HARC-New Hybrid Method with Hierarchical Attention Based Bidirectional Recurrent Neural Network with Dilated Convolutional Neural Network to Recognize Multilabel Emotions from Text

Md Shofiqul Islam [1,2], Sunjida Sultana [3], Uttam Kumar Roy [4], Jubayer Al Mahmud [5], SM Jahidul Islam [6]

[1]Faculty of Computing, Universiti Malaysia Pahang, 26300, Kuantan, Pahang, Malaysia
[2]IBM Centre of Excellence (Universiti Malaysia Pahang), Cybercentre, Pahang Technology Park, 26300 Kuantan, Pahang, Malaysia
[3]Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh
[4]Assistant Programmer at Bangladesh Bank-The Central Bank of Bangladesh.Head Office, Motijheel, Dhaka 1000
[5]Senior Software Engineer at Charja Solutions Limited,129-Kha/1, Elephant Road, New Market, Dhaka-1205
[6]Senior Software Engineer at Oscillo Soft Private Limited, Dhaka-1205

ARTICLE INFO

Article history:

Received April 13, 2021
Revised April 24, 2021
Accepted April 29, 2021

Keywords:

BiGRU;
BiLSTM;
Convolutional Neural Network;
Sentiment Analysis;
Text Classification;
Natural Language Processing (NLP)

ABSTRACT

We present a modern hybrid paradigm for managing tacit semantic awareness and qualitative meaning in short texts. The main goals of this proposed technique are to use deep learning approaches to identify multilevel textual sentiment with far less time and more accurate and simple network structure training for better performance. In this analysis, the proposed new hybrid deep learning HARC model architecture for the recognition of multilevel textual sentiment that combines hierarchical attention with Convolutional Neural Network (CNN), Bidirectional Gated Recurrent Unit (BiGRU), and Bidirectional Long Short-Term Memory (BiLSTM) outperforms other compared approaches. BiGRU and BiLSTM were used in this model to eliminate individual context functions and to adequately manage long-range features. Dilated CNN was used to replicate the retrieved feature by forwarding vector instances for better support in the hierarchical attention layer, and it was used to eliminate better text information using higher coupling correlations. Our method handles the most important features to recover the limitations of handling context and semantics sufficiently. On a variety of datasets, our proposed HARC algorithm solution outperformed traditional machine learning approaches as well as comparable deep learning models by a margin of 1%. The accuracy of the proposed HARC method was 82.50 percent IMDB, 98.00 percent for toxic data, 92.31 percent for Cornflower, and 94.60 percent for Emotion recognition data. Our method works better than other basic and CNN and RNN based hybrid models. In the future, we will work for more levels of text emotions from long and more complex text.

Md Shofiqul Islam,
Faculty of Computing, Universiti Malaysia Pekan, 26600, Kuantan, Pahang, Malaysia.
Email: shafiqcseiu07@gmail.com

1. INTRODUCTION

Recently, as the social media horizon continues to broaden, the impact it has on people has grown significantly. Many businesses use social media to sell themselves to a specific market group. This is intended to detect and analyze feelings (emotions, sentiments, and opinions) in social media texts on any topic. Methods in emotion perception are used to explain the feeling. Emotion's research enables a thorough understanding of customer opinions expressed on social media or in other forms of feedback [1]. In the last year, the study of emotions, perspective processing, emotion recognition, and document analysis has been a significant area of research. In the NLP context, feeling analyses are needed to define emotion polarity in the initial article, such as positivity, negativity, or neutrality. The importance of Internet users is thanks to the proliferation of online-based social networks [1]. Consumers on the internet are continuously sharing their thoughts and, in the end, creating a full text. Such texts provide a lot of knowledge that all advertisers can find useful when evaluating social networks. In the entire field of sentiment research, there are only a few studies that have yielded positive results. While previous methods looked impressive, they were limited in their ability to handle coherence, qualitative, and semantic knowledge [1]. In the entire area of sentiment research, there are only a few studies that have yielded good outcomes. While previous approaches performed admirably, they were restricted in their ability to handle coherence, qualitative, and semantic information. Opinion mining presents a significant challenge. The majority of machine learning techniques (SVM, LR, NB, K-means, and so on) depend heavily on handcrafted application, which is time-consuming and costly to create and adapt.

Although deep learning has significantly improved the problems in recent years, these artificial neural network approaches will fail to encrypt and understand the part-all relation in the short document [2]. In recent years, a variety of neural network models have developed with better outcomes than other approaches, especially for video classification [3], speech recognition [4], text classification [5], and image classification [6]. In recent years, the Convolutional Neural Network (CNN) [5], Recurrent Neural Network [7][8], Long Short Term Memory (LSTM), BiLSTM [9], and Gated Recurrent Unit (GRU) [10] models have proven to be more effective for textual recognition. After all, the CNN and its recursion developed by Kim in 2014 are used in the deep learning-based sentiment analysis process [11], and also Xu's LSTM job [9].

These methods, which were mentioned in the previous troubles of sentiment analysis, are widely used in several sentiment classification activities, and as a result, they have shown impressive game in experiments. However, modern approaches are often restricted in their ability to use linguistic skills, and text features can be difficult to read in-depth. Furthermore, when encoding features, the training and learning phase isn't sufficient in traditional models. The calculation's findings do not match up. However, decoding emotions is proving to be a difficult task at the moment.

With attention and BiGRU, BiLSTM, and dilated CNN, the HARC model can overcome some of the limitations described above for multilevel sentiment analysis. This model extracts grammatical as well as syntactic features to improve the network's generalization capacity and multilevel sentiment classification accuracy. Initially, textual words are represented as vectors using the Glove vector embedding method, which represents semantic content. The vectors that result is sent to the BiGRU-BiLSTM. The BiGRU and BiLSTM reduce the time it takes for semantic features to travel a long distance. Then CNN learned contextual and local features independently.

Existing approaches have certain drawbacks when it comes to dealing with context and long-ranged characteristics. So it's a difficult job to propose a new approach for multilevel and multilabel data mining. Six deep learning-based tasks for multilevel and multilabel emotion systems are discussed as well as discussed here. These are the proposed system's concentrated activities in conjunction with our process. Their approaches, including their working process, methodology, conclusions, and shortcomings or study holes.

