text.

A. PROBLEM STATEMENT

The detection and classification of emotional states from the formal text (poetry) faces different challenges, such as

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif Naeem.

the small size of baseline datasets tagged with a limited set of emotional classes. Furthermore, classical machine learning classifiers are used to detect and classify emotional states expressed by the poet [1], [4]. However, recently deep learning-based neural network models for emotional state detection have shown promising results in different areas like multimedia sentiment and emotion analysis from images, audio, video, and text. We consider the task of emotional state detection from literary (poetry) text as a multi-mode classification issue. A training dataset Trd={trd1, trd2, trd3...trdn} is tagged with emotional-states ={love, hate, fear, joy, nature, suicides sad, surprise, sad, and alone}. Every poetry text is given an emotional-state. The aim of the work is to design a computational model, trained over a training dataset and which can classify a new poetry text according to the different emotional states.

B. RESEARCH OBJECTIVES

We propose an emotional state classification system from poems, with the following objectives.

RO1: Classification of emotional states from poems by applying the attention-based C-BiLSTM neural network model.

RO2: Evaluating the efficiency of the proposed approach with respect to multiple machines and deep learning techniques with an extension in emotional classes and dataset size.

RO3: Comparing the efficiency of proposed technique with respect to similar works.

C. CONTRIBUTIONS

The contributions made in this study are as follows:

1) To classify emotions in the poem using the Attention-based C-BiLSTM deep learning model.

2) The proposed method's efficiency is evaluated with respect to multiple machine and deep learning techniques with an extension in emotional classes and dataset size.

3) The effectiveness of the proposed method is evaluated with similar works.

4) The proposed method produced significant improvement over similar works.

The remaining article is arranged as follows: In section II presents related work; In section III presented a proposed methodology; in section IV, we present the analyses of experimental results, and finally conclusions along with a possible recommendation for the future work, are presented in section V.

II. RELATED WORK

In this section, a brief review of relevant studies on emotion classification in poetry text is presented.

In recent years a considerable number of works have been carried out by a number of researchers in the field of ''Emotion recognition'' by applying Machine learning techniques. Sreeja and Mahalakshmi [1] developed an emotional state recognition system from poetry text by classifying poetry text on the basis of different emotion categories. For this purpose, the Naïve Bayes machine learning classifier is applied. Promising results are achieved, despite the fact that some poems are not classified accurately. To analyze tweets with respect to different linguistic dimensions like sarcasm, metaphor, and irony, Ghosh (2015) [5] applied multiple the ML classifiers like Decision Tree and Naive Bayes. The 11-point scale was used to perform the comparison and up to the mark, results are achieved. Similar to the Sreeja and Mahalakshmi [1] work, in which they face an issue of limited dataset size, Ghosh [5] also used a limited dataset. Both studies aimed to handle this limitation in the future by the extension of the dataset. Rakshit *et al.* [6] presented a comparison of SVM and Naïve Bayes algorithms to classify poems according to their emotion category. A small dataset yielded promising results. The inclusion of phonemic features can make the system more versatile. People's feelings about music can be indicated by their moods and themes. Bischoff *et al.* [7] proposed a system for analyzing audio input. For this purpose, a support vector machine classifier is implemented, achieving the best results. However, further refinement can be made for mood categorization.

For the computational understanding of poetry, automatic analysis has been previously done in various languages like Arabic, Chinese, German, Malay, Persian etc. Arabic poetry is investigated by Alsharif *et al.* [8] in terms of different types of poems, such as Fakhr, Heja, Retha, and Ghazal. The input Arabic poem is classified into the aforementioned poem categories by applying different machine learning. SVM and Ss-CNN algorithms are implemented by Zehe *et al.* [9] to perform sentiment analysis at sentence-level based on German novels. The SVM performed better than S-CNN classifier. Emotions are classified in Quevado's poetry into different emotion categories. For this purpose, Barros [10] suggested a supervised ML technique. Multiple classifiers were used in experiments. However, further, enhancement can be made to improve a set of emotions for achieving more promising results. Soumya *et al.* [11] conducted a classification of Persian poems based on sonic visualization. Emotions can be detected from audio streams. Satisfactory results were achieved in terms of improved accuracy. In another work, the classification of Punjabi poems was performed by Kaur and Saini [2] by applying multiple machine learning classifiers. Additional variations of poetry text can be investigated to achieve improved results. Another study in the Arabic language is conducted by Almuhareb *et al.* [12] to recognize writing style and word usage in poetry. More types of Arabic poetry, such as verse, hemistich or poem, can be considered for improving the system.

While working in Persian poetry, Rahgozar and Inkpen [13] developed a system Persian for Persian poems' classification into different types. The results obtained are promising, depicting that the suggested approach performed significantly better than the compared methods. However, further enhancement can be made by introducing word embedding based features. While working on supervised machine learning techniques, Can *et al.* [14] developed a system to automatically classify 10 Ottoman's poetry into different categories. The experimental results show that the SVM model outperformed the Naïve Bayes classifier. However, more work is required to add visual features for investigating the ottoman language diachronically.

Automated classification of lyrics in the Thai language is performed by Jareenpon *et al.* [15]. For this purpose, they used supervised machine learning approaches, namely Naive Bayes and Random Forest classifiers and obtained satisfactory results. Further extension can be made by introducing melody. In contrast to Jareenpon *et al.* [15]'s work, Jamal *et al.* [16] developed a system for semantics from Malay poetry at a deeper level by applying machine learning classifiers, namely SVM. Resultantly, 10 themes were classified. More promising results can be attained by considering proper poetry-related features. Recently, researchers have also started to investigate Deep learning models for emotion classification of the text of poetry. Mohanty *et al.* [3] developed an Odia poetry-based emotion classification system. Due to limited corpus size, the system didn't produce promising results. To receive more satisfactory results, the size of the corpus may be extended, as deep learning models generally yield efficient results on large datasets.

The aforementioned studies have applied different Machine learning techniques like Naive Bayes, Support Vector Machine, Decision Tree and others, and deep learning-based approaches, for emotion recognition in poetry text. However, further, inspection is required to replace the traditional techniques of machine learning with more robust deep learning models for efficient recognition of emotions in the poetry text.

III. METHODOLOGY

The suggested methodology is composed of multiple modules: which are: (i) System Overview, (ii) System Architecture, and (iii) Applied Example. Every module is briefly explained as follows:

A. SYSTEM OVERVIEW

In this section, there are 7 major steps in building the proposed deep learning model for emotion classification from the given poetry text. This is accomplished by constructing a multi-label classification model i.e. 13 emotion classes. The proposed system is composed of multiple modules (see Fig. 1): i) Data Acquisition, ii) Preprocessing, iii) Feature representations, iv) Feature extraction v) Feature encoding, vi) Context information generation, and vi) classification. Each of these steps are described below:

1) DATA ACQUISITION

To train a deep learning model one of the most important steps is to collect the data. For this purpose, we have used the dataset acquired from [1] which consists of 9142 posts.

