https://doi.org/ 10.33472/AFJBS.6.4.2024.1076-1086

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Deep Hybrid models with BERT for Cross Domain Suggestion Classification

Nandula Anuradha[1,a)] and Dr P Vijaya Pal Reddy[2]

[1]Research Scholar, Department of CSE, University College of Engineering, Osmania University, Hyderabad, Telangana, India.
[2]Professor, Department of CSE, Matrusri Engineering College, Hyderabad, Telangana, India.
Corresponding author : evanianuradha2002@gmail.com

Article History

Received:16 April 2024

Accepted: 18 May 2024

Published: 17 June 2024

ABSTRACT

Suggestion analysis involves computationally analysing textual expressions like tips, advices, recommendations etc in online reviews. This work utilizes a deep learning framework using models, such as CNN, LSTM, LSTM-CNN, and CNN-BiLSTM-GRU architectures, with BERT representations. The evaluation of these models is conducted using Software Forum Domain related to electronics reviews and TripAdvisor domain associated with hotel reviews. To assess the model accuracy, precision, recall and F1score metrics are calculated. Notably, transfer learning hybrid models have demonstrated superior performance in suggestion analysis and particularly BERT-CNN-BiLSTM-GRU model.

Keywords: BERT, Machine learning, Deep Learning, Transfer Learning, Natural Language Processing, Suggestion Mining

INTRODUCTION

Opinion reviews always play a crucial role in service or product utilization, and decision-making [1]. Online reviews make a reliable resource on social

platforms for diverse knowledge [2][3]. These reviews often contain suggestions, hints, and recommendations that benefit fellow consumers and product owners [4]. Suggestion Mining, a contemporary field within Natural Language Processing involves identifying and extracting suggestions from customer reviews, focusing on understanding review content rather than just determining sentiments [5][6].

In the age of NLP and sentiment identification, the concept of suggestion mining emerges as a recent endeavour [6]. [7][8][9] introduced the concept of suggestion mining, employing rule-based methodologies to extract insights from reviews. [10][11] aimed to discern insightful feedback from opinion texts, treating them as suggestions, and developed various linguistic rules and patterns for this purpose.

In [10][11], explored the utilization of linguistic features, n-gram analysis, and POS tagging to identify sentences expressing suggestions within customer opinions. They extended their analysis to encompass reviews from both TripAdvisor and Yelp, covering hotels and electronic devices. Till 2015, an adequate definition of suggestion mining was missing. [12] filled this gap by curating annotated datasets and defining the problem through a compilation of reviews from various restaurants, electronic products, Microsoft Windows Phone tweets, and software forums.

To garner further research interest and foster rigorous investigation into suggestion mining, Suggestion Mining was introduced as Task 9 in SemEval-2019. This initiative involved classifying reviews into suggestions and non-suggestions. The datasets are collected from electronics forums domain and trip advisor domain. The task consists of in-domain review classification and cross-domain suggestion review classification. Subtask-A entailed training and evaluating models for classifying reviews into suggestions and non-suggestion within Electronics domain. In subtask-B, the training and evaluation come from different domains. 33 teams made submissions as a part of this competition. Most of the teams incorporated pre-trained models and transfer learning techniques.

In presenting the outcomes, [14][15] utilized a range of foundational Machine learning approaches and hand-crafted rules to classify reviews as either suggestive or not Primarily included partitioning methods like logistic regression, SVM and ensemble methods such as random forests, along with statistical models such as Bayesian inference and genetic algorithms are employed. A number of neural network models such as RRN are implemented for NLP tasks. These models significantly advanced suggestion analysis capabilities. The initial step in suggestion analysis involves converting textual inputs into numerical vector representations. Different methods, such as the bag

of words and sequence of word methods, are employed for text representation, addressing issues such as word order and semantics. [16][17] also devised a hybrid model to identify review texts with suggestion words, incorporating semi-supervised learning methods for extracting customer-to-customer suggestions. [18] A rule-based system incorporating manually crafted features serves as input to Bi-LSTM models, with results reported [13]. [19] initialized deep models with Word2Vec and Glove word embeddings, finding LSTM to be more effective. In both the subtasks, BERT-based models emerged as topmost performers.