However, for tasks where serial modelling is more relevant, CNN [5] works well. CNN is deep learning as well as a feed-forward neural network with no loops in its node connections. When dealing with long-range functions, this model is ineffective. The LSTM [12] method was created with improved memory usage and recall control in mind. Where classification is determined by long-range semantic dependency rather than any local key phrases, LSTM is easier to use. The LSTM of RNN [13] is a linear classifier in which nodes shape a directed graph over a chain of correlations. One of them is sending a message to a specific person. The LSTM training time is still longer. When compared to traditional window-based machine learning, the Bi-LSTM [9] approach can gather as much contextual knowledge as possible when studying language model, resulting in significantly less noise. For some datasets, this method has a high level of accuracy, but it takes a long time to train compared to other methods. Attention to the GRU [10], the adaptive attention function in GRU will obtain the attention attributes of contextual terms and then produce the target representation dynamically. It works well for target-based emotion, but it can overlook important contextual emotions. BiGRU [14], it works in dual BiGRU and does well enough for target-based sentiments. To handle the multilevel as well as multiclass sentiment analysis, it does not work well. For sentiment classification for short texts, CNN-LSTM [12] is a jointed CNN as well as RNN architecture that takes full advantage of fine-grained feature vectors created by CNN and long-distance dependency learned by RNN. Long-range aspects in a long and complex statement are not handled well. This approach does not allow for individual text visualization. Just a few multi-denominated datasets are available for production.

From the preceding discussion, it can be inferred that conventional systems for multilevel sentiment classification of emotion classification do have limitations, such as handling context adequately [9][12], dealing with time and space complexity [9][15], dealing with lexical dependency [9][12] or semantics [5][12][15], visualization for single input text, functions for big data, not just short data [5], but works for a wide range of data type [12][15]. Furthermore, none of the current methods have sufficient flexibility, adaptability, or coherence to consider sentiment in the sense of context, long-ranged attributes, negation, intensifier, yet clause, and other sentence modifiers. As a result, a new unsupervised, information- or attention-based, highly performed, adaptable, scalable, and better cohesiveness multilevel sentiment analysis methodology remains an open research problem for the proposed system's new work.

The following are the contributions of the proposed model. In the HARC model, a dual framework of BiGRU and BiLSTM layers are used to reduce the gap between grammatical as well as syntactic elements, allowing for a better understanding of meaning. For the specific dependency on terms as well as the internal arrangement of the sentence, we use a dilated CNN with such a dynamically permuted form of vectors. The adjusted algorithm of hierarchical attention is used to optimize feature weights in order to receive better textual information as well as improve hierarchical adaptive tuning performance to increase the results of multilevel sentiment analysis. The following is a breakdown of how this paper is structured. Section 2 lists existing research. Section 3 discusses the architecture, details, deep learning framework theory, and dataset. Section 4 delves into the result and discussions, as well as a discussion. Finally, there is a conclusion in section 5 and an acknowledgment in section 6.

2. RESEARCH METHOD

Multiclass classification is indeed a grading technique for even more than two classes; each mark is exclusive to the other. The classification means that for each data set, only one level is set. For each sample of the Multi-level classification, a number of goal marks are allocated. A data item not mutually exclusive must be identified for deterrence. For example, Tim Horton is frequently listed as a bakery or coffee shop. Multi-label classification has many application forms in real life, such as classification for yelp companies or film labeling for one or more user groups. We focused on multi-level text classification. We conducted our research on multi-level D datasets. For education and text tasks. We perform four tests on that very problem for multi-level text categorization data sets. For example, our tested ER data remarks to assess the effectiveness of our model. For further information on this dataset, Subject I of this technology segment. The data contains information on the comments and every declaration placed on the vector stage.

$$D = \{(C, E) | \ C \ \in \ Documents, E \ \in \ (0,1)^L \tag{1}$$

Within which C is the statement or textual opinion of the entire text E being a vector in six stages and the emotional categories show an increased level L. The purpose of our profound HARC model of a neural network is to evaluate the opinion level for four data groups: Kaggle, Crowdflower, IMDB, ER. The following are the names. Our process includes the following steps: Text integration, BiLSTM, BiGRU, dilation, and evaluation of CNN learning models.

2.1 Text embedding

The concept developed cannot directly accept the input, which is why data must be incorporated. We use the embedding of Glove Vector. This technique of the glove vector is unsupervised, and the phrases are represented by the vector. It maps word for word correlation of distance and semantics on valuable space [16]. Each tokenized word is mapped from the embedding matrix by a matrix of the appropriate word index. We use this Glove vector prior to mapping in our document embedding method. The output of the transformed vector from the Glove vector fed into the designed learning model using BiGRU-BiLSTM layer, CNN layer, and hierarchical attention layer.

Let a sentence S = w_1, w_2, …,w_N with the range N. The aim of this layer is to represent every word in S with such a d-dimensional vector. Every word by each input text is taken from the pre-trained embedded address bus as a static vector for embedding. The series of word vectors X = [x_1, x_2, …., x_N] ϵ R^{N*d} is the output of this embedding layer.

2.2 Learning Model

Two types of neural networks in the approach proposed are: the first is the CNN and BiGRU-BiLSTM of the RNN. Our HARC model includes certain layers called BiLSTM-BiGRU, Dilated CNN, and Hierarchical Attentive Layer during the learning phase. At the beginning of our model, we have an embedding layer with such tokenization. Then BiGRU-BILSTM needs to take a one-dimensional input sequence to retrieve long-distance dependencies. It then transferred its output to the expanded CNN layer, which is processed dynamically, to permute vector output to hierarchical attention from the previous CNN output. Finally, the performance of the hierarchical focus layer is associated with binary cross-entropy, Adam optimizer, and ROC-AUC assessments to estimate the multilevel feeling. Two types of neural networks in the approach proposed are: the first is the CNN and BiGRU-BiLSTM of the RNN. Our HARC model includes certain layers called BiLSTM-BiGRU, Dilated CNN, and Hierarchical Attentive Layer during the learning phase. At the beginning of our model, we have an embedding layer with such tokenization. Then BiGRU-BILSTM needs to take a one-dimensional input sequence to retrieve long-distance dependencies. It then transferred its output to the expanded CNN layer, which is processed dynamically, to permute vector output to hierarchical attention from the previous CNN output. Finally, the performance of the hierarchical focus layer is associated with binary cross-entropy, Adam optimizer, and ROC-AUC assessments to estimate the multilevel feeling.