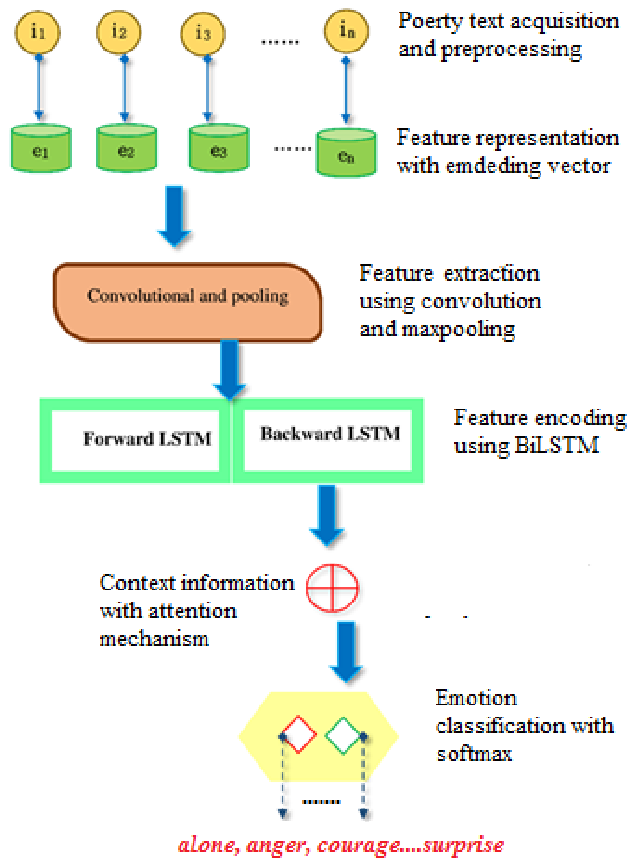


FIGURE 1. Overview of the proposed system.

2) PREPROCESSING

To implement the deep learning model, the next step is to transform the words into numbers. So some of the basic preprocessing steps such as stop-words removal, conversion to lowercase, and tokenization, are performed. After tokenization, a vocabulary is built which transforms the sequences of words into the sequences of integers, where each integer represents a specific word in a vocabulary [18].

3) FEATURE REPRESENTATION

To enable the model to learn, each word is further transformed into an embedding vector by using the Keras embedding layer [19].

4) FEATURE EXTRACTION

In this module, the proposed model extracts the n-gram features from the input received from the previous embedding layer [20].

5) FEATURE ENCODING

A Bidirectional LSTM model [21] is used in the proposed system [21] to capture both the backward and forward dependencies of a word.

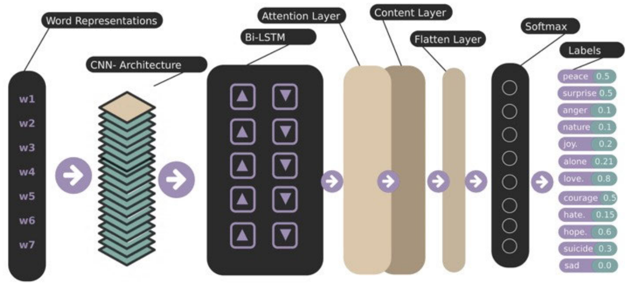


FIGURE 2. Proposed system architecture.

6) CONTEXT INFORMATION

To make the system capable of retaining contextual information, we exploited the attention mechanism in the proposed model. Using this capability, the system can understand, which tokens are useful. The attention mechanism determines which words to focus on the most, by considering only the most relevant input feature and removing the unnecessary or irrelevant information from the input feature [22].

7) CLASSIFICATION LAYER

The final step is to apply the softmax function, implemented in CNN architecture so that the given input text is classified into one of the 13 emotion types.

B. SYSTEM ARCHITECTURE

In the present section, we describe the system architecture of the suggested model for classifying poetry content into several emotion classes, such as *anger, alone, hate, hope, courage, fear, love, joy, sadness, nature, peace, surprise and suicide*. The proposed architecture (Attention-based C-BiLSTM) for emotion classification in poetry text, is composed of seven layers, described as follows:: i) Embedding layer, ii) Convolution layer, iii) Maxpooling layer, iv) BiLSTM layer, v) Attention layer, and vi) classification layer (see Fig. 2)

1) EMBEDDING LAYER

There are two main problems that occur in conventional word representations (one-hot vector): losing words order along with high dimensionality. In comparison to one hot representation, the word embedding is more powerful and suitable [23]. So this work utilizes the word embedding representation scheme. Take a text t with a n length of words (tokens), each word is mapped to its corresponding word embedding e. Resultantly, the text looks like a concatenation of the word embedding, t = $[e_{1,} e_2, e_{3,} \ldots . e_n]$, where e_i is the word embedding for the *i-th* word, projected to a vector $e_i \in R^d$. So for every input text, A sentence matrix $\in R^{dxn}$, is built, where d is the embedding dimension and n is the sentence length. Now, the sentence matrix is further sent to the Convolutional layers.

Two Operations are performed by the CNN layer: Convolution and Maxpooling, to extract the features. These layers are described viz:

2) CONVOLUTION OPERATION

The convolutional layer involves a convolutional operation "$*$", between a poetry text matrix $S \in \mathbb{R}^{dxn}$ and a filter matrix $F \in \mathbb{R}^{mxk}$ [24], which results in an output matrix O, known as feature map. For instance, the feature map is learned as follows:

$$O_{uv} = (S * F) = f(W \circ s_{u:u+k-1,v+d-1} + b) \quad (1)$$

where the bias vector is the b, the weight matrix is W and f represents the nonlinear activation function of the convolutional operation. We used Relu (Rectified nonlinear activation function) as a non-linear activation function because it speeds up the training and produces significantly better results [25].

3) MAXPOOLING OPERATION

The output of the Convolutional layer is now passed on to the pooling layer. This layer aims to further reduce the representation by choosing the maximum value from the pool of numbers, Thereby, discarding the unnecessary information [26]. The pooling operation is formulated as:

$$M_{uv} = \max(O_{u+k-1,v+d-1}) \quad (2)$$

Algorithm 1 Pseudocode for CNN Model for Feature Extraction

Input: L filters, Poetry(text) Matrix S
Output: M_{uv} Pooled Feature Map
Begin
i. **For** each filter l = 1 to L **do**
ii. $O_{uv} = (S * F) = f(W \circ s_{u:u+k-1,v+d-1} + b)$ // Applying operation of convolution to generate the feature map being convolved and introducing of non-linearity using Relu activation iii $M_{uv} = \text{maximum}(O_{u+k-1,v+d-1})$ // Applying maximum pooling operation for the generation of the features significant.
iv. **End For**
End

4) BIDIRECTIONAL LSTM LAYER

To achieve exact predictions, it is necessary that the model should learn the long term dependency in text data. The convolutional layer lacks this capability [26] Therefore, to include this functionality to the proposed model, we applied BiLSTM. BiLSTM allows the model to learn data from both right to left and left to right directions. Hence BiLSTM improves the classification accuracy. There are two independent LSTMs in the Bidirectional LSTM. The forward LSTM computes the hidden state "$\vec{h}$" on the base of past hidden state "h_{t-1}" and the vector input "x_t", while the rearward LSTM computes the hidden state "$\overleftarrow{h}$", on the base of future hidden state "h_{t+1}" and the vector input "x_t". Finally, the vectors of these two directions (forward & backward) are concatenated as a final hidden state of the

BiLSTM to generates an output sequence of hidden vectors H= $[h_1, h_2, h_3, \ldots .h_n]$ using the following Equation 3.

$$\overleftrightarrow{h} = \vec{h} \oplus \overleftarrow{h} \tag{3}$$

The BiLSTM cell is implemented using the following Equations [27].

Forward LSTM:

$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$

$$\sigma f_t = \left(W_f.X + b_f\right) \tag{4}$$

$$\sigma i_t = (W_i.X + b_i) \tag{5}$$

$$\sigma o_t = (W_o.X + b_o) \tag{6}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c.X + b_c) \tag{7}$$

$$h_t = o_t \odot \tau\,(c_t) \tag{8}$$

Backward LSTM:

$$X = \begin{bmatrix} h_{t+1} \\ x_t \end{bmatrix}$$

$$\sigma f_t = \left(W_f.X + b_f\right) \tag{9}$$

$$\sigma i_t = (W_i.X + b_i) \tag{10}$$

$$\sigma o_t = (W_o.X + b_o) \tag{11}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c.X + b_c) \tag{12}$$

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

W_i, W_f, W_o are the weight matrices and b_i, b_f, b_o are biases, which are the parameters of the input gate, forget gate and output gate respectively. σ denotes the sigmoid activation function and $\odot$ is the element-wise multiplication. h_t represents a hidden state vector, and x_t denotes the input vector. Table 1 gives a description of each of the aforementioned mathematical terms.