In Later years, Transformer models, notably BERT, have revolutionized text representation, learning contextual word relationships effectively. BERT, a pre-trained language model, utilizes deep learning to understand the contextual meanings among words in a text. It has demonstrated superiority in various NLP tasks, including suggestion classification, where a fully connected layer, called BERT-NN, is added for classification purposes.

This paper evaluates the effectiveness of CNN, LSTM, LSTM-CNN, and CNN-BiLSTM-GRU models with BERT to enhance performance of suggestion classification, utilizing Electronics reviews and Hotel reviews datasets taken from SemEval 2019 Task 9.

2. THE PROPOSED MODEL ARCHITECTURE

2.1 BERT-CNN-BiLSTM-GRU Model

Figure 1 depicts the ensemble model, with CNN, BiLSTM, and GRU models. Firstly, each word is transformed into a feature matrix using a BERT pretrained model. Then, feature extraction is carried out by CNN and a max pooling layer. BiLSTM produces predictive label sequences directly for input sentences. Additionally, GRU, a LSTM variant with fewer parameters and simpler training, is employed. These models are combined using a voting approach to yield the final output.

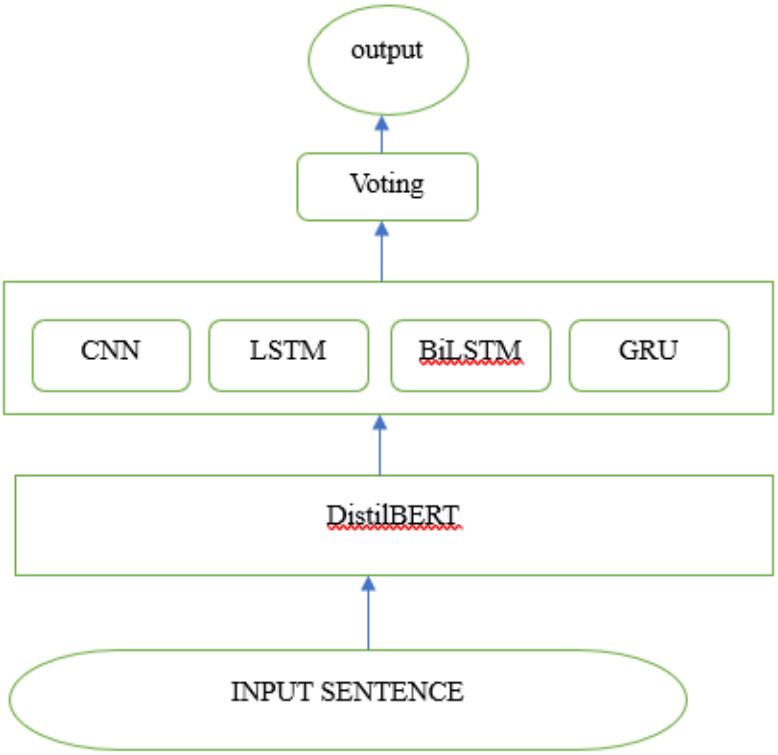

Figure 1. BERT-CNN-BiLSTM-GRU Architecture

3. METHOD

This section outlines the methodology adopted for text representation and its integration into deep learning frameworks for the analysis of suggestions within online reviews. The study investigates the effectiveness of employing BERT for textual representation in deep learning models. Before employing deep learning models for suggestion analysis, textual reviews are converted into numerical representations using BERT. The evaluation of deep learning models, including CNN, LSTM, LSTM-CNN, and LSTM-CNN, follows the analysis of textual representation in online reviews containing customer suggestions.

3.1. Bidirectional Encoder Representation from Transformer

The BERT architecture incorporates a multi-layer bidirectional transformer encoder, employing transformers exclusively up to the encoder phase. Work by [23], the transformer adopts an encoder-decoder architecture depicted on both sides of Figure 2, analyses contextual relationships among words in text. In the transformer architecture, multi-head attention is employed at each encoder and decoder layer. This mechanism, known as self-attention, maps each word to all others within the sentence, aiding the model in understanding contextual text representation. The multi-head attention mechanisms integrate self-attention to produce a more nuanced representation.

