International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(2): 822-829

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i02.0659

An Evaluation of the CNN-LSTM Model's Efficacy in Sentiment Analysis Using the Bert and Attendance Mechanisms

Shivaranjini[1], A Srinivasan[1]*, Anjappa SB[2], R Vasanth Selvakumar[3], E Purushotham[1]

[1]Department of Computer Science and Engineering, Sreenivasa Institute of Technology and Management Studies, Chittoor, Andhra Pradesh-517325, India, [2]Department of Mechanical Engineering, Viswam Engineering College, Madanapalle, Andhra Prades-517325, India, [3]Department of Computer Science and Engineering, Viswam Engineering College, Madanapalle, Andhra Pradesh-517325, India.
*Corresponding Author's Email: mechvism@gmail.com

Abstract

Social networking and e-commerce platform demographics have skyrocketed along with the Internet's rapid development. Online user from all around the world the world exchange their opinions and thoughts, which has become a new custom. The volume of data that people express on various platforms has increased as a result of the Internet's expansion. Sentiment analysis is made possible by the availability of these various worldviews and peoples emotion. However, in the NLP domain, the lack of consistent labeled data makes sentiment analysis even more difficult. This article proposes a Convolution Neural Network-long and Short-term Memory model based on BERT and attention. The technique is due to the shortcomings of the existing model in managing long-term dependencies in natural languages. First, feed the Convolution Neural Network's long- and short-term memory with the text vector encoded using Bert. Second, the attention layer receives the output of the Convolution Neural Network and Short-Term Memory model. It uses weighting vectors to extract significant features and the most pertinent information from the input. The findings indicate this. Attention-based models, Constitutional Neural Networks (ABCNN), Hierarchical Attention Networks (HAN), and Bi-LSTM-ATT are contrasted. The macroaverage F1 index and F1-score accuracy both see notable gains from the approach. The suggested model has strong performance in both the macro average F1 indicator and the F1 accuracy scores.

Keywords: Attention Mechanisms, BERT-Convolution Neural Networks, LSTM Neural Network, NLP, Sentiment Analysis.

Introduction

Social networking and e-commerce are two examples of platforms that have seen a rise in popularity recently. On these platforms, a growing number of individuals are sharing their thoughts and experiences on goods and services. The emotional content of these user-generated comments is abundant, offering insightful information on emotional inclinations and mood intensity that is crucial for both consumers and retailers. As such, sentiment analysis methods are essential for examining these reviews (1). In the field of deep learning, models like long short-term memory (LSTM) are frequently used, while convolutional neural networks (CNN) have become commonplace in emotional analysis. Nevertheless, these models have difficulties managing persistent dependencies in natural language, which places restrictions on sentiment analysis assignments (2).

In order to improve CNN-LSTM models' performance in data review and sentiment analysis tasks, this work suggests utilising BERT and attention mechanisms. In particular, this work supports the integration of BERT and attention mechanisms with the CNN-LSTM model, which combines the benefits of CNN with the contextual knowledge of BERT and the focus processes of attention mechanisms. Additionally, the LSTM model is used to enhance the model's capacity for generalisation as well as its capacity to learn representations (3). The addition of BERT and attention mechanisms greatly improves the model's functionality.In order to improve CNN-LSTM models' performance in data review and sentiment analysis tasks, this work suggests utilising BERT and attention mechanisms. In particular, this work supports the integration of BERT and attention mechanisms with the CNN-

(Received 23rd February 2024; Accepted 24th April 2024; Published 30th April 2024)

LSTM model, which combines the benefits of CNN with the contextual knowledge of BERT and the focus processes of attention mechanisms. Additionally, the LSTM model is used to enhance the model's capacity for generalization as well as its capacity to learn representations. The addition of BERT and attention mechanisms greatly improves the model's functionality. Because sentiment analysis is so important to natural language processing, sentiment analysis research is very important. Finding and extracting emotional information from text is the main goal. Conventional methods of sentiment analysis mostly depend on dictionaries or machine learning models (4). Techniques that work well for managing intricate text preprocessing, feature extraction, and model training include random forests, support vector machines, and ordinary Bayes (5).