TABLE 1. Mathematical terms for context information extraction strategy.

Mathematical terms	Definition
a_t	present input
h_{t-1}	past input
h_{t+1}	future input
H	new review matrix created through BILSTM
$\overleftrightarrow{h}$	final review matrix $(\overrightarrow{h_t} + \overleftarrow{h_t})$
i_t, f_t, o_t	input, forget, and output gate
$c\sim_t, c_t, c_{t-1}, c_{t+1}$	candidate value, cell state, past and future cell state
b_i, b_f, b_o, b_c	bias vectors
W_i, W_f, W_o, and W_c	weight metrics regarding input gate, forget gate, output gate, and cell state.
$\circ,\ \sigma, \tau$	product, sigmoid, and tangent function

5) ATTENTION LAYER

Inside a sentence, there are some words, which are irrelevant, while others are decisive. To attend such informative words, the attention mechanism is introduced. Therefore, we added this layer to automatically mine the significant words [28]. The word importance vector u_t is computed using Eq. 14, in which the attention mechanism takes the whole BiLSTM hidden states h as input. W is the weight and b is the bias and tanh is the activation function.

$$u_t = tanh(W_h h_t + b_h) \tag{14}$$

After that, the normalized word weight a_t is obtained through the softmax function using Eq. 15.

$$a_t = softmax(u_t) \tag{15}$$

Finally, to generate the attention mechanism output. A weighted summation is computed using Eq.16.

$$c_t = \sum_{t=1}^{n} a_t h_t \tag{16}$$

The output of Attention layer i.e $\mathbf{c} = [\mathbf{c1}, \cdots, \mathbf{cT}]$ is made input to the Flatten layer.

6) FLATTEN LAYER

To transform the context matrix obtained from the previous layer into a context vector, and to prepare the input for the final classification layer, we applied the flatten layer [29]. The flatten layer operation is performed using the below Eq. 17.

$$f = [c_1 * c_2 * ..c_T] \tag{17}$$

7) OUTPUT LAYER

It is the final layer of our architecture that determines, either the emotion expressed in poetry text is *anger, courage, hate, surprise, alone, joy, fear, peace, love, sadness, nature, surprise and suicide*. A function of softmax activation receives the output of the flatten layer and computes the probability of the emotion class label. It is computed as follows:

$$y_j = \sum_{i}^{m} w_i f_i + b \tag{18}$$

C. APPLIED EXAMPLE

This section presents a step-wise detail of the proposed Attention-based C-BiLSTM model (Fig. 3) for classifying poetry text into following emotion classes: *anger, courage, hate, surprise, alone, joy, fear, peace, love, sad, nature, surprise and suicide.*

1) INPUT PREPARATION AND EMBEDDING LAYER

We take an example of input poetry text: "God made a beauteous garden." The proposed model takes this text as input, analyzes its content, and then automatically assigns a relevant emotion label: anger, courage, hate, surprise, alone, joy, fear, peace, love, sadness, nature, surprise and suicide. Fig. 4 shows a diagrammatic representation.

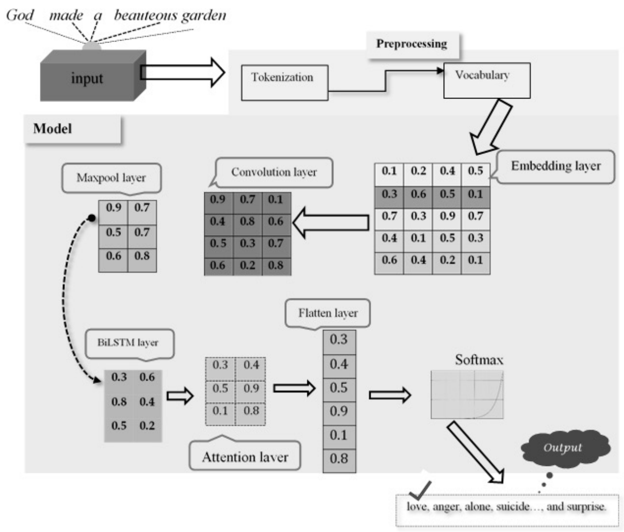


FIGURE 3. Detailed diagram for proposed model.

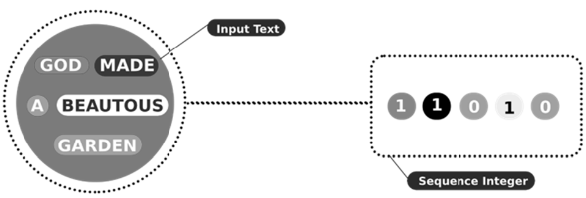


FIGURE 4. Input text with integer sequence.

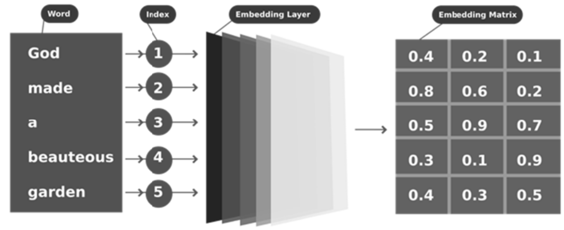


FIGURE 5. Embedding layer.

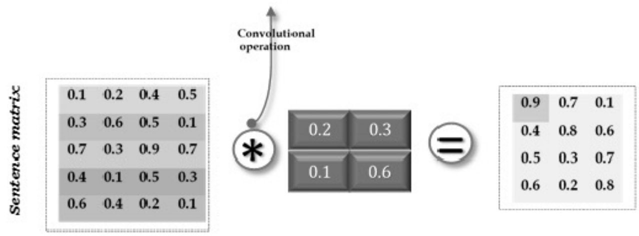


FIGURE 6. Convolutional operation.

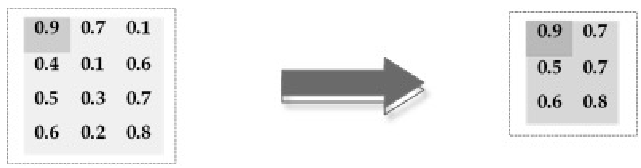


FIGURE 7. Maxpooling operation.

Input Preparation: The first step towards training a deep learning model is Tokenization which breaks down the text into a list of words, such as {God, made, a, beauteous, garden}. Next, a vocabulary is created from all the words in the input text. There are 5 words in our case, so we have a vocabulary of 5 words in which a unique number is allocated to every word, represented as {*God:0, made:1, a : 2, beauteous: 3, garden:4*}

Now we have a sequence of integers (Fig. 4).

Embedding Layer: After encoding the text data into integers (numerical representation), it can be applied to the deep learning model. In this step, each integer is transformed into a vector, resultantly an embedding matrix gets created (Fig. 5)

2) CONVOLUTION NEURAL NETWORK

To facilitate the input of the CNN received from the embedding layer. The CNN has two operations, described as follows:

Convolutional Operation (Feature Extraction): To perform a convolutional operation (Fig. 6), we have two elements namely, (i) embedding matrix (generate from the previous layer), and (ii) feature detector (randomly initialized matrix). Firstly, these two matrices are aligned together, then the feature detector is moved over the selected patch of the embedding matrix to perform element-wise multiplication, and finally, an element of the feature map (output matrix) is obtained through the convolutional operation using Eq. 1.

Mathematical Operation:

$$f(1,1) => 0.1 * 0.2 + 0.2 * 0.3 + 0.4 * 0.1 + 0.5 * 0.6 = 0.9$$

Maxpooling Operation (Feature Selection): This layer aims at selecting the most powerful features. For this purpose, it first selects the window size (commonly 2 or 3). Then a stride of size 1 is selected to control the movement of the window over these feature map, and resultantly the maximum value is selected and inserted into the pooled feature map (Eq. 2) which is the output matrix generated by the Maxpooling operation.

$$P(1,1) => Max(0.9, 0.7, 0.1) = 0.9, \tag{19}$$

Fig. 7 shows the objects of the tuned convolved feature within the output matrix, with the maximum value = 0.9, depicting the first object of the pooled convolved feature. In the same way, other items are received for the input poetry text.

