IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 May 2024, accepted 30 June 2024, date of publication 11 July 2024, date of current version 2 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3426604

RESEARCH ARTICLE

Reciprocating Encoder Portrayal From Reliable Transformer Dependent Bidirectional Long Short-Term Memory for Question and Answering Text Classification

M. SUGUNA AND K. S. SAKUNTHALA PRABHA
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India
Corresponding author: M. Suguna (suguna.m@vit.ac.in)

This work was supported by the Vellore Institute of Technology, Chennai, India.

ABSTRACT Diversity in use of Question and Answering (Q/A) is evolving as a popular application in the area of Natural Language Processing (NLP). The alive unsupervised word embedding approaches are efficient to collect Latent-Semantic data on number of tasks. But certain methods are still unable to tackle issues such as polysemous-unaware with task-unaware phenomena in NLP tasks. GloVe understands word embedding by availing information statistics from word co-occurrence matrices. Nevertheless, word-pairs in the matrices are taken from a pre-established window of local context, which may result in constrained word-pairs and also probably semantic inappropriate word-pairs. SemGloVe employed in this paper, refines semantic co-occurrences from BERT into static GloVe word-embedding with Bidirectional-Long-Short-Term-Memory (BERT- Bi-LSTM) model for text categorization in Q/A. This method utilizes the CR23K and CR1000k datasets for the effective text classification of NLP. The proposed model, with SemGloVe Embedding on BERT combined with Bi-LSTM, produced better results on metrics like accuracy, precision, recall, and F1 Score as 0.92, 0.79, 0.85, and 0.73, respectively, when compared to existing methods of Text2GraphQL, GPT-2, BERT and SPARQL. The BERT model with Bi-LSTM is better in every way for responding to different kinds of questions.

INDEX TERMS Bidirectional encoder representations from transformer, natural language processing, question and answering, SemGloVe.

I. INTRODUCTION

The span of around twenty years has a notable increase globally in the frequency of internet searches related to specific topics. Grading the search outcomes and time usage on dragging millions of web-pages are most exciting research subjects [1]. Open Domain Question Answering (ODQA) is a recent subfield that has garnered increased attention and aims to identify answers within a large volume of diverse topic documents [2], [3]. The ODQA ability can be enlarged for evolving an effective backup system to overcome the problem of retrieving pertinent answer to the specific question from the acquisition of data [4]. The chatbot program imitates human communication or chat, invoking Artificial Intelligence (AI) to take part in a dialog with human administered Natural Language Processing (NLP). Now, a Chatbot can be executed on number of functionalities such as mobile applications or websites by employing communication applications [5]. The Chatbots are developed day by day in various fields such as education, healthcare, e-commerce, marketing, customer services and so on. In all of those fields, the Chatbots have demonstrated to be more prompt in number of contexts to automate tasks and to enhance user activities [6], [7].

Natural language processing has two widely recognized architectures: BERT and Bi-LSTM. Both have unique

The associate editor coordinating the review of this manuscript and approving it for publication was Agostino Forestiero.

advantages for comprehending textual material and obtaining contextual information. In 2018, Google transformed natural language processing by introducing BERT, a deep bidirectional transformer model pre-trained on extensive text data. BERT uses a transformer design, which allows it to gather context information from neighbouring words in a phrase in both directions, unlike earlier versions. BERT effectively learns nuanced contextual meanings of words through pre-training on large corpora and adapting to specific tasks. This leads to notable enhancements across various NLP tasks, such as named entity recognition, sentiment analysis, and question answering.

Conversely, Bi-LSTM is a type of RNN that combines the ability to process in both directions, meaning that when it processes sequential input, it may record context information from the past as well as the future. Through analyzing sequences of inputs in forward as well as backward directions, Bi-LSTM overcomes the drawbacks of conventional LSTM networks and improves its ability to collect context information and long-range relationships. Because of this, Bi-LSTM is especially well suited for sequential data-related tasks including machine translation, text classification, and sequence labeling.