Researchers are becoming more interested in using deep learning algorithms for sentiment analysis as a result of recent developments in the field. An RNN model for sentence-level sentiment analysis was presented by P.K. Singh et al. (6), and it performed well on a variety of datasets. Furthermore, Kim's CNN model, which was first put forth in 2014 (7), continues to be a fundamental technique in sentiment research. A unique method centred on user and product considerations was developed by Chen et al. (8). Their approach uses attention mechanisms to record how users and products interact, improving sentiment analysis's weighting of user and product ratings.

Sun et al. (9) presented a novel method for feature extraction from input comments in a recent study. For classification, their approach blends a multi-layer perceptron with a convolutional neural network. In addition, the researchers presented a novel model that identifies positive and negative feelings in text by combining BERT and RCNN. To find sentences pertaining to the subject, they also used BERT's Next Sentence Prediction (NSP). The study also emphasises the importance of attention mechanisms in natural language processing, which enables the model to focus on relevant textual portions. In conclusion, sentiment analysis has greatly improved because to deep learning. This work presents a CNN-LSTM model that combines attention processes with BERT to improve sentiment analysis's accuracy and efficacy.

Theoretical Justification and empirical data from earlier research

BERT's Language Representation: Because BERT can capture highly contextualized language representations, it has shown state-of-the-art performance in a variety of NLP tasks. Because BERT is bidirectional, it may infer a word's context from the words around it. This is important for sentiment analysis, as context is a major factor. By allowing models to concentrate on particular segments of the input sequence, attention mechanisms facilitate a more sophisticated comprehension and interpretation of sentiment-related characteristics seen in textual data. When assessing sentiment, the CNN-LSTM model can choose to attend to significant words or phrases by integrating attention processes.Long-range dependencies and sequential patterns are best captured by LSTMs, whereas CNNs are good at capturing local patterns in text (e.g., n-grams). Utilizing the advantages of both architectures, CNNs, and LSTMs can be combined to provide thorough text analysis. Improving sentiment analysis tasks may be achieved by combining CNN-LSTM with BERT and attention methods (10). To effectively comprehend sentiment in textual data, our strategy combines the strengths of CNN and LSTM architectures, leverages attention for targeted feature extraction, and makes use of sophisticated NLP algorithms to gather contextual information. By expanding on this body of knowledge, the proposed research seeks to push the boundaries of sentiment analysis techniques (10).

Clearly articulate the aims and hypotheses of the research, such as determining if CNN-LSTM's integration of BERT and attention mechanisms enhances sentiment analysis accuracy in comparison to standalone models. Using cutting-edge neural network designs and language models, elucidate how the suggested evaluation fills in gaps in current research and advances sentiment analysis approaches (11).

Methodology

To improve accuracy, we apply an attention-based CNN-LSTM model. In order to feed the CNN-LSTM model with text vectors encoded by BERT, preparing the data into training, validation, and test sets is necessary for testing, parameter optimization, and training. Figure 1 shows the structure diagram for the model evaluating the Efficacy of CNN-LSTM Architectures.

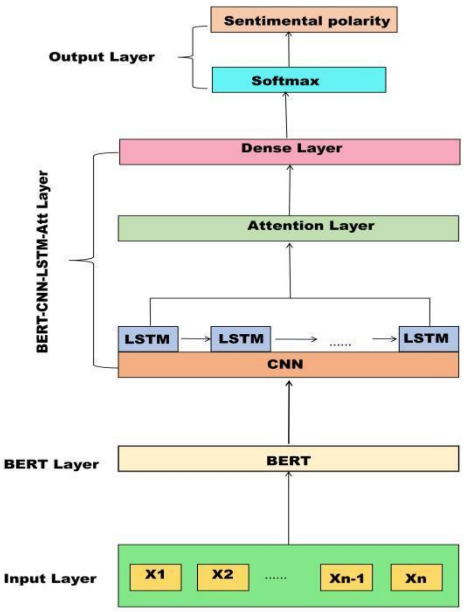

Figure 1: Structure of BERT CNN-LSTM- Att mode

Though the word vector's input layer differs, the classification layers of BERT-CNN and generic neural network models are comparable. While the classic neural network language model employs Word2Vec as a representation of text information, the BERT-CNN model uses a pre-trained language model called BERT as the feature representation of text information. Furthermore, the robustness of the model is increased by BERT-CNN, which uses the word vectors produced by BERT to extract additional features from the Convolutional Neural Network.