3) BIDIRECTIONAL LSTM LAYER

The BilLSTM receives an input from CNN to preserve long term dependency of the input text. To improve the model efficiency on the emotion classification problem, we have used the Bidirectional LSTM, which is an extension of classical

LSTMs. To understand the word context, the BilLSTM not only considers the previous word but also the coming word. Therefore, it processes information in dual directions: left to right (LSTM Forward) and right to left (LSTM backward).

Mathematical Computation for Forward & Backward LSTM: Using Eq. 4 to Eq. 8 the forward LSTM generates a vector $\vec{h} = [0.10.2]$ and the backward LSTM utilized Eq. 9 to Eq. 13 to generates a vector $\overleftarrow{h} = [0.20.4]$. Finally, the forwardlstm $\vec{h}$ and backwardlstm$\overleftarrow{h}$ are accumulate (Summation of element wise) using Eq. 3 for receiving the ultimate representation $\overleftrightarrow{h}$, which is represented as below:

$$\overleftrightarrow{h} = \vec{h} \odot \overleftarrow{h}$$

After putting the vector in above Equation, we get the following output vector.

$$\overleftrightarrow{h}1 = [0.30.6]$$

Similarly, $\overleftrightarrow{h}2$ and $\overleftrightarrow{h}3$ are calculated.

To further enhance our architecture, the encoded representation is forwarded to the attention layer, which is applied after the BilLSTM layer.

4) ATTENTION LAYER

The attention mechanism consists of the following steps, described below.

Step 1 (Compute the Attention Vector): Using Eq. 14, it computes the word importance vector or attention vector *ut* as follows:

$$u_1 = \tau([0.10.2] . \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix} + [0.3])$$

After inserting the values, we get an output as follows:

$$u_1 = [0.4] , \tag{20}$$

Similarly, we compute $u_2 = [0.7]$, $u_3 = [0.8]$

Step 2 (Compute the Attention Weights): In this step, the attention weights are obtained using Eq. 15 such as

$$a_1 = \frac{e^{[0.4]}}{e^{[0.4} + e^{[0.7]} + e^{[0.8]}}$$
$$a_1 = 0.29,$$

Similarly, we compute $a_2 = 0.4$, $a_3 = 0.8$

Step 3 (Compute the Context Vector): To compute the context vector "c" concatenate each of the attention weights *at* with Bidirectional hidden state ht(Eq. 16).

$$c1 = a1 * h1 + a2 * h2 + a3 * h3$$
$$c1 = [0.30.4]$$

Likewise $c2$ and $c3$ will be computed by putting values in Eq. 16.

Finally, we concatenate the contextual information of each step to provide input to the next layer i.e. Flatten layer.

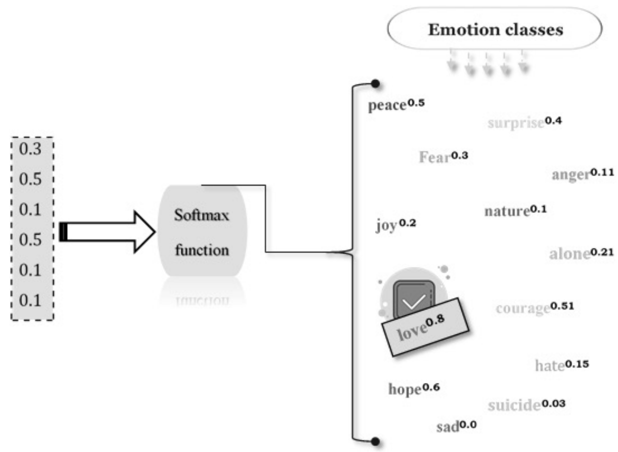


FIGURE 8. Emotion classification of poetry text using softmax function.

5) FLATTEN LAYER

The context matrix acquired from the last layer is converted to a context vector by applying Eq. 17 as shown in Fig. 2.

6) CLASSIFICATION LAYER

The context vector acquired from the last layer (Flatten layer) which is given as input to the classification layer, whereupon a softmax activate function is implemented to compute the probability of the emotion classes: anger, hate, surprise, alone, joy, peace, love, sadness, nature, suicide, courage, fear, and surprise). For the Attention-based C-BiLSTM model, firstly, the input could be calculated (using Eq. 18) as under:

$$y_j = \sum\nolimits_i^m w_i f_i + b$$

Using above formula, the net input is computed as follows:

$$y_{1=}0.7, \quad y_{2=}0.2, \ldots\ldots\ldots, y_{13=}0.5.$$

Applying Softmax Activation Function: Using softmax function, the probability of y is obtained as given below;

$$\sigma\left(y_j\right) = \frac{e^{y_j}}{\sum_{n=1}^{13} e^{y_n}}$$

For Emotion Class Label = 1:
put $j = 1$: $\sigma\,(y_1) = e^1/e^1 + e^2 + \ldots e^{13} = \mathbf{0.8}$
For Class Label = 2:
put j=2: $\sigma\,(y_2) = e^2/e^1 + e^2 + \ldots e^{13} = 0.6.$
For Class Label = 13:
put j=1: $\sigma\,(y_{13}) = e^{13}/e^1 + e^2 + \ldots e^{13} = 0.4$

The computation mentioned above revealed that "love" emotion class attained the highest probability among other emotion classes. So, it is found that the given poetry text "*God made a beauteous garden*" which predicted as *"love"* emotion (see Fig. 8).

The algorithm design of the proposed Attention-based C-BiLSTM for emotion classification of poetry text is presented in Algorithm 2.

Algorithm 2 Steps of pseudo code for Poetry text emotion classification using CNN+BILSTM

Input: Pt_{train}, Pt_{test} from Pt text dataset Pt

Output: Poetry text emotion class for Pt_{test}(*'anger', 'hate', 'surprise', 'alone', 'joy', 'peace', 'love',' sad', 'nature', 'suicide', 'courage, 'fear' and 'surprise'*)

Start

// Encoding Module

1: **while** single input poetry text T ε Pt:
{
2: **while** single word W ε T:
3: Integer(number) allocation to associated word;
};
4: *Parameters initialization of CNN+BILSTM neural network*
5: P_{train} size = 90%, P_{test} Size = 10%, max_features = 80000, ephocs = 7, batch_size = 16, max_length = 53;

// Building Attention-based CNN+BISLTM neural network

6: **while** single input poetry text T ε Pt_{train}:
{
7: Learn the set of word embedding vectors regarding entire words in T = $[t_1, t_2, \ldots, t_{n-1}, t_n]$;
8: Reducing overfitting through the inclusion of dropout layer O ε (x,y) = y (if x=1) else y-1(if x=0)
9: Obtaining features from input text by exploiting convolutional layer (Eq. 1-2);
10: Extracting important features by exploiting pooling layer (Eq. 3-4);
11: Generating new text representation by exploiting BISLTM layer (Eq. 5-17);
12: Adding Attention Layer to focus on significant features (Eq. xx to Eq. xx)
};

// Poetry text emotion classification through trained Attention-based CNN+BISLTM neural network

12: **while** single input text T ε Pt_{test}:
{
13: Trained neural network is developed;
14: Classification of input poetry text by exploiting softmax function (Eq. 19) into different emotion classes;
};
End

TABLE 2. Summary of dataset.

Dataset	Size	Train set	Test Set
Poetry dataset	9142	7313	1829

IV. EXPERIMENTAL RESULTS AND EVALUATIONS

In this section, we give a short description of the experimental setting, Dataset splitting for the suggested model, besides a thorough answer for every question posed in this work.