Custom BERT configurations tailored to specific tasks enable the representation of sentence pairs either as sequence of tokens or as single phrase. The creation of a representation for a particular token involves amalgamating relevant tokens, segments, and embeddings. In classification tasks, a unique token is assigned for the first word in the sequence, with the fully connected layer connected to the final encoder layer. Sentence or sentence pair classification is accomplished using a voting layer.

BERT analyzes contextual information both before and after each layer, offering a representation adept at incorporating information from both directions.

The implementation of the BERT model can be utilized in two ways: the feature-based approach and the fine-tuning approach [21]. This study utilzes the feature-based approach, which involves analysing and representing text using pre-trained models, particularly those trained on large datasets. Typically, the BERT model comes in two sizes: BERTBASE and BERTLARGE [20].

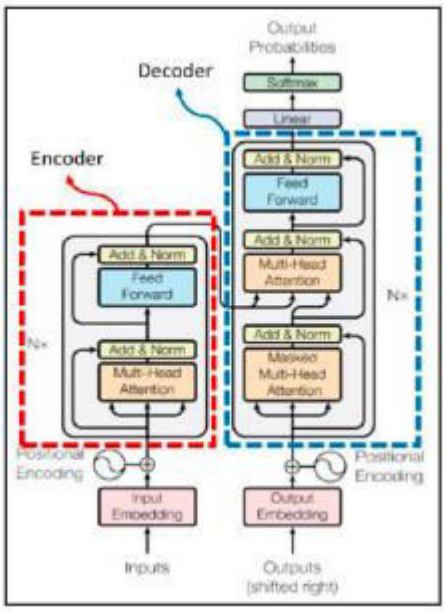

Figure 2. Transformer Architecture

3.2. Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed for processing structured grid-like data, such as images and text. In text processing, CNNs use filters to convolve over sequences of words or characters, extracting local patterns and features which can capture hierarchical representations of text data. By leveraging these hierarchical features, CNNs excel in tasks such as sentiment analysis, document classification, and text generation.[24]

3.3. Long-Short Term Memory

Long Short-Term Memory (LSTM) networks are recurrent neural network (RNN) architecture known for their capacity to capture long-distance relationships in sequential data, making them particularly suitable for sentiment classification tasks. The difference between conventional RNNs and LSTMs is that, LSTMs incorporate memory cells and gates that empower them to retain and selectively update information over time, effectively tackling the issue of vanishing gradients. In sentiment classification, LSTMs adeptly acquire contextual dependencies and subtleties in text, enabling them to discern nuanced emotional expressions and precisely categorize sentiment across varied text lengths. Their proficiency in modelling sequential information across extensive contexts renders LSTMs a potent choice for sentiment analysis in various applications, spanning from tracking social media sentiment to analysing product reviews [25].

4. RESULTS AND DISCUSSION

The process initiates with datasets as input where reviews from Electronics and Hotel datasets undergoes data pre-processing steps such as removing special characters, lower casing etc. Following this, BERT is utilized to represent the data, and the resultant representations serve as inputs for suggestion analysis using CNN, LSTM, LSTM-CNN, and CNN-BiLSTM-GRU models. The efficiency of these models is then estimated using a confusion matrix, and metrics such as accuracy, precision, and recall. The evaluation aims to determine the efficacy of the BERT-based deep learning model in achieving optimal performance. Figure 9 illustrates the sequential phases of the training process.

4.1. Data Description

The dataset supplied by the SemEval-2019 organizers includes reviews from the Electronics and TripAdvisor domains. Specifics regarding the dataset are outlined in Table 1.