2.2.1 BiGRU-BiLSTM layer

This layer consists sequence of action by BiLSTM and BiGRU concurrently. GRU cell processing is seen below with certain calculations. In this example, x is used input for it and r for the reset gate as well as h for the hidden state.

$$r_t=\sigma(W_{xr}x_t + W_{hr}h_{t-1} + br) \tag{2}$$

$$z_t= \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \tag{3}$$

$$\hat{h}_t=\tanh(W_{xh}x_t + W_{hh}(r_t \odot h_{\ t-1}) + b_h) \tag{4}$$

$$\hat{h}_t=z_t \odot h_{\ t-1} + (1-z_t) \odot \hat{h}_t \tag{5}$$

In this case, W stands for matrices, b stands for model parameters,¨wise sign mouth function, and $\odot$ for wise proliferation elements. LSTM is usually an RNN extension, requiring the long-term saving of inputs. In contrast to RNN's internal quality memories, LSTM does have an advanced memory. Below are fundamental LSTM operational formulas.

$$i_t = \sigma(w_i[h_{t-1}, x_t]+b_i \tag{6}$$

$$f_t = \sigma(w_f[h_{t-1}, x_t]+b_f \tag{7}$$

$$o_t = \sigma(w_t[h_{t-1}, x_t]+b_o) \tag{8}$$

Here, i_t indicates input gate, o_t for output gate and f_t denotes forget gate. σ represents the activation function, w_x is used to weights of different gates (x), h_{t-1} is the output from the previous LSTM with timestamp t-1, x_t is the current timestamp, b used for bias value in different gates.

Where the context for its feedback is essential, BiLSTMs are especially useful. It is commonly used for the characterization of feelings. Data does not just move backward and backward in two hidden countries with bi-directional LSTM. The Bi-LSTMs are thus more familiar with the context [17]. BiLSTMs are used to scale up the network-usable entry information chunks.
BiLSTM sequences are as follows:

$$\underset{\mathrm{blstm_t}}{\longrightarrow} = \underset{\mathrm{LSTM}}{\longrightarrow}{}^{*}(\underset{(\mathrm{h_{t-1}})}{\longrightarrow}, \mathrm{xt}) \tag{9}$$

$$\underset{\mathrm{blstm_t}}{\longleftarrow} = \underset{\mathrm{LSTM}}{\longleftarrow}{}^{*}(\underset{(\mathrm{h_{t-1}})}{\longleftarrow}, \mathrm{xt}) \tag{10}$$

$$\mathrm{blstm_t}=(\underset{(\mathrm{h_t})}{\longrightarrow}, \underset{\mathrm{h_t}}{\leftarrow}) \tag{11}$$

Thus, the output of BiGRU and BiLSTM encoder is a sequence of vectors as follows:

$$\mathrm{BLSTM} = [\mathrm{lstm_1},\ldots.,\mathrm{lstm_N}] \epsilon R^{N*d'} \tag{12}$$

$$\mathrm{BG} = [\mathrm{bg1},\mathrm{bg2},\ldots.,\mathrm{bgN}\ ] \epsilon R^{N*d'} \tag{13}$$

The output of BiLSTM is sent to BIGRU and then sent to the dilated convolutional layer.

2.2.2 Dilated CNN layer

This is one of the basic layers within our model. The aim of this deeply expanded layer of convolution is to recover hierarchical features of the multi-granularity as a consent form based on text. In contrast to conventional CDNs, which directly use convolution operations at training set embeddings, our dilated CNN relies on Bi-GRU output Vectors which contain contextual details. We state this as followed for the very first block of matrix multiplication.

$$\mathrm{UHV0} = \mathrm{BG} = [\mathrm{bg1},\ldots.,\mathrm{bgN}] \in R^{N*d'} \tag{16}$$

The input matrix containing the input vector variable d. The results within each intermediate structure may thus be viewed as the end block.

$$\mathrm{UHV}^{l}=[\ uhv_1^l\ ,\ldots\ldots,uhv_N^l]\epsilon \mathrm{R}^{\mathrm{N*k}}(\mathrm{l}\epsilon[1,\mathrm{L}]) \tag{15}$$

where L represents the overall number of blocks and k indicates the filter quantity for each block. Now consider the block of l-th numbers, let

$$\mathrm{W}^{l} \in R^{k*w*k} \quad \mathrm{W}^{1} \in R^{k*1*d'} \tag{16}$$

This really is the layer matrix also with k filter kernel's fully convolutional w input vector data. It is necessary to transition two nearby blocks in the following way

$$\mathrm{UHV}_{=}\mathrm{f}(\mathrm{W}^{l}{}_{,}\mathrm{UHV}^{\ l-1}) \tag{17}$$

As f is being used as a refined function, selector filters are also used as a window over the w-length entry. Usually, $uhv_t^l \epsilon \mathrm{UHV}^{l}$ is done as follows

$$uhv_t^l = \mathrm{ReLU}(\mathrm{Wl}\oplus[uhv_{t+1r}^{l-1}]_{t=0}^{w-1} \tag{16}$$

where, ⊕ is used for concatenation operator, the ⊕ do the convolution operation, and r for dilated rate with CNN. Rectified linear units (ReLU) are just an activation function.
We utilize a dilation mechanism after the BiGRU-BiLSTM layer where the dilation rate is twice at each block with a zenith rate 2^{L-2} and every field of bock is increased by maximum width $(w-1)2^{L-1}$. Know that r = 1 for standard dilation [18]. Finally, hierarchical maps of UHV^1, UHV^2, …., UHV^L are extracted.

2.2.3 Attention layer

This is just the attention layer for catching and retaining dense Hierarchical depictions with 1D and dilated convolution of the CNN layer, parallel to various kernel dimensions. Dropout [19] is used to make the model extremely unsuitable for overfitting in the mid-stage cycle.

W_j is used for attention layer 1D convolution and transition matrix. For the spatial relations between low-level and higher-level properties, weight W_{ij} codes are important. The designated position is the highest feature of this 3D curved tensor. The dilatated CNN number is n, and the CNN scale is d. Can the feedback be determined by the series brand size by (1, i[-1], n*d). Suppose that u be (none, s, i[-1]) specified, then $\widetilde{uhv}_{\mathrm{j|ı}}$= conv1d (uhv, W). This output vector is sent to the hierarchical attention layer. One convolutional block's output feature considers for instance and formally comes as the dilated convolutional operation in detail. Recall that $\mathrm{UHV}^{l}= [uhv_1^l,\ldots\ldots,uhv_N^l] \in \mathrm{R}^{\mathrm{N*k}}(\mathrm{l}\epsilon[1,\mathrm{L}])$, use as the output of l-th convolutional block.