Input Layer

1) Information Preprocessing: The first content information is cleaned up, part transformed into words, then turned off to produce a data format that the model could use.
2) Content insertion layer: The content layout is converted into a high-dimensional vector representation following word segmentation. To capture the semantic information of each word, a vector $\{X_1, X_2, ..., X_{n-1}, X_n\}$ is matched with each word according to the job at hand. This article embeds material using a pre-trained BERT model.

Embedding Layer

The embedding layer, which also records the semantic information of words, first converts the text data into a high-dimensional vector space (12). To be more precise, the pre-trained Bert model processes the input text first and gives each word a vector representation. The length of the input text is represented by S, and the dimensionality of each word vector is indicated by d, after these vectors are combined with manually created word vectors. This matrix is used as the input for the next model and as the output of the embedding layer. The BERT architecture, which includes the Bert model, is shown in Figure 2. The token n in the input sequence is represented by En, the transformer block is denoted by Trm, and the matching output embedding is denoted by Tn.

Convolutional Layers

The feature map for the CNN's input layer is the preprocessed text matrix, or S*d. Here, this feature map is processed using the convolutional kernel in order to extract local features. The convolutional layer of the CNN is made up of convolutional kernels of different sizes; the size utilised in this work was h*d. How high is the convolutional layer, h

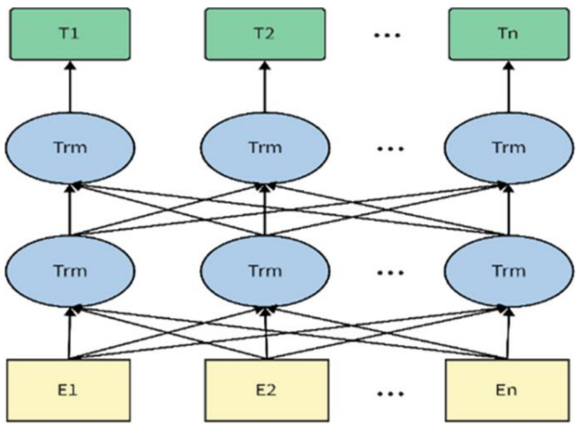

Figure 2: BERT Architecture

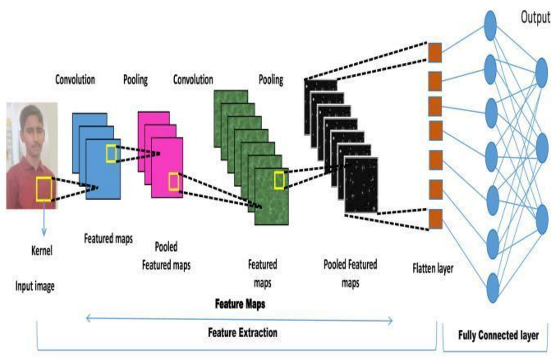

Figure 3: CNN structure diagram

Each of the 128 kernels in the convolutional layers has a size of 3 by 3, 4 by 4, and 5 by 5, respectively. While maintaining the same number of feature maps, the pooling layer—also known as the down sampling layer—serves to reduce the dimensionality of feature maps. It also helps in choosing the most essential feature values. Figure 3 shows the CNN's architectural layout.