A. EXPERIMENTAL SETTING AND DATASET SPLITTING

We implement our model based on the Keras -a python library using, OS 64-bit with a memory 4GB, and Intel Core i3. To train our suggested model, we divide the dataset into two parts using the test-train division method of Scikit-learn: *(i)Train data:* We have utilized 80% of the data for training set through which the model learns and is used to fit the model (see Table 3). It contains both the class label and input, and *(ii)Test data*: To evaluate the results of the model on the unseen data. we have applied 20% of the dataset as testing data (see Table 4). When the model is completely trained, then the test set is used. It includes only the input. The description of the dataset is available in Table 2

B. ADDRESSING RESEARCH OBJECTIVES

In this section, we present detailed experiments conducted for addressing research objectives.

1) ADDRESSING RO1

To address RO1: *"Classification of emotional states from poems by applying attention-based C-BiLSTM neural network model*", we implemented various Attention-based C-BiLSTM models for emotion classification from the poetry text inspired by Rani *et al.* [25].

The experimentation is performed with parameter tuning to adjust the classifier efficiency. The parameters related to each layer are as follows: *CNN layer*: a variant of filters from 16 to 64, kernel size such as 2 × 2, 3 × 3, and Pool_size is kept unchanged; *BiLSTM layer:* Units varied from 80 to 30. The parameters for other layers are kept with fixed sizes, such as embedding dimension, dropout, activation function, batch size, and a number of epochs (see Table 5).

Table 6 presents the configuration settings of the selected parameters (kernel size, BiLSTM unit Size, and number of filters) for the 6 Attention-based C-BiLSTM models. In different models, BiLSTM unit size has been tuned with the numbers of filter size and kernel size.

After performing experimentation with parameter tuning, the training time and total accuracy of all the models is included in Table **7**. It is noted that among the other DL variants, the proposed model with parameter settings of kernel_size=2, Number of filters=64 and BiLSTM

TABLE 3. A sample Train Set acquired from the dataset.

No	Poetry	Emotion
1	Out of the cradle endlessly rocking Out of the mocking-birds throat the musical shuttle	Alone
2	Let me confess that we two must be twain Although our undivided loves are one;	Anger
3	Sometime this world was so steadfast and stable That mans word was held	Courage
	obligation;	
4	She knocked the gate Stood out side and preferred to wait	Fear
5	I look down toward earth I had anxiety to know after death	Hate
6	People have their own perception As it is going on since long from the inception	Hope
7	All the days are not equal but I will always have my love for you;	Joy
8	Teach me The Truth! Answer me with your joy along the line	Love
9	We pray the God for well being Ask from heart a little favor to bring	Nature
10	What a happy surprise, You were firm on promise	Peace
11	I am no more like infant now, I have learnt to tackle how	Sad
12	It drives me back forty years, When i had heardbeautiful voice in ears	Suicide
13	But alas! You failed in test ,The story speaks for the rest	Surprise

TABLE 4. A sample Test Set acquired from the dataset.

No	Poetry	Emotion
1	My heart aches and a drowsy numbness pains, My sense as though of hemlock I had drunk	Alone
2	Let not my love be called idolatry, Nor my beloved as an idol show	Anger
3	Take since you bade it should bear,These of the seed of your sowing	Courage
4	it has nothing to do with ups or down you must analyze mistakes and try to own	Fear
5	I had no courage to ask for any change It was assuming with some thought with your age	Hate
6	I did not want to become wealthy man At least I did not dream about pleasing a woman	Hope
7	It was like a dream! And i did see your nakedness beside me; ,For you were very attractive and very beautiful to see.	Joy
8	It is not simple alarm or aimless signal It coveys powerful message in original	Love
9	Happiness in home and in life,No trouble whatsoever even if to walk on edge of knife	Nature
10	To walk hand in hand,And not mere as friend	Peace
11	With all the sense in order With all that accuracy and nothing on border	Sad
12	She was to be my partner for long journey, We had to march ahead without source and money	Suicide
13	The rain God, Readied to share your load	Surprise

units=80, has attained the maximum accuracy of 88%. The results depict that the accuracy result of the model is improved significantly. through increasing the number of filters and BiLSTM units, and by decreasing the kernel size.

Fig. 9 shows the test accuracy and training times of all DL variants in which, the X-axis indicates test accuracy and the y-axis indicates the training time. It‘s noted that the test accuracy result is gradually increased with the increase in the training time with some variations in the symmetry of training time. Also, the learning curve (Fig. 9) depicts that the model is learning from the data.

Table 8 reports the results of assessment metrics, like as recall, f-measures, and precision for different variations of the suggested model. Moreover, the Attention-based C-BiLSTM (Proposed model#6) achieved a maximum (0.88%) recall, (0.88%) precision, and (0.88%.) f1-score.

The accuracy-based comparative results are presented in Fig. 10 for all the models (Attention-based C-BiLSTM). Afterward performing the experimentation with multiple parameters variations of the proposed model, it‘s evident that the proposed Attention-based C-BiLSTM (Proposed model #6) outperformed the other variants of DL models.

2) ADDRESSING RO.2

To address RO2: ‘‘Evaluating the efficiency of the proposed approach with respect to multiple machine and deep learning techniques with an extension in emotional classes’’. ‘‘⊙’’ We have conducted an experiment to investigate the performance of the proposed model with respect to classical deep learning and machine learning techniques. The comparative experimental results of various machine and deep learning techniques are summarized in Table 9.

Part A vs C: During experimentation, we applied different classical classifiers of machine learning using traditional feature representation schemes i.e. (BOW) Bag of Word approach (Tf-IDF CountVectorizer), and the suggested deep learning model (Attention-based C-BiLSTM) based on the basis of task-specific word embedding approach. It is observed that the proposed model yields improved results as compared to ML classifiers including KNN, DT, LR, and RF.

TABLE 5. Attention-based C-BiLSTM model with parameters tuning.

Parameter	Value
Max_features	1000
Max_length	1000
Embedding dimension	128
Dropout bit size	0.5
Number of filters	16,32,64
Convolutional layer Number	1
Pool size	2
Kernel_size	2, 3
BiLSTM unit size	80-30
Attention Layer	1
Activation function	Softmax
Hidden layers Numbers	8
Batch size	32
Epochs Numbers	8

TABLE 6. Parameter tuning for variations of the Proposed model.

Variations of Proposed Attention based C-BiLSTM model	Kernel size	#. of filters	BiLSTM unit size
Proposed model #1	3	16	30
Proposed model #2	2	16	40
Proposed model #3	3	32	50
Proposed model #4	2	32	60
Proposed model #5	3	64	70
Proposed model #6	2	64	80

Some of the causes because of which the performance results of the classifiers in part A is degraded from the proposed methodology in part C, which is listed below.

(i). The traditional BOW method has the limitation of data sparseness along with lack of capturing complex linguistic details which lead to the degraded results, (ii) Secondly, it cannot capture the semantic information of the words, and (iii) The BOW technique takes every word as an independent item, without considering the contextual information about given text [30].

The proposed methodology in C Part performed better as compared to the traditional classifiers in A Part, because the suggested technique has some of the advantage of the words embedding over a traditional bag of words, described as under:

(i) The Attention-based C-BiLSTM is based on the word embedding feature representation approach that is encoded in a low and dense dimensional vector, thus overcoming the issue of data sparseness, (ii) It can capture both the semantic and syntactic information of an input sentence, and (iii) Another reason for the better performance of proposed model is that the word embedding feature representation approach provides a similar representation for the words

TABLE 7. Training time and Test accuracy of Attention-based C-BiLSTM models.