Let us uhv_i in source attention layer and also the prediction vector $\widetilde{uhv}_{\mathrm{j|ı}}$ for viewing of raw feature to transfer and is got from uhv_i by the multiplication operation with the transformation matrix Wj.

$$\widetilde{uhv}_{J|\iota} = uhv_i * W_j \tag{16}$$

Inside the "softmax routing" feature, b_{ij} is set with zero and changed with the aij scale agreement. Its agreement a_{ij} is calculated on the scale as follows

$$a_{ij} = \widetilde{uhv}_{J|\iota} \tag{16}$$

$$b_{ij} = b_{ij} + a_{ij}$$

2.2.4 Attention aggregation layer

The purpose of this layer is to produce a given and aggregated conceptual vector by agreeing each CNN output is being used as inputs. Following process occurs at focus layer.

$$a_{ij} = \frac{exp(e_i)}{\Sigma_k \exp(e_k)} \tag{22}$$

$$ei = a(q, uhvi) \tag{23}$$

$$a(q, uhvi) = qT * uhvi \tag{24}$$

Where the q is a pattern vector that can be trained. After that, we achieve a set long attention vector by calculating and transmitting the weighted total to the application classification overall goal lengths.

$$o = \sum_i a_i \, uhv_i \tag{25}$$

2.2.5 Prediction layer

The aim of this layer is only to estimate the probability distribution with $p(y|S)$ where y is the class target. The fixed-length as well as attention-based vectors o is supplied by softmax function to the MLP (Multi-Layer Perceptron) category.

$$p(y|S) = p(y|o) = softmax\,(MLP(o)) \tag{26}$$

Fig. 1 shows the overall functioning mechanism of our model. The structure of each framework from the input phrase to the prediction layer is clearly outlined in this figure. In one form, the produced output passes through the next input layers to be processed.

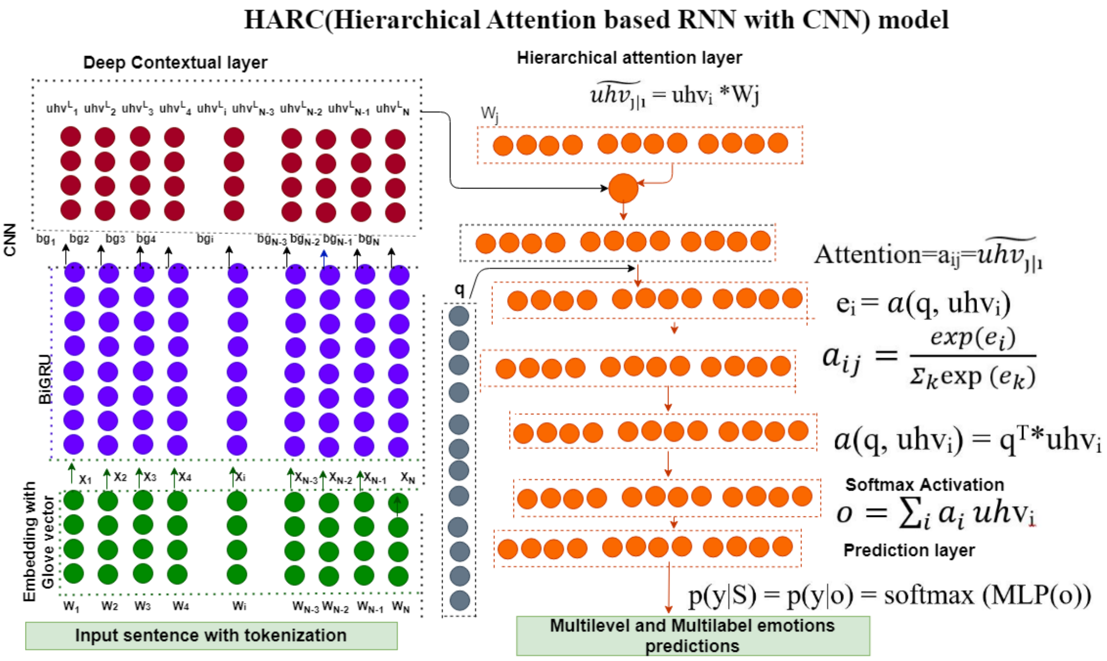

HARC(Hierarchical Attention based RNN with CNN) model
Deep Contextual layer
Hierarchical attention layer
$\widetilde{uhv}_{j|1} = uhv_i * W_j$
W_j
uhv^L_1 uhv^L_2 uhv^L_3 uhv^L_4 uhv^L_i uhv^L_{N-3} uhv^L_{N-2} uhv^L_{N-1} uhv^L_N
CNN
bg_1 bg_2 bg_3 bg_4 bg_i bg_{N-3} bg_{N-2} bg_{N-1} bg_N
BiGRU
x_1 x_2 x_3 x_4 x_i x_{N-3} x_{N-2} x_{N-1} x_N
Embedding with Glove vector
W_1 W_2 W_3 W_4 W_i W_{N-3} W_{N-2} W_{N-1} W_N
q
Attention=$a_{ij}=\widetilde{uhv}_{j|1}$
$e_i = a(q, uhv_i)$
$a_{ij} = \frac{exp(e_i)}{\Sigma_k \exp(e_k)}$
$a(q, uhv_i) = q^T * uhv_i$
Softmax Activation
$o = \sum_i a_i \, uhv_i$
Prediction layer
$p(y|S) = p(y|o) = softmax\,(MLP(o))$
Input sentence with tokenization
Multilevel and Multilabel emotions predictions