LSTM Layer

Particularly unique among recurrent neural networks (RNNs) is the LSTM layer, which is mainly intended to process sequential data organized in a temporal sequence, like text data. The LSTM layer, which consists of three gating units (forget, output, and input gates), controls information flow in the network (13). The LSTM layer and the application expression's operational flow are broken down as follows:

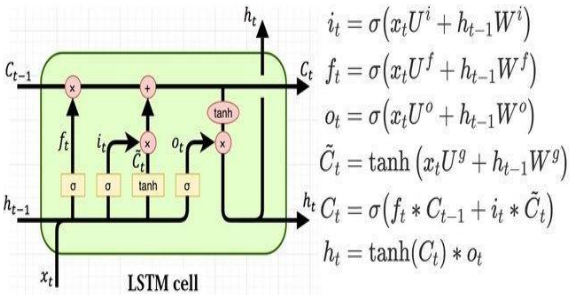

Figure 4 Structure of the LSTM cell and equations

The forgetting gate in Figure 4 determines how much data should be lost, whereas the input gate controls how much information enters the LSTM cell. The output gate controls the cell state's output, and equations are used to compute modifications to the cell state. In this case, I stands for the input gate's output, Cs for the current unit state, and σ for the sigmoid function.

Attention Layer

To improve the model's ability to focus on significant portions of the input sequence, the attention mechanism is applied after the LSTM layers in our BERT-CNN-LSTM-Att architecture. One of the key components of deep learning is the attention layer, which helps the model focus on important parts of the input sequence. In the model covered in this article, relevant information is extracted from the input more easily by adding an attention layer to the network's output that consists of convolutional long short-term memory (LSTM) units. The steps involved are as follows:

$U_i = \tan h\,(W_w\, h_i + b_w)$

$a_i = \text{soft max}\,(u_i)$

$H_a = \sum a_i h_i$ [1]

Output Layer

Predicting the results of comment sentiment categorization is the primary duty of the output layer. The resulting output vector is normalized by applying a softmax activation function, which produces a probability distribution for every class (14).

y=soft max (Whfc+b) [2]

Methods to utilised

Many techniques are used in the BERT-CNN-LSTM-Att model to achieve optimal performance, such as parameter tuning, hyper parameter optimization, and model fine-tuning.

Hyper parameter optimization:

The configuration parameters, known as hyper parameters, control the model's behavior and functionality. The number of layers in each component (BERT, CNN, LSTM), batch size, dropout rate, and learning rate are a few examples. Finding the best combination of these parameters to maximize the model's performance on a validation dataset is the goal of hyper parameter optimization.

Model fine-tuning:

Large-scale corpora provide the initialization parameters for pre-trained models such as BERT. To adjust the model's parameters to the intended job, fine-tuning entails retraining the model using task-specific datasets. Back propagation is used during fine-tuning to update the pre-trained model's parameters based on gradients calculated from task-specific loss functions. To avoid catastrophic forgetting of the pre-trained knowledge, the fine-tuning procedure usually entails freezing certain layers (e.g., the BERT layers) and allowing other layers (e.g., the classification layer) to adjust to the new task.

Parameter tuning:

The process of parameter tuning entails maximizing the internal parameters of each part of the model architecture, including the attention mechanism parameters, the hidden state size and number of LSTM units, and the filter sizes and numbers in the CNN layers. To effectively explore the parameter space and assess how various parameter configurations perform on a validation set, grid search or random search can be applied. The gradients of the loss function for the parameters are used to update the parameters using strategies such as gradient-based optimization techniques (e.g., stochastic gradient descent, Adam).

Results and Discussion

Dataset

The Stanford Sentiment Tree Bank (SST), a popular dataset in sentiment analysis research, is the dataset used in this work. It is a benchmark for assessing sentiment analysis models and consists of 2,210 test examples, 1,016 validation samples, and 11,855 training samples. A sample from the SST dataset is shown in Figure 5, which shows a parse tree with labeled nodes.