Table 1. Datasets Description

Subtask	A	B
Domain	Electronics	Hotel
Training	8200	0

Trial	592	808
Test	833	824

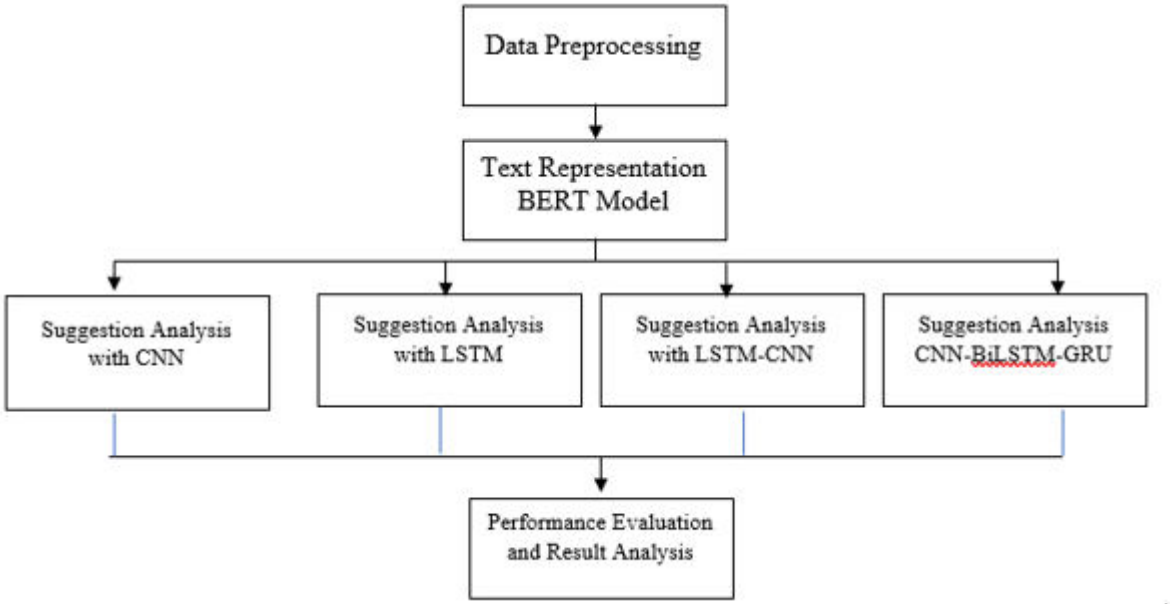

Figure 9. General Phases for Suggestion Analysis

4.2. Data Pre-processing

The aim of data pre-processing is to improve the model's understanding of data representation and enhance overall performance.

4.3. Text Representation Using BERT Model

Prior to processing of textual data by BERT, it must be formatted into a representation that BERT can accept. The first step in this process is Tokenization, where each word in a given sentence is converted into a sequence of tokens. BERT utilizes the WordPiece for tokenization, augmented with specific [CLS] and [SEP] tokens placed at the sentence's beginning and end.

Following tokenization, the next step involves ensuring uniform input sentence's length through padding. Padding entails the adding of a special token, denoted as [PAD], until the sentence's length reaches 128 tokens. Next step is transforming tokens into integers to facilitate input reading by the BERT model. This stage employs the vocab of the WordPiece model, consisting of 30,522 pairs of tokens and their corresponding unique integers. Numericizing generates a tokenID, which BERT utilizes in converting every input token into a numeric representation vector.

4.4. Suggestion Analysis Steps

The dataset in this work is divided into training, eval and testing subsets, as shown in table 1, respectively. The training subset is employed for model

training, followed by evaluating the model using eval and testing subset to test the model. After representing text data using BERT , both training and testing subsets are fed as inputs for the CNN, LSTM, LSTM-CNN, and CNN-BiLSTM-GRU models.

The model is trained over 5 epochs, with 5 iterations utilizing the training dataset during the learning phase. Adam optimization strategy is adopted is with a batch size of 32 and a learning rate of 1×10-3. Further, binary cross-entropy is utilized as loss function. The rectified linear unit (ReLU) is used as activation function in CNN, and LSTM employs the sigmoid activation function in each gate and tanh in the recurrent output. Also, the voting is applied in the fully-connected layer.

4.5. Model Performance Analysis

Upon constructing the model with training data, the subsequent phase involves evaluating the model's performance using testing data. Across the dataset and the four suggestion analysis models employed, four metrics are considered. Table 2 displays the results of Suggestion classification utilizing BERT as Text representation in Subtask A. Table 3 displays the results of suggestion analysis evaluation utilizing BERT as text representation in Subtask B. Leveraging the 8500 Electronics review dataset for subtask A and 800 hotel reviews as test dataset for Subtask B. Performance evaluation of BERT-CNN, BERT- LSTM, BERT-LSTM-CNN, Analysis CNN-BiLSTM-GRUis conducted using a confusion matrix comprising F score, precision, and recall metrics.