Fig. 1. An overall diagram of our HARC model

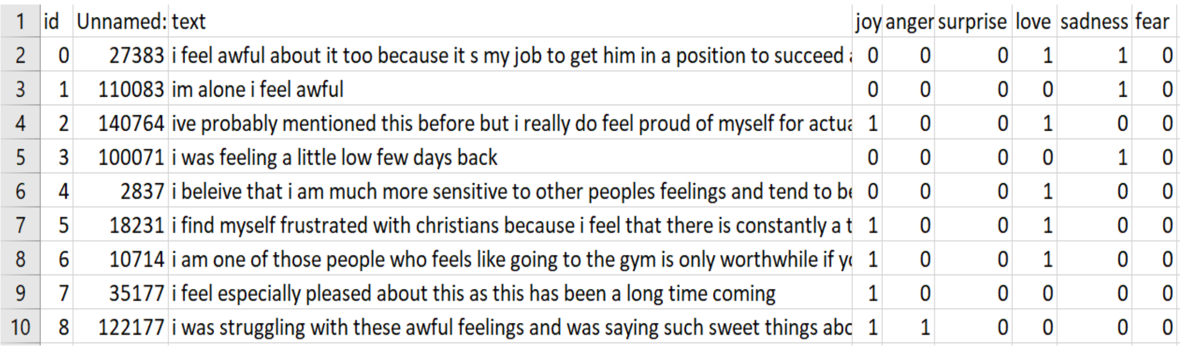
2.3 Data Information

We take out four tests for the issue of multi-level text-based sentiment classification in order to assess the performance of our proposed system. For these data sets, the following is a short description: ER is a dataset for emotion recognition. This dataset consists of multiclass and multilabel information. Six emotions (anger, passion, terror, sorrow, joy, surpass) of 414810 are included in this text. Any emotion is marked with a single emotion. MR is a collection on www.rottentomatoes.com, an English film review by Cornell University. There were 5331 positive movie comments and 5 331 negative film critical thoughts. The usual word period is twenty. The toxic data set includes input from modifications classified as individual toxic conduct via the Wikipedia pages of debate. Six toxicity groups exist. There are 159571 responses in this dataset. A single commentary may include any degree of toxicity throughout one or more of the six levels of toxicity. IMDB [20] is the Internet, and Serious Division ratings compose of 50,000 films [20]. This collection of dates contains 25,000 positive outcomes plus 25,000 negatives. The data collection of Crowdflower includes 40,000 emotional opinions of 13 individual feelings. There are twelve feelings in this dataset. The overall summary of the dataset is given below in Table 1.

Table 1. Overall summary of Dataset

Dataset	Number of classes	Average sentence length	Maximum sentence length	Dataset Size	Test Size	Type
ER	6	160	250	414810	CV	Emotions
Toxic	6	150	200	159572	153165	Sentiment
IMDB	2	120	270	50000	CV	Sentiment
Crowdflower	12	90	200	40000	CV	Emotions

Our approach works for multilevel analysis of feelings (text data with a multi-level sentiment but one text with emotion) and multilevel analysis (one text with many feelings). From the Fig. 1, it is shown that each text is labeled with different emotions. Our method will analysis on this data to find multi-level emotions. See an example of the grouping in feelings of multilevel and multi-class in Fig. 2.



1	id	Unnamed:	text	joy	anger	surprise	love	sadness	fear
2	0	27383	i feel awful about it too because it s my job to get him in a position to succeed	0	0	0	1	1	0
3	1	110083	im alone i feel awful	0	0	0	0	1	0
4	2	140764	ive probably mentioned this before but i really do feel proud of myself for actua	1	0	0	1	0	0
5	3	100071	i was feeling a little low few days back	0	0	0	0	1	0
6	4	2837	i beleive that i am much more sensitive to other peoples feelings and tend to be	0	0	0	1	0	0
7	5	18231	i find myself frustrated with christians because i feel that there is constantly a t	1	0	0	1	0	0
8	6	10714	i am one of those people who feels like going to the gym is only worthwhile if yo	1	0	0	1	0	0
9	7	35177	i feel especially pleased about this as this has been a long time coming	1	0	0	0	0	0
10	8	122177	i was struggling with these awful feelings and was saying such sweet things abc	1	1	0	0	0	0

Fig. 2. Multilevel and multilabel text of Emotion Recognition data

3. RESULTS AND DISCUSSION

The implementation of the proposed HARC model is evaluated and compared to that of some other standard models. We test a broad variety of classification purposes both on small and large datasets. Experiments show that our proposed method outperforms a variety of competitive baselines while still achieving a few state-of-the-art outcomes.

I. CNN [5]: However, for tasks where sequential modelling is more relevant, CNN works well. CNN is deep learning and feed-forward neural networks with no loops in their node connections. When dealing with long-range functions, this model is ineffective.

II. LSTM [12]: This method was created with improved memory and recall control in mind. Where classification is determined by long-range semantic dependency rather than any local key phrases, LSTM is easier to use. Since the text is usually sequential, LSTM displays temporal behaviour and gathers long-range sequential features, making it a "simpler" approach when dealing with text data. The LSTM of RNN is an artificial neural network one where peers form a directed graph over a chain of connections. It's basically a cord made up of neural network blocks connected together. One of them is sending a message to a specific person. The LSTM training time is still longer.

III. Bi-LSTM [9]: As compared to traditional window-based neural networks, this technique can gather as much contextual information as possible when examining word representations, resulting in significantly less noise. This method also employs a max-pooling layer, that automatically decides the words in the expression classification play critical roles in catching key components in the text. For certain datasets, this method has a high level of accuracy, but it takes a long time to train compared to other methods.

IV. Attention GRU [10]: The dynamic attention function in GRU will achieve the attention functions of contextual words and then produce the target expression dynamically. The method's efficiency is improved by dynamically changing the weights. It works well for target-based emotion, but it overlooks important contextual sentiments.

V. BiGRU [13]: It performs well for target-based opinions when used with dual BiGRU. Only the Senti-Drugs dataset performs well. Does not work well for sentiment analysis at multiple levels or across multiple classes.

VI. CNN-LSTM [14]: This seems to be an architecture joined by CNN and RNN and uses the coarse feature points generated by CNN as well as the dependencies learned through the RNN to observe the feelings of short texts. Cannot manage long-range, complicated sentence characteristics.

Our new proposed HARC framework is a dual structured, attentive hybrid BiGRU and BiLSTM network using the expanded CNN hybrid procedure. We propose a new hybrid model that improves the structure and, indeed, the extraction of features by dynamically connecting its hierarchical system to something like an attentive layer to improve the structure of representational learning. In the handling of the BiGRU-BiLSTM layers, and with a hierarchical attention mechanism, we establish an effective, dilated CNN-based structure to automatically extract the hierarchical imagery of a given text. Our suggested HARC model can acquire implicit semantic knowledge effectively. The BiGRU-BiLSTM is used for long distances and interdependent features in all parts of the model. The hierarchical system also has a careful layer capable of extracting rich textual data to enhance its expressive capacity. A hierarchical system to an attentive layer of this hybrid model has the advantage of a less training time as well as a clear network structure to achieve a good output, in comparison with an attentive model which integrates hierarchical self-care procedures and dilated convolutions neural networks (CNNs). It helps in extracting the characteristics of turns of phrase using a fully convolutional n-gram layer-based method and collects dependences on the aspects of the care system to show the rich structure of the sequence of features. The updated hierarchical careful network is used to reconfigure weight parameters for some more reliable text data and to make the adaptive configuration tuning more efficient. Our approach works for multilevel (Text data though one text with emotion) as well as multi-label (one text with multiple feelings).