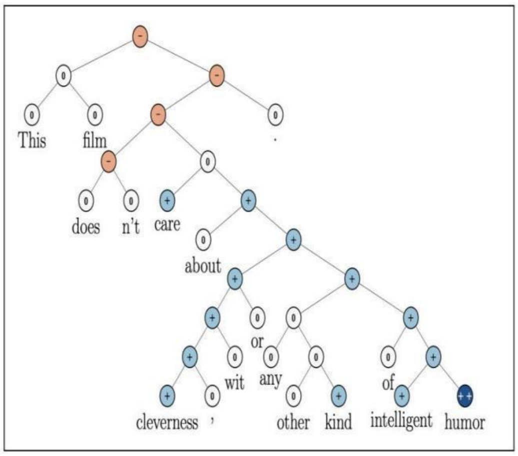

Figure 5. Example SST dataset

Experimental parameter settings

An example of a recursive neural network that can reliably predict five sentiment classes from extremely negative to very positive (- -, -, 0, +, + +), at each parse tree node. It also captures the negation and the extent of it in this particular sentence. Table 1 displays the model's precise parameters along with their explanations.

Table 1: Model parameter configurations.

The experimental parameter's name	Values for parameters
Embedding size	300
hidden size	150
attention size	150
Size of the window	3, 4, and 5
Epochs	10
rate of dropouts	0.5
Optimizer	Adam
Batch Size	32
rate of Learning	0.001

Evaluation Metrics

The evaluation measures in this document are macro F1, accuracy, and F1 score. A clear evaluation of the model's performance is provided by accuracy. The F1 score is the total sum of recall and precision when taking into account both for a classification model. Large F1 is merely a summary of the total performance rating, calculated by averaging the F1 values for every category. Below is the computation formula (14).

$$Accuracy = \frac{TP+TN}{TP+TN+EF+FN} \quad [3]$$

$$Precision(P) = \frac{TP}{TP+EP} \quad [4]$$

$$Recall\ (R) = \frac{TP}{TP+FN} \quad [5]$$

$$F1 = 2 \times \frac{P \times R}{P \times R} \quad [6]$$

The variables TP, TN, and EP represent the number of positive and correct sentiment predictions, negative and correct sentiment forecasts, and negative category errors anticipated to be negative, respectively.

Results and Analysis of the Experiment

As shown in Table 2, the suggested model in this work was tested in trials against five more popular sentiment analysis models. CNNs, LSTMs, and models with attention mechanisms (BiLSTM-ATT, Hierarchical Attention Networks (HAN), and ABCNN) are examples of these models. The experimental results show that, with an accuracy of 90.4% on the test set, the suggested model performs better than the others on the SST-2 dataset. This accuracy is around 1 percentage point higher than the highest that the comparison models could attain. Additionally, with weighted F1 and macro-average F1 scores of 89.5% and 88.2%, respectively, the model covered in this article showed excellent performance. These outcomes outperformed the top-performing alternative models by 0.5% and 0.7%, respectively. To be more precise, the model in this study outperformed the CNN model in terms of accuracy, F1 score, and macro-average F1 score by 1.3%, 1.5%, and 0.9%, respectively. The model showed 1.4%, 0.4%, and 0.9% gains in accuracy, F1 score, and macro-average F1 score, respectively, when compared to the LSTM model. Additionally, the model described in this article performed 0.7% better in accuracy, 0.8% better in F1 score, and 0.9% better in macro-average F1 score when compared to the HAN model.The model demonstrated improvements of 1.1%, 0.4%, and 0.4%, respectively, when comparing to the ABCNN model in terms of accuracy, F1 scores, and macro-average F1 metrics. These findings suggest that the model covered in this article performs better on the SST-2 dataset. Baselines already discussed. Two-thirds of the data are set aside for inference and the remaining 80% for training. The one-shot learning task receives two inputs at the inference step (just like in training). We also perform further ablation tests to illustrate the importance of each element in our model. When we identify an LSTM, the accuracy decreases from 90.2% to 75.74%, indicating the significance of the LSTM in encoding the sequence. Initially, we analyze the significance of the LSTM that encodes the data before feeding it to the other components of the network. The accuracy of the ablation (CNN-LSTM) decreases from 90.1% to 87.74%. The model presented in this research study works well in terms of accuracy, score F1, and macro F1, according to experimental results. The model presented in this research study enhances the macro F1 value by around 0.7% points when compared to the best benchmark model. This demonstrates that the model put forward in this study performs admirably when it comes to sentiment analysis.