Table 2. Performance Comparison Among Different Models on Electronics Reviews

Model	Accuracy	Precision	Recall
BERT-CNN	0.62	0.6	0.6
BERT-LSTM	0.74	0.71	0.76
BERT-LSTM-CNN	0.73	0.72	0.73
BERT-CNN-BiLSTM-GRU	0.77	0.78	0.79

Table 3. Performance Comparison Among Different Models on Hotel Reviews

Model	Accuracy	Precision	Recall
BERT-CNN	0.76	0.77	0.74
BERT-LSTM	0.74	0.71	0.7
BERT-LSTM-CNN	0.8	0.83	0.8
BERT-CNN-BiLSTM-GRU	0.85	0.86	0.85

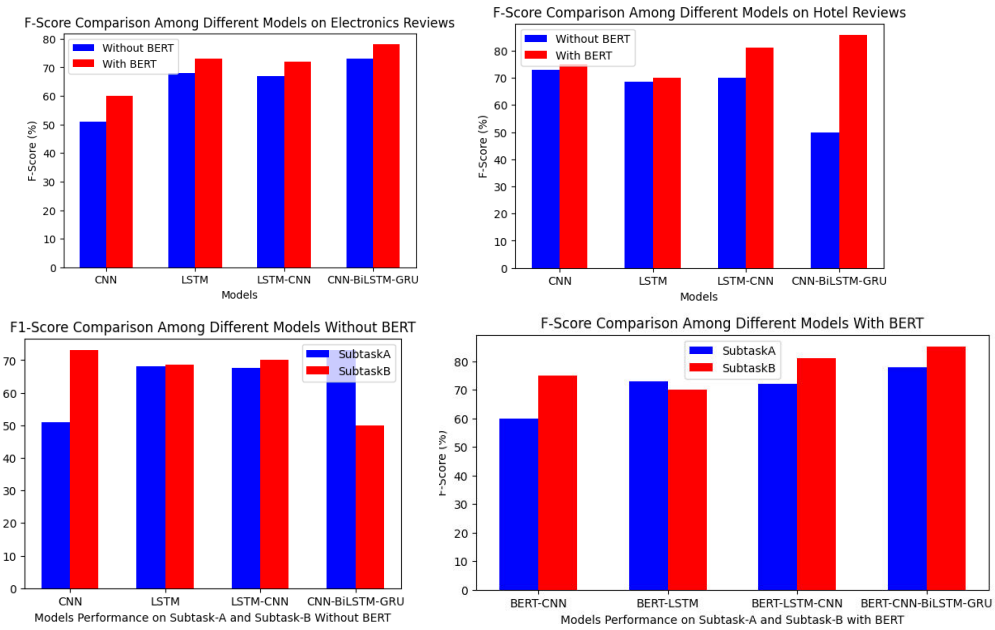

Figure 10. Performance Comparisons of Models with BERT and Without BERT

4.6. Discussion

The evaluations of the deep learning model based on BERT for suggestion analysis concerning online customer suggestions regarding the Electronics and Hotel reviews reveals that the BERT- CNN-BiLSTM-GRU model exhibits the highest performance. The BERT- CNN-BiLSTM-GRU model achieves an F1 score of 0.78% on Subtask A and 0.85.7% of F score on Subtask B. Following, the hybrid BERT-LSTM-CNN deep learning model demonstrates an F1 score of 0.72% on Subtask A and 0.81% on Subtask B. The BERT- LSTM model records an F1 score of 0.73% on Subtask A and 0.70% on Subtask- B. Subsequently, the BERT-CNN model presents F1 score of 0.60% on Subtask-A and 75% on Subtask B

5.CONCLUSION

The performance analysis of the BERT-hybrid model was conducted to analyse the suggestions expressed in online reviews of Electronics domain and Trip Advisor domain. The evaluation of methods with various text representation employed with CNN, LSTM, LSTM-CNN, CNN-BiLSTM-GRU and the same models in conjunction with BERT is made. Through the examination of the Electronics and TripAdvisor datasets, it is evident that the deep learning models enhance the performance of suggestion analysis conducted by the BERT model. A noteworthy performance is the CNN-BiLSTM-GRU model, which achieves remarkable improvements with BERT, with an accuracy of 0.77%, precision of 0.78%, and recall of 0.79% on Subtask A. Also, CNN-BiLSTM-GRU model,

which achieves remarkable improvements in BERT model performance, with an accuracy of 0.85%, precision of 0.86%, and recall of 0.85% on Subtask B.