From the preceding discussion, it can be inferred that conventional systems for multilevel sentiment classification of emotion classification do have limitations, such as handling context adequately [9][12], dealing with time and space complexity [9][14], dealing with lexical dependency [9][12] or semantics [5][12][14], visualization for single input text, functions for big data, not just short data [4], but works for a wide range of data types [12][14], Furthermore, none of the current methods have sufficient flexibility, adaptability, or coherence to consider sentiment in the sense of context, long-ranged attributes, negation, intensifier, yet clause, and other sentence modifiers. As a result, a new unsupervised, information- or attention-based, highly performed, adaptable, scalable, and better cohesiveness multilevel sentiment analysis methodology remains an open research problem for the proposed system's new work.

Table 2 Presents the project training outcome (accuracy and loss) of the comparison of several methods to deep learning with our HARC method and basic guidelines on four text classification benchmarks, small and large, to identify text feelings at a multi-level and multiple-label level. This table sequentially provides the key results of various methods.

Table 2. Experimental result (training set accuracy and loss)

Authors	MODEL	IMDB		Toxic Comment		Crowd flower		ER	
		Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
[5]	CNN	69.94	58.89	96.30	19.90	92.31	74.17	83.33	46.60
[12]	LSTM	67.07	60.01	95.28	18.45	92.29	22.52	82.64	42.49
[9]	Bi-LSTM	82.35	39.89	95.50	16.45	92.31	22.23	82.98	35.00
[10]	Attention GRU	81.45	41.28	96.99	0.951	92.30	22.39	90.64	22.27
[13]	BiGRU	78.60	46.36	97.86	06.08	92.31	22.16	90.32	22.18
[14]	CNN-LSTM	76.87	45.42	95.58	13.12	82.77	41.71	93.88	13.81
Proposed	**HARC** (Hierarchical Attention - BiGRU-BiLSTM-CNN)	82.50	25.42	98.00	13.11	92.31	22.20	94.60	13.76

We compared the performance of our proposed methods with some other new, advanced and recent multilevel text classification with the neural network model. Fig. 3 - Fig. 6 illustrates the graph for training accuracy and loss for four datasets named Toxic, Crowdflower, ER, IMDB, respectively. On each figure, it is clearly shown that training performance is slightly lower than the testing performance. We have implemented all the related methods to get the actual performance from the methods. Raining is colored as blue, and testing is colored as yellow in the analysis figure is given.