Model Name	Training time(s)	Test accuracy
Attention based C-BiLSTM model [1]	784s	82.67%
Attention based C-BiLSTM model [2]	1565s	83.98%
Attention based C-BiLSTM model [3]	1431s	85.46%
Attention based C-BiLSTM model [4]	1233s	85.57%
Attention based C-BiLSTM model [5]	928s	86%
Attention based C-BiLSTM model [6]	1472s	88%

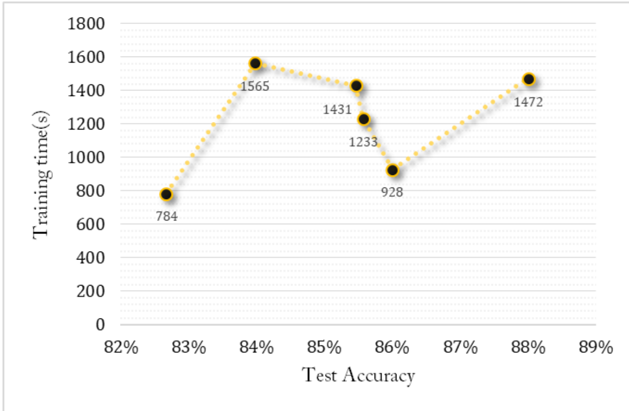


FIGURE 9. Learning curve of training time and accuracy.

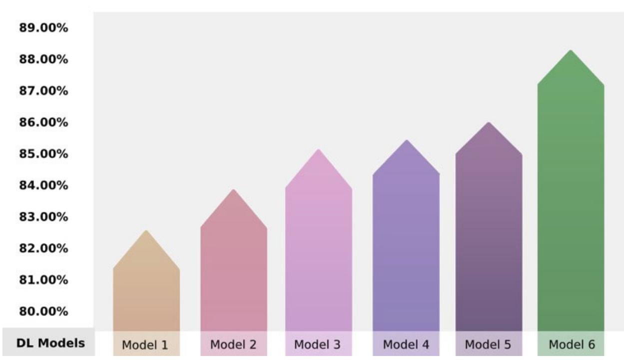


FIGURE 10. Comparison of accuracy for all Attention-based C-BiLSTM models.

having similar meaning [31]. The comparison of the suggested model with the Machine Learning algorithms is represented in Fig.11.

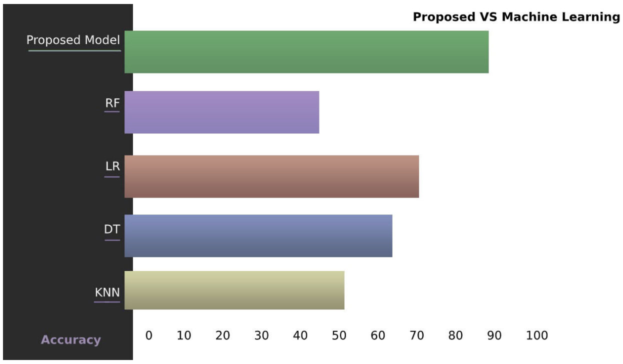


FIGURE 11. Proposed vs ML classifier performance comparison.

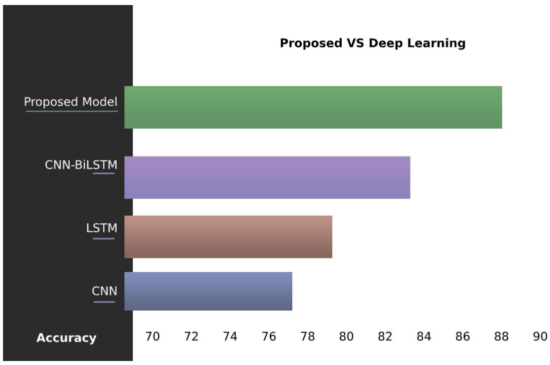


FIGURE 12. Proposed vs DL classifier performance comparison.

Part B vs C: To evaluate the performance results of the suggested Attention-based C-BiLSTM model, we compare it with different models of deep learning, such as LSTM, CNN and the hybrid CNN-BiLSTM models. The results depict the proposed model outperformed the other deep learning models in terms of improved performance due to the following reasons.

CNN vs Proposed: In this experiment, the performance of our proposed model is compared with an individual layer CNN model as an emotion classification of poetry text. Results depict that a single layer CNN model yielded poor results (A: 77%, P:0.78%, R: 0.77% and F:0.77%) as compared to the proposed Attention-based C-BiLSTM model. The cause for the poor performance of CNN because it cannot maintain the sequencing order in the text, which is necessary for the text classification issue to keep track of the ordering detail for yielding enhanced classification results [32].

LSTM vs Proposed: In next experiment, the performance of the LSTM model as compared with the suggested model. Experimental results reveal that the LSTM produces degraded results than the proposed model (A: 79%, P: 0.79%, R: 0.78% and F:0.78%). The major reason why the proposed model gives satisfactory results is that the unidirectional LSTM only preserves past information without retaining future information. Therefore, the unidirectional LSTM is not a suitable choice to perform the feature extraction [33].

TABLE 8. Proposed model evaluation metrices.

Proposed Model	Precision	Recall	F-score
Proposed model #1	0.83	0.83	0.82
Proposed model #2	0.86	0.84	0.84
Proposed model #3	0.87	0.85	0.85
Proposed model #4	0.86	0.85	0.85
Proposed model #5	0.86	0.85	0.85
Proposed model #6	0.88	0.88	0.88

TABLE 9. Performance result evaluation of proposed model with deep and machine learning techniques.

	Methods	P	R	F	A
Part-A Machine learning using traditional features	KNN	0.47	0.52	0.46	51.50 %
	DT	0.62	0.62	0.62	62.49 %
	LR	0.74	0.70	0.70	70.20%
	RF	0.48	0.51	0.47	51.06 %
Part-B Deep learning using Word embedding	CNN	0.78	0.77	0.77	77%
	LSTM	0.79	0.79	0.78	79%
	CNN-BiLSTM	0.83	0.83	0.83	83.05%
Part-C ***Proposed*** *methodology*	**Attention based C-BiLSTM model**	**0.88**	**0.88**	**0.88**	**88%**

CNN-BiLSTM vs Proposed: Finally, for comparing the performance result of the suggested method and CNN-BiLSTM model, we performed an experiment. It is observed that the CNN-BiLSTM exhibit poor results (83.05 % Accuracy, 0.83% Precision, 0.83% Recall, and 0.83% F1-score) with respect to the proposed method. CNN-BiLSTM produces degraded performance as it doesn't contain the attention mechanism as compared to the proposed model [34].

The comparison of the proposed model with the DL models is shown in Fig.12

3) ADDRESSING RO3

While addressing the objective RO3: "Comparing the efficiency of the proposed technique with respect to similar works", we analyze the performance result of the suggested model (Attention-based C-BiLSTM model) for emotion classification in poetry text. Table 8 depicts the comparative

TABLE 10. Comparability with baseline studies.

Study	Method	Precision	Recall	F-score	Accuracy
Baseline-1 P.S and Mahalakshmi [1] *Fine-grained*	NB	0.52	0.54	0.49	54.40%
Baseline-2 Mohanty et al. [3] *Binary*	SVM	0.62	0.62	0.59	62.11%
	BiLSTM	0.81	0.81	0.81	81%
Proposed Methodology ***Fine-grained***	**Attention-based C-BiLSTM model**	**0.88**	**0.88**	**0.88**	**88%**

results. We performed a comparison with the baseline methods and findings are reported in Table 10.

Using different performance evaluation metrics, namely recall, precision, accuracy, and f-measure, the performance results of the suggested system is compared with that of baseline studies [1], [3]. Results reported in Table 8, present that the suggested model outperformed the baseline works with an improved recall, accuracy, precision, and f-measure.

We compare the following Baseline works with our approach.

Baseline-1 vs Proposed: The effectiveness of our suggested model is compared with the work conducted for emotion classification in poetry text [1]. The prior work [1] used classical machine learning methods (NB) for emotion recognition from poetry content, applied on a dataset with a small size and tagged with a small number of emotion categories. We implemented the Attention-based C-BiLSTM model using Anaconda-based Jupiter notebook [35]. However, the proposed model shows significant improvement over the state-of the-art-work by extending the set of emotions.