This study will also carry out additional analysis. It was found that the attention mechanism may efficiently focus on important information, improving the model's sensitivity to sentiment words, through visualisation of the attention weight matrix. Additionally, each model component's importance was assessed, and it became clear that the LSTM layer and the attention mechanism are essential to the model described in this article. These results provide insightful information that can be used to improve and optimise the model that is suggested in this study article (Figure 6).

Table 2: Experimental results Comparison

Model	F 1 score	Macro F 1	Accuracy
Text CNN	88.2%	86.7%	89.5%
LSTM	89.1%	87.3%	89.0%
Bi LSTM-ATT	88.7%	87.4%	89.7%
Hierarchical Attention Network(HAN)	88.7%	87.4%	89.7%
ABCNN	89.1%	87.8%	89.3%
Ablation (CNN-LSTM)	89.98%	88%	90.1%
Baseline (CNN-LSTM)	88.6%	87.5%	90.2%
BERT-CNN-LSTM-Attention	89.5%	88.2%	90.4%

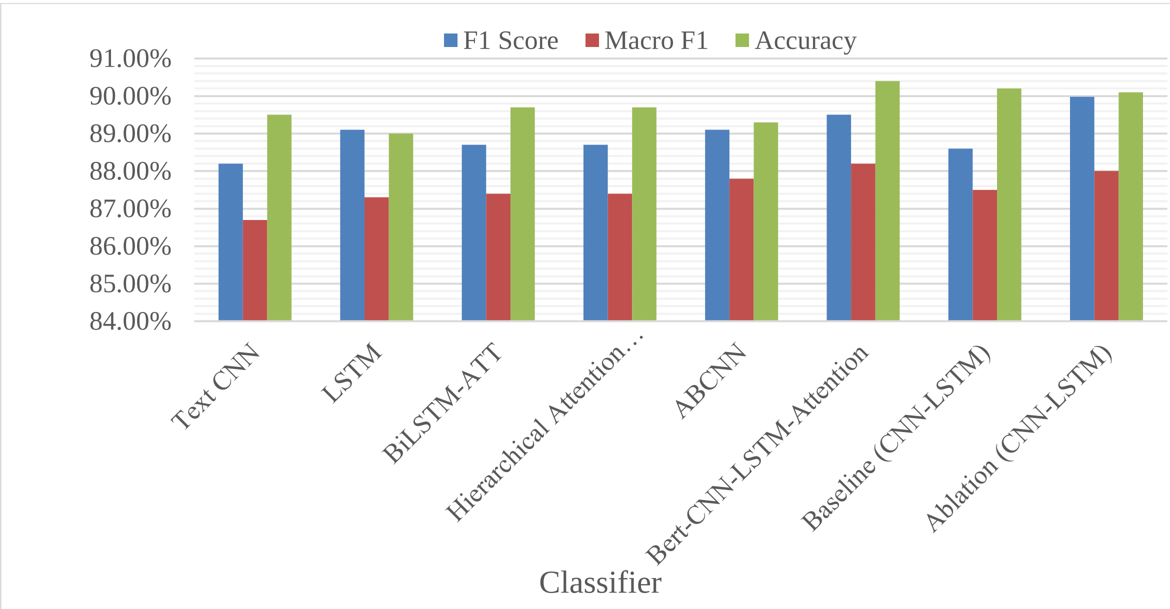

Figure 6: Comparison of experimental results

Conclusion

In conclusion, this work provides an in-depth analysis of attention mechanisms and BERT for sentiment analysis, emphasising data validation. We provide a CNN-LSTM model with these mechanisms integrated. Our tests show that our model achieves the best performance on the SST dataset, outperforming both the conventional CNN-LSTM model and the approach that only uses BERT. We find notable improvements in a number of sentiment analysis tasks at different difficulty levels, especially in negative sentiment analysis that is sophisticated, based on our study of the experimental data. However, we note certain drawbacks, including possible inefficiency with extremely semantically complicated content. To improve performance in this sentiment analysis task, future research should look at larger datasets and more effective model architectures.