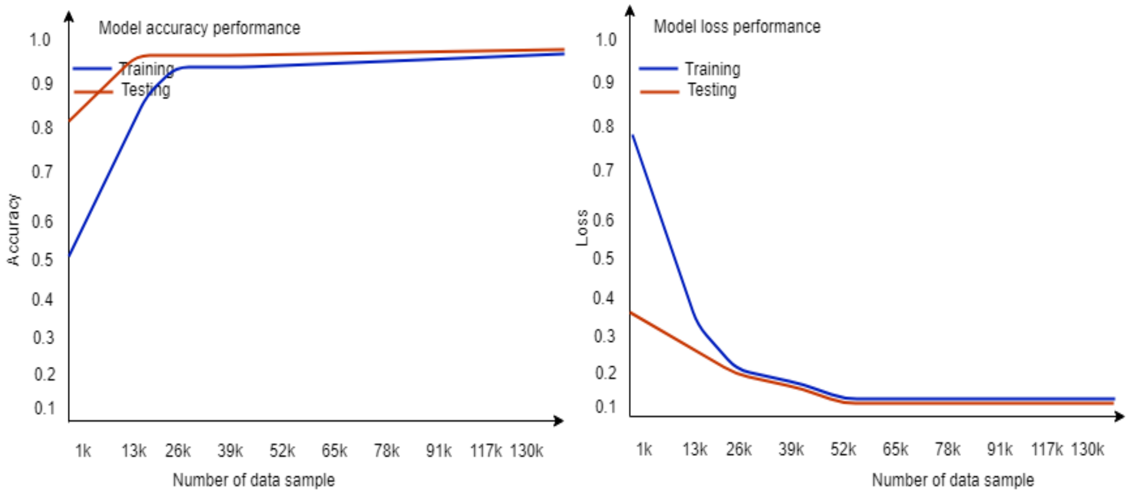


Fig. 3. Train accuracy and loss for toxic dataset

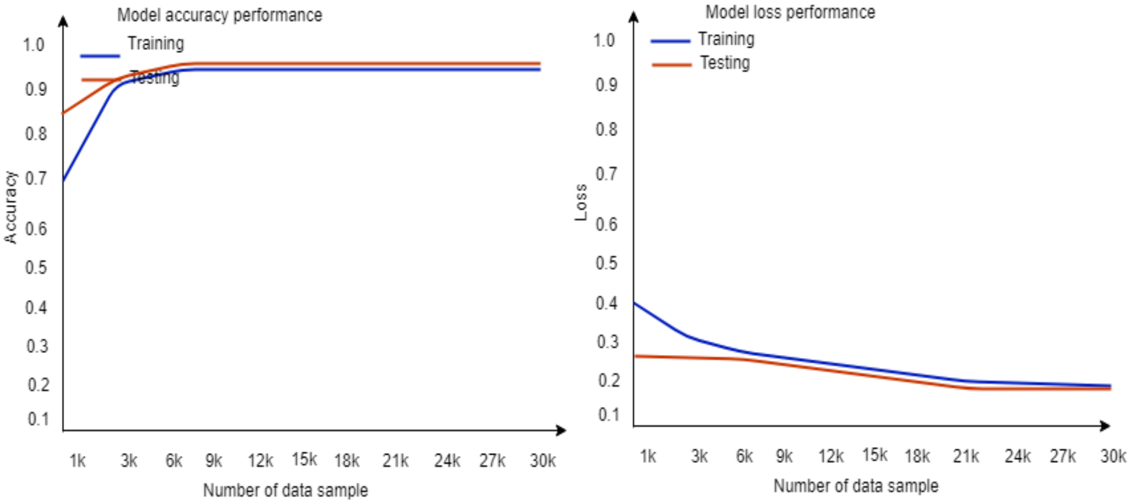


Fig. 4. Train accuracy and loss for Crowdflower dataset

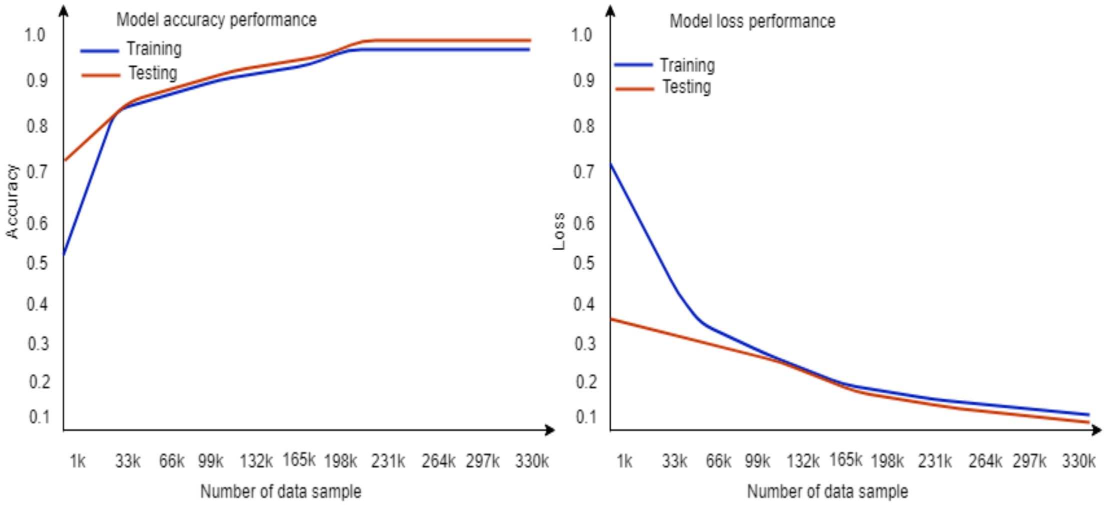


Fig. 5. Train accuracy and loss for ER dataset

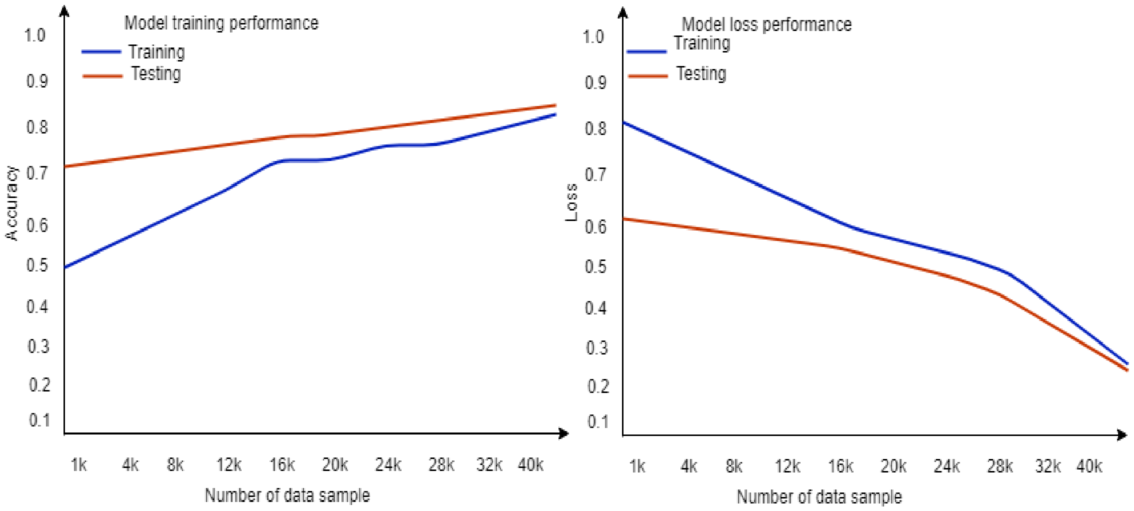


Fig. 6. Train accuracy and loss for IMDB dataset

As well as our ER data set for input text, we also analyze our model performance. We use different colors and bar character for multiple emotions to analyze the classification results. There are six type of multilabel emotions in the Emotion Recognition dataset shown in Fig. 7. The 'Joy' and 'Sadness' are the most part of emotions in the dataset.

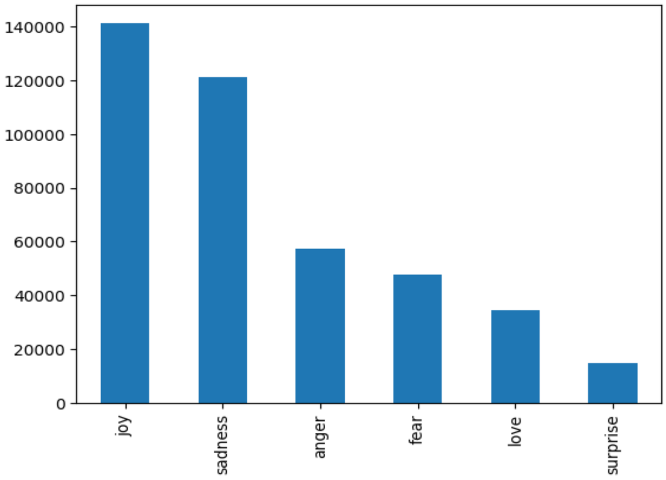


Fig. 7. Emotion Recognition (ER) data overview

We visualize the result for the text input to test this same performance of the proposed method. At last, we added Fig. 8 to show our visualization of predicted emotion with our model. We show text-based input results for emotional prediction here in Fig. 8. The essential words or phrases to assist in getting accurate feelings are colored text on the displayed result. Here we show input text as input, and then the text processed by our model and shows colored which word is more important to recognize emotion.

Input text: i love romantic movie but do not like action movie.
Model processed input text: i love romantic movie but do not like action movie.

4. CONCLUSION

This article described a method for persistent multi-level text classification using a hierarchical attention network-based architecture and a recurrent, dilated convolutional neural network throughout our experiments. The proposed HARC method had an IMDB accuracy of 82.50 percent, 98.00 percent for toxic data, 92.31 percent for Crowdflower data, and 94.60 percent for Emotion recognition data. By dynamically mapping typical hierarchical features to attentive mechanisms, we implement a different hierarchical architecture for enhancing structure learning algorithms and extraction of features within the sentiment analysis mission. We add a dilated CNN-based framework with dual structured bidirectional LSTM and GRU to retrieve hierarchical representations of text automatically, as well as a hierarchical attention mechanism to allow full use of information. Our proposed attention-based hybrid neural network model is capable of effectively obtaining implicit semantic knowledge. This model uses bidirectional GRU and bi-directional LSTM to achieve interdependent features over a long distance.