Baseline-2 vs Proposed: In this experiment, the performance of the proposed model is evaluated with the work performed by [3], in which they used SVM and BILSTM models to perform sentiment analysis for poems. We implemented Attention-based C-BiLSTM and report the results (Table 8).

The results indicates that the proposed model outperformed the baseline work. The SVM employ the hand-crafted features (TF-IDF); hence, limited with respect to learning deep latent features and their correlations [46]. Also the single layer BiLSTM model yielded a performance result degeneracy of up to 6% than the suggested model. We attribute this as

TABLE 11. Evaluation of Performance result Difference between the Proposed model (Attention-based C-BiLSTM) and the Baseline (NB).

	Baseline study [1]	
	Correct Classification	Incorrect Classification
Proposed model Correct Classification	192	42
Proposed model Classification	20	25

the single-layer BiLSTM model is not sufficient for achieving significant results for emotion classification from the poetry text.

The above experiments indicate that the proposed Attention-based C-BiLSTM model has presented improved performance result (88% accuracy) in with respect to the state the art studies [1], [3], conducted on the emotion classification in poetry text. One of the major reasons for obtaining significant results is that we have used a combination of several deep learning models. Firstly, we applied CNN which receives the input from the embedding layer for feature extraction, followed by the BiLSTM layer, which stored the information of both from the future and past inputs. Thus, it has the ability to maintain the contextual detail over a long duration to predict efficiently. Finally, a new phenomenon, known as an attention mechanism, is also utilized. The attention layer determined which words should be given more attention, with respect to other words [37].

Therefore, it is concluded that the overall performance of the proposed methodology is promising because it takes the benefits of CNN, BiLSTM, and Attention mechanism as compared to traditional ML classifier (NB and SVM), and single DL models like BiLSTM.

Statistical Analysis

In order to analyze whether the proposed Deep learning model (Attention-based C-BiLSTM) is statistically significant than the ML classifier (NB) and does not occur accidentally, we have conducted two experiments. For this purpose, we randomly chose 278 poetries from the dataset. Each poetry is classified by both models: Attention-based C-BiLSTM and NB.

The null and alternate hypothesis is formulated as follows: H0: Both models have the same error rate, and HA: The error rate of two models is significantly different.

The McNemar's test statistic (‘‘chi-squared’’) with one degree of freedom is calculated following:

$$\chi^2 = \frac{(|a-b|-1)^2}{(a+b)} \tag{21}$$

Significant results are shown in Table 11.

The results reported in Table 8 are obtained to estimate the baseline performance [1] with the suggested model. It is observed that the baseline method using traditional classifiers has shown poor performance (recall, accuracy, f-measure, and precision) for emotion classification in poetry text.

However, the experiment conducted on emotion classification with a Deep learning approach has performed significantly better than the baseline method [1] by achieving an accuracy of 88%.

The significant test validated that the difference between the performance of the two models: Attention-based C-BiLSTM and NB are statistically different. Table 9 depicts that 62 input poetries are discordant, i.e. the two models behaved differently to the misclassification. After applying McNemar's test, we obtain a chi-squared value of 7.1 and a two-tailed p-value of 0.008 with 1 degree of freedom. Hence, the null hypothesis is rejected (p-value < 0.5) and alternate hypothesis is accepted, i.e. the proposed Deep learning model has received a statistical significance than the ML classifier.

From the aforementioned discussion it is evident that after applying the deep learning model, the performance has been improved for emotion classification in poetry text than the classical machine learning classifier.

V. CONCLUSION

To categorize English poetry text within multiple emotion classes, we have exploited a deep learning technique namely, Attention-based C-BiLSTM model. For experimentation, a benchmark dataset is used with an extension into the emotion classes: Alone, Hope, Nature, and Surprise along with their respective poems. Then this dataset is passed through the following modules, i) Data Acquisition, ii) Pre-Processing, iii) Feature representation, iv) Feature extraction v) Feature encoding, vi) Context information generation, and vii) classification. Results depict that the proposed approach attained highest performance in terms of better (0.88%) precision, (0.88) Recall, (0.88%) f-measure, and(88%) accuracy, as compared to the state of the art studies.

The limitations of this work include: (i) limited size of the poetry dataset, which is needs to be increased to get the model trained with improved accuracy, (ii) the system predicts only one emotion category from 13 classes, which needs to extended to predict multiple emotions with proper ranking mechanism, and (iii) we used only random word embedding model, whereas, there is a need to investigate other pre-trained word embedding models. Furthermore, only text-based features are considered, whereas inclusion of other features like smiles may produce more efficient results. We also plan to validate the effectiveness of the proposed work on multiple datasets.

ACKNOWLEDGMENT

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant No. G:49-830-1441. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

REFERENCES

[1] P. S. Sreeja and G. S. Mahalakshmi, "Emotion recognition from poems by maximum posterior probability," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, pp. 36–43, 2016.

[2] J. Kaur and J. R. Saini, "Punjabi poetry classification: The test of 10 machine learning algorithms," in *Proc. 9th Int. Conf. Mach. Learn. Comput. (ICMLC)*, 2017, pp. 1–5.

[3] G. Mohanty and P. Mishra, "Sad or glad? Corpus creation for Odia poetry with sentiment polarity information," in *Proc. 19th Int. Conf. Comput. Linguistics Intell. Text Process. (CICLing)*, Hanoi, Vietnam, 2018.

[4] Y. Hou and A. Frank, "Analyzing sentiment in classical chinese poetry," in *Proc. 9th SIGHUM Workshop Lang. Technol. Cultural Heritage, Social Sci., Hum. (LaTeCH)*, 2015, pp. 15–24.

[5] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, "SemEval-2015 task 11: Sentiment analysis of figurative language in Twitter," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 470–478.

[6] G. Rakshit, A. Ghosh, P. Bhattacharyya, and G. Haffari, "Automated analysis of Bangla poetry for classification and poet identification," in *Proc. 12th Int. Conf. Natural Lang. Process.*, Dec. 2015, pp. 247–253.

[7] K. Bischoff, C. S. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo, "Music mood and theme classification-a hybrid approach," in *Proc. ISMIR*, Oct. 2009, pp. 657–662.

[8] O. Alsharif, D. Alshamaa, and N. Ghneim, "Emotion classification in Arabic poetry using machine learning," *Int. J. Comput. Appl.*, vol. 65, p. 16, May 2013.

[9] A. Zehe, M. Becker, F. Jannidis, and A. Hotho, "Towards sentiment analysis on german literature," in *Proc. Joint German/Austrian Conf. Artif. Intell.* Cham, Switzerland: Springer, 2017, pp. 387–394.

[10] L. Barros, P. Rodriguez, and A. Ortigosa, "Automatic classification of literature pieces by emotion detection: A study on Quevedo's poetry," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 141–146.

[11] S. Soumya, S. Saju, R. Rajan, and N. Sebastian, "Poetic meter classification using TMS320C6713 DSK," in *Proc. Int. Conf. Signal Process. Commun. (ICSPC)*, Jul. 2017, pp. 23–27.

[12] A. Almuhareb, I. Alkharashi, L. A. Saud, and H. Altuwaijri, "Recognition of classical Arabic poems," in *Proc. Workshop Comput. Linguistics Literature*, 2013, pp. 9–16.

[13] A. Rahgozar and D. Inkpen, "Bilingual chronological classification of Hafez's poems," in *Proc. 5th Workshop Comput. Linguistics for Literature*, 2016, pp. 1–21.

[14] F. Can, E. Can, P. D. Sahin, and M. Kalpakli, "Automatic categorization of ottoman poems," *Glottotheory*, vol. 4, no. 2, pp. 40–57, Jan. 2013.