Abbreviation

ATT Attention
BERT Bidirectional Encoder Representations from Transformers
Bi LSTM Bidirectional LSTM
CNN Constitutional Neural Network
HAN Hierarchical Attention Network
LSTM Long and Short-term memory
NSP Next Sentence Prediction
RNN recurrent neural network
SST Stanford Sentiment Tree

Acknowledgment

This investigation was carried out by Shivaranjini et al. We thank Sreenivasa Institute of Technology and Management Studies, Chittoor, Andhra Pradesh-517325, India, for granting us access to the facilities we needed to finish this project.

Author Contributions

Shivaranjini: writing- original draft, Validation, Visualization, Investigation, Data curation.

A.Srinivasan: Writing- review and editing, Conceptualization, supervision, Resources.

Anjappa SB: writing Conceptualization, supervision, Resources

R Vasanth Selvakumar: writing Data curation,review and editing.

E Purushotham: Validation, Visualization.

Conflict Interest

The authors declare that none of their personal or financial conflicts could have influenced the work presented in this paper.

Ethics Approval

Not relevant

Funding

No funding from the governmental, private, or nonprofit sectors was given to this study.

References

1. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal. 2014 Dec;5(4):1093–113. https://doi.org/10.1016/j.asej.2014.04.011
2. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2018 Mar 30;8(4). https://doi.org/10.1002/widm.1253.
3. Yue W, Li L. Sentiment Analysis using Word2vec-CNN-BiLSTM Classification [Internet]. IEEE Xplore. 2020. p. 1–5. Available from: https://ieeexplore.ieee.org/document/9336549. https://doi.org/10.1109/SNAMS52053.2020.9336549.
4. Nawaz H, Ali T, Al-Laith A, Ahmad I, Sridevi Tharanidharan, Nazar A. Sentimental Analysis of Social Media to Find Out Customer Opinion. Communications in computer and information science. 2019 Jan 1;110–5. https://doi.org/10.1007/978-981-13-6052-7_10.
5. Iglesias CA, Moreno A. Sentiment Analysis for Social Media. Applied Sciences. 2019 Nov 22;9(23):5037. https://doi.org/10.3390/app9235037.
6. Vartika Bhadana, Garg H. The Sentimental Analysis of Social Media Data: A Survey. Lecture notes in networks and systems. 2021 Jan 1;241–52. https://doi.org/10.1007/978-981-16-0733-2_17
7. Kim Y. Convolutional Neural Networks for Sentence Classification.2014Aug25;https://doi.org/10.48550/arXiv.1408.5882
8. Zhou J, Huang JX, Chen Q, Hu QV, Wang T, He L. Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision, and Challenges. IEEE Access. 2019;7:78454–83. https://doi.org/10.1109/ACCESS.2019.2920075
9. Huang L, Sun X, Li S, Zhang L, Wang H. Syntax-Aware Graph Attention Network for Aspect-Level Sentiment Classification. 2020 Jan 1; https://doi.org/10.18653/v1/2020.coling-main.69
10. Bhuvaneshwari P, Rao AN, Robinson YH, Thippeswamy MN. Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model. Multimedia Tools and Applications. 2022 Feb 19; https://doi.org/10.1007/s11042-022-12410-4
11. Jia K. Sentiment classification of microblog: A framework based on BERT and CNN with attention mechanism. Computers and Electrical Engineering. 2022 Jul;101:108032. https://doi.org/10.1016/j.compeleceng.2022.108032.

12. Huang F, Li X, Yuan C, Zhang S, Zhang J, Qiao S. Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis. IEEE Transactions on Neural Networks and Learning Systems.2021;1–14. https://doi.org/10.1109/TNNLS.2021.3056664
13. Jayakarthik R, Srinivasan A, Goswami S, Shivaranjini, Mahaveerakannan R. Fall Detection Scheme based on Deep Learning Model for High-Quality Life. 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC). 2022 Aug 17; https://doi.org/10.1109/ICESC54411.2022.9885675
14. Prottasha NJ, Sami AA, Kowsher M, Murad SA, Bairagi AK, Masud M, et al. Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. Sensors. 2022 May 30;22(11):4157. https://doi.org/10.3390/s22114157.