REFERENCES

[1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics, 2002, pp. 79–86.

[2] B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool, 2012.

[3] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: a survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8, no. 4, pp. 2018, doi: 10.1002/widm.1253.

[4] A. Viswanathan et al., "Suggestion mining from customer reviews," in AMCIS, 2011.

[5] C. Brun and C. Hagège, "Suggestion Mining: Detecting Suggestions for Improvement in Users' Comments," Research in Computing Science, vol. 70, no. 1, pp. 199–209, 2013, doi: 10.13053/rcs-70-1-15.

[6] S. Negi and P. Buitelaar, "Towards the extraction of customer-to-customer suggestions from reviews," in Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2015, doi: 10.18653/v1/d15-1258.

[7] A. B. Goldberg et al., "May All Your Wishes Come True: A Study of Wishes and How to Recognize Them," in The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2009, pp. 263–271.

[8] A. F. Wicaksono and S.-H. Myaeng, "Mining advices from weblogs," in Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12, ACM, 2012, pp. 2347–2350.

[9] S. Moghaddam, "Beyond sentiment analysis: mining defects and improvements from customer feedback," in European Conference on Information Retrieval, Springer, 2015, pp. 400–410.

[10] S. Negi and P. Buitelaar, "Suggestion Mining From Opinionated Text," in Sentiment Analysis in Social Networks, Elsevier Inc., 2017, doi: 10.1016/B978-0-12-804412-4.00008-5.

[11] S. Negi and P. Buitelaar, "Inducing Distant Supervision in Suggestion Mining through Part-of-Speech Embeddings," 2017, arXiv:1709.07403.

[12] S. Negi, M. de Rijke, and P. Buitelaar, "Open domain suggestion mining: Problem definition and datasets," 2018, arXiv:1806.02179.

[13] S. Negi, T. Daudert, and P. Buitelaar, "Semeval-2019 task 9: Suggestion mining from online reviews and forums," in Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019).

[14] T. N. Fatyanosa et al., "DBMS-KU at SemEval-2019 Task 9: Exploring Machine Learning Approaches in Classifying Text as Suggestion or Non-Suggestion."

[15] I. Markov and E. Villemonte De la Clergerie, "Inria at semeval-2019 task 9: Suggestion mining using svm with handcrafted features," in Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), 2019.

[16] R. A. Potamias, A. Neofytou, and G. Siolas, "Ntua-islab at semeval2019 task 9: Mining suggestions in the wild," in Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), Minneapolis, Minnesota, 2019.

[17] Y. Ding, X. Zhou, and X. Zhang, "Ynu dyx at semeval-2019 task 9: A stacked bilstm model for suggestion mining classific," in Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), 2019.

[18] S. Pecar et al., "NL-FIIT at SemEval-2019 Task 9: Neural Model Ensemble for Suggestion Mining," 2019, arXiv:1904.02981.

[19] R. S et al., "Ssn-sparks at semeval2019 task 9: Mining suggestions from online reviews using deep learning techniques on augmented data," in Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), 2019.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[21] A. Vaswani et al., "Attention is all you need," in Advances in neural information processing systems, vol. 30, Jun. 2017.

[22] Z. Gao et al., "Target-dependent sentiment classification with BERT," IEEE Access, vol. 7, pp. 154290–154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[23] C. C. Aggarwal, "Training deep neural networks," in Neural Networks and Deep Learning, Springer International Publishing, 2018, pp. 105–167.

[24] "Understanding LSTM networks," colah's blog, 2015, http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed Oct. 25, 2023).

[25] P. Anki et al., "High accuracy conversational AI chatbot using deep recurrent neural networks based on BiLSTM model," in 2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020, Nov. 2020, pp. 382–387, doi: 10.1109/ICOIACT50329.2020.9332074.