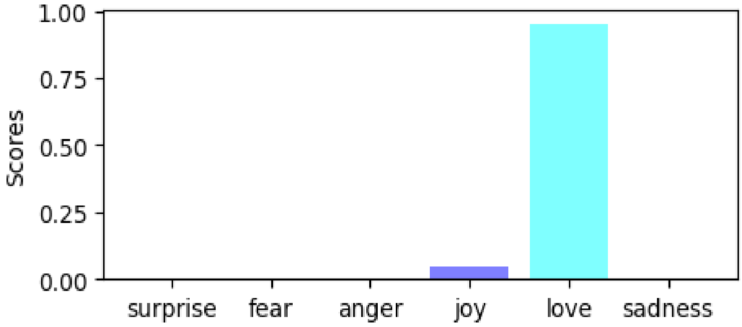


Fig. 8. Output visualization for input text

Furthermore, the hierarchical attention network can extract richer semantic information to enhance expression ability. The hierarchical attention-dependent hybrid model outperforms the attention-based model, which incorporates self-attention mechanisms as well as convolutional neural networks (CNN) to offer a more precise class of emotion by managing long-ranged features and functionalities, and to reduce training time to extract features. In the future, we hope to develop additional data pre-processing techniques such as data expansion by translation and misspelled word techniques. We will then attempt to solve more level of sentiment classifications using different sets of data in order to identify emotions from the text. Our future challenge will also be to handle more multiple sentences and more complicated sentences accurately. Our method's next creation will also be based on handling live data.

Acknowledgments

This research is funded in part by funds from the Malaysian Government’s Fundamental Research Grant Scheme (FRGS) for Universiti Malaysia Pahang (UMP). FRGS/1/2018/ICT02/UMP/02/15 and FRGS/1/2019/ICT02/UMP/02/8 are the grant numbers.

REFERENCES

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1-167, 2012. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

[2] M. S. I. Shofiqul, N. A. Ghani, and M. M. Ahmed, "A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations," *COMPUSOFT: An International Journal of Advanced Computer Technology*, vol. 9, no. 7, pp. 3768-3776, 2020.

[3] M. S. Islam, S. Sultana, U. K. Roy, J. A. Mahmud, "A review on Video Classification with Methods, Findings, Performance, Challenges, Limitations and Future Work," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, Vol 6, No 2, pp. 47-57, 2020. https://doi.org/10.26555/jiteki.v6i2.18978

[4] B. T. Atmaja and M. Akagi, "Deep Multilayer Perceptrons for Dimensional Speech Emotion Recognition," *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2020, pp. 325-331.

[5] S. Lai, L. Xu, K. Liu, J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*. Vol. 29 No. 1, 2015.

[6] G. Algan, I. Ulusoy, "Image classification with deep learning in the presence of noisy labels: A survey," *Knowledge-Based Systems*, vol. 215, 106771, 2021. https://doi.org/10.1016/j.knosys.2021.106771

[7] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," in *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673-2681, Nov. 1997. https://doi.org/10.1109/78.650093

[8] Y. Du, X. Zhao, M. He and W. Guo, "A Novel Capsule Based Hybrid Neural Network for Sentiment Classification," in *IEEE Access*, vol. 7, pp. 39321-39328, 2019. https://doi.org/10.1109/ACCESS.2019.2906398

[9] J. Xu, D. Chen, X. Qiu, X. Huang, "Cached long short-term memory neural networks for document-level sentiment classification," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016. https://doi.org/10.18653/v1/D16-1172

[10] L. LI, A. Zhou, Y. Liu, S. Qian, H. Geng, "Aspect-based sentiment analysis based on dynamic attention GRU," *Scientia Sinica Informationis*, vol. 49, no. 8, pp. 1019-1030, 2019. https://doi.org/10.1360/N112018-00280

[11] Y. Kim, "Convolutional neural networks for sentence classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. https://doi.org/10.3115/v1/D14-1181

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

[13] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014. https://doi.org/10.3115/v1/D14-1179

[14] Y. Han, M. Liu and W. Jing, "Aspect-Level Drug Reviews Sentiment Analysis Based on Double BiGRU and Knowledge Transfer," in *IEEE Access*, vol. 8, pp. 21314-21325, 2020. https://doi.org/10.1109/ACCESS.2020.2969473

[15] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review*, vol. 53, pp. 4335-4385, 2020. https://doi.org/10.1007/s10462-019-09794-5

[16] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014. https://doi.org/10.3115/v1/D14-1162

[17] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint*, arXiv:1511.07122, 2015. https://arxiv.org/abs/1511.07122

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Slakhutdinov "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[19] B. Pang, and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in *Proceedings of the 43rd annual meeting on association for computational linguistics*, Association for Computational Linguistics, P05-1015, p. 115–124, 2005. https://doi.org/10.3115/1219840.1219855

[20] A. L. Maas, R. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th annual meeting of the association for computational linguistics*, pp. 142-150, 2011.

BIOGRAPHY OF AUTHORS

Md Shofiqul Islam, Currently, he is doing Masters (Research-based), a student at University Malaysia Pahang (UMP), Pahang, Malaysia. He has completed my B. Sc. in 2014 in CSE from Islamic University, Kushtia, Bangladesh. Now he is a research assistant at University Malaysia Pahang (UMP). He is also a teacher at CSE under the faculty of FST at ADUST university, Dhaka. He is also in the teaching profession since 2015. His research field is Deep learning, Machine learning, Natural Language Processing, Image Processing. He has published a lot of papers in my field. Email: shafiqcseiu07@gmail.com

Sunjida Sultana was completing master’s degree and completed bachelor’s degree from the Department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. She is working in the field of image processing, video processing, and text processing. Her email is sunjidasultana51984@gmail.com

Uttam Kumar Roy has completed bachelor's and master’s degrees from the Department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. Now he is working as Assistant Programmer at Bangladesh Bank-The Central Bank of Bangladesh, Dhaka 1000. He is also doing his research work in the field of Machine learning, image processing, video processing, and text processing. His email is cseuttamiu@gmail.com

Jubayer Al Mahmud has completed master's and bachelor’s degrees from the Department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. Now he is working as a Senior Software Engineer at Charja Solutions Limited, Dhaka-1205. He is also doing his research work in the field of Machine learning, IoT, image processing, video processing, and text processing. His email is jubayear.iu0708@gmail.com

SM Jahidul Islam has completed bachelor’s degrees from the Department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. Now he is working as Senior Software Engineer at Software firm: Oscillo Soft Private Limited, Dhaka-1205. He is also doing his research work in the field of Machine learning, IoT, image processing, video processing, and text analysis. His email is mjahidiu@gmail.com