[15] C. Jareanpon, W. Kiatjindarat, T. Polhome, and K. Khongkraphan, "Automatic lyrics classification system using text mining technique," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–4.

[16] Rang, "Poetry classification using support vector machines," *J. Comput. Sci.*, vol. 8, no. 9, pp. 1441–1446, Sep. 2012.

[17] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1742–1749, Mar. 2014.

[18] *TensorFlow Text Classification Using Attention Mechanism*. Accessed: Jan. 25, 2020. [Online]. Available: https://androidkt.com/tensorflow-text-classification-attention-mechanism/

[19] *Keras Documentation: Embedding*. Accessed: Jan. 2, 2020. [Online]. Available: https://keras.io/layers/embeddings/

[20] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 225–230.

[21] Z. Zhang, Y. Zhang, and T. Zhou, "Medical knowledge attention enhanced neural model for named entity recognition in Chinese EMR," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Cham, Switzerland: Springer, 2018, pp. 376–385.

[22] Q. Zhou and H. Wu, "NLP at IEST 2018: BiLSTM-attention and LSTM-attention via soft voting in emotion classification," in *Proc. 9th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2018, pp. 189–194.

[23] D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, and I. Chorbev, "Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages," *Multimedia Tools Appl.*, vol. 77, no. 24, pp. 32213–32242, 2018.

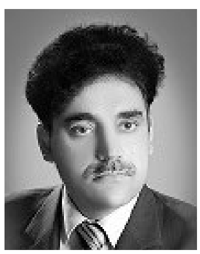
[24] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," 2016, *arXiv:1611.06639*. [Online]. Available: http://arxiv.org/abs/1611.06639

[25] S. Rani and P. Kumar, "Deep learning based sentiment analysis using convolution neural network," *Arabian J. Sci. Eng.*, vol. 44, no. 4, pp. 3305–3314, Apr. 2019.
[26] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2015, pp. 959–962.
[27] X. L. Yao, "Attention-based BiLSTM neural networks for sentiment classification of short textsv," in *Proc. Int. Conf. Inf. Sci. Cloud Comput.*, Dec. 2017, pp. 110–117.
[28] Y. Zhang, J. Wang, and X. Zhang, "YNU-HPCC at SemEval-2018 task 1: BiLSTM with attention based sentiment analysis for affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 273–278.
[29] *What is the Mathematical Formula(Equation) for the Flattening Layer for Deep Learning Model?* Accessed Dec. 19, 2019. [Online]. Available: https://www.researchgate.net/post/What_is_the_mathematical_formulaequation_for_the_flattening_layer_for_deep_learning_model
[30] X. Ding, T. Liu, J. Duan, and J. Y. Nie, "Mining user consumption intention from social media using domain adaptive convolutional neural network," in *Proc. 29th AAAI Conf. Artif. Intell.*, Feb. 2015, pp. 1–10.
[31] E. Rudkowsky, M. Haselmayer, M. Wastian, M. Jenny, Š. Emrich, and M. Sedlmair, "More than bags of words: Sentiment analysis with word embeddings," *Commun. Methods Measures*, vol. 12, nos. 2–3, pp. 140–157, Apr. 2018.
[32] [Online]. Available: https://www.quora.com/What-are-some-of-the-limitations-or-drawbacks-of-Convolutional-Neural-Networks
[33] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," 2015, *arXiv:1511.08630*. Accessed: Feb. 6, 2020. [Online]. Available: http://arxiv.org/abs/1511.08630
[34] Y. Zhu, X. Gao, W. Zhang, S. Liu, and Y. Zhang, "A bi-directional LSTM-CNN model with attention for aspect-level text classification," *Future Internet*, vol. 10, no. 12, p. 116, 2018.
[35] *Anaconda*. Accessed: Nov. 20, 2020. [Online]. Available: https://www.anaconda.com/
[36] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K. F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proc. IJCAI*, Jul. 2016, pp. 3818–3824.
[37] T. Liu, S. Yu, B. Xu, and H. Yin, "Recurrent networks with attention and convolutional networks for sentence representation and classification," *Int. J. Speech Technol.*, vol. 48, no. 10, pp. 3797–3806, Oct. 2018.

MUHAMMAD ZUBAIR ASGHAR is currently an Assistant Professor with the Institute of Computing and Information Technology (ICIT), Gomal University, Dera Ismail Khan, Pakistan, and approved Ph.D. supervisor recognized by Higher Education Commission (HEC), Pakistan. His Ph.D. research interests include game-ai, first-person shooter games, third-person shooter games, computational intelligence, computational linguistics, machine learning, text mining, opinion mining, sentiment analysis, and big data solutions for social networks. He has published more than 40 publications in journals of international reputation (JCR and ISI indexed). He has more than 15 years of University teaching and laboratory experience in artificial intelligence and intelligent systems design. He is a Guest Editor of special issues in the journal of *Social Computing in Health Informatics and Business Intelligence*. He is also a Reviewer of many impact factor journals. He is an Associate Editor of IEEE ACCESS and Plos One.

FAHAD MAZAED ALOTAIBI was born in Saudi Arabia, in 1978. He received the master's degree in computer science from King Abdulaziz University, Jeddah, Saudi Arabia, in 2008, and the Ph.D. degree in computer science and information system from Montfort University, Leicester, U.K., in 2014. In his Ph.D. research, he worked on Simulating Classroom for Training of Trainers using 3D Virtual Environments Supported by Simulation of Smart Board Behaviors. He is currently working as an Associate Professor with the Faculty of Computing and Information Technology, Rabigh, King Abdulaziz University. He is a Certified Trainer from the Canadian Board of Training and Human Resources. He has been a member of Saudi Computer Association and Engineering Association, since 2015. He has an outstanding teaching career with proficient research background, reflecting more than 11 years of teaching and research experience in machine learning, neural networks, sentiment analysis and text mining, and e-learning. He has produced many publications in Journal of international repute.

SHAKEEL AHMAD received the B.Sc. degree (Hons.) from Gomal University, Pakistan, in 1986, the M.Sc. degree in computer science from Qaid-e-Azam University, Pakistan, in 1990, and the Ph.D. degree in computer science from University Science Malaysia (USM), in 2010. In 1990, he started his career as a Lecturer with the Institute of Computing and Information Technology (ICIT), Gomal University, for 11 years. Then, he served as an Assistant Professor, an Associate Professor, a Professor, and the Director Institute of Computing and Information Technology (ICIT), Gomal University, from 2001 to 2014. He is currently serving as a Professor with the Faculty of Computing and Information Technology at Rabigh (FCITR), King Abdulaziz University, Jeddah, Saudi Arabia. He has an outstanding teaching career with proficient research background, reflecting more than 27 years of teaching and research experience in performance modeling, sentiment analysis and text mining, optimization of congestion control techniques and electronic learning. He has produced many publications in Journal of international repute and presented articles in International conferences.



SHERAFZAL KHAN was born in Swabi, Pakistan, in 1970. He received the M.Sc. and M.Phil. degrees in mathematics from Quaid Azam University, Islamabad, Pakistan, in 1995 and 1997, respectively, and the Ph.D. degree from the University of Central Punjab, Pakistan, along with research at the University of Queensland, Brisbane, Australia, in formal methods. In his Ph.D. thesis, he worked on extension of Object-Z towards the fuzzy multi agent specification language (FMASL) for specification and verification of open-block railway interlocking system. He is currently working as a Professor with the Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan. He also worked as a Professor with the King Abdul Aziz University, Saudi Arabia, and Usman Ghazi University, Turkey. He has organized a series of conferences International Conference on Computing and Social Sciences (ICCSS). He was the Chair of ICCSS-1 held in Mardan and ICCSS-2 in the Turkey of Tyeb Ordouan University. His current research interests include formal methods, fuzzy logic, soft sets, and machine learning.