With the rapid growth of Information-Retrieval (IR) as well as Reading-Comprehension (RC) approaches, the most ODQA structure embrace a Retriever-Reader-pipeline [8]. Retriever task can be generally decomposed into various subtasks: 1. Retrieve applicable documents, 2. Extracts answer applicants from retrieved documents and 3. Re- grade answer applicants to find the right answer [9], [10]. Regarding the first task, the NLP approach named SemGloVe with Elasticsearch is used. With respect to second task, it contains language model applications [11], [12]. The language-models are pre-trained on wide-reaching text collections with existing method to obtain better performance in number of NLP operations, particularly those fine-tuned for the provided purposes. The number of existing Q/A approaches often depend on the pre-trained language models named BERT-based Q/A approach that is trained using Natural Questions. As in third task, the existing methods utilize Neural Network (NN) to re-grade every extracted applicant simply based on input question as well as context over applicant answer [13], [14], [15].

Through improving word meanings and resolving semantic ambiguity and polysemy, the ''Reciprocating Encoder'' improves text comprehension. It enhances text analysis by integrating memory skills and semantic comprehension with BERT and Bi-LSTM. Better performance metrics show that this strategy provides more accurate answers than previous approaches when it comes to answering questions. As the suggested study shows, this strategy enhances model performance and produces better text categorization outcomes than other approaches like ''Text2GraphQL, BERT, and SPARQL.''

In comparison with present techniques, the suggested model—which integrates SemGloVe, BERT, and Bi-LSTM—presents several theoretical advances. SemGloVe allows for a more accurate representation of word meanings by refining semantic co-occurrences, in contrast to conventional word embedding methods like GloVe. The model attains improved text comprehension and classification accuracy by fusing Bi-LSTM's contextual learning with BERT's capacity to collect syntactic and semantic information. Beyond approaches that only use word embeddings or pre-trained language models like BERT, this combination of SemGloVe, BERT, and Bi-LSTM produces an effective framework for NLP problems, giving theoretical advances in semantic representation and contextual comprehension.

The primary contributions of this research are as follows; SemGloVe, refines static GloVe word embedding using semantic co-occurrences culled from BERT. Particularly, the model presented extract co-occurrence statistics based on multi-head attention weights of BERT. This technique derives word-pairs that are constrained by the local window assumption and determines co-occurrence weights by straightway capturing semantic distance between word-pairs into direct consideration.

Semantically significant sentence embedding can be compared using cosine similarity and are best produced using SemGloVe and BERT [2].The suggested system incorporates a BiLSTM neural network with a fully connected neural network architecture.

• The structure of the design consists of a feature extractor called BiLSTM-Attention that generates more meaningful vectors, a BERT embedding layer that transforms words to their BERT embedding, and a similarity score that ranges from 0 to 1.

The paper is organized as that: Section II discusses Literature survey. Section III presents the proposed method. The classification of results and discussion are included in Section IV and then Section V concludes with conclusion.

II. LITERATURE SURVEY

Ni et al. [16] developed the encoder-decoder pipeline frame of language model by ''schema-utterance'' knowledge establishment system as well as Pointer Network by difficult computing systems. The Text2GraphQL task approach was majorly developed according to enhanced pipeline contained language model, pre-trained Adapter plug-in and Pointer Network. A whole language approach was considered as smarter encoder of entire pipeline. However, the model had led to an insufficient data due to its expensive annotation.

Qiu et al. [17] implemented a robust end-to-end method for effectively retrieving the queries referred to Mineral Exploration (ME) terms. An automated process was initially developed for building the Q/A databases according to the names and definitions in ME existence. The BERT was trained to test the answers developed from input of user question. Eventually, building a prototype chatbot model on WeChat platform estimates the mechanisms introduced. However, the pre-trained BERT data source approach was developed from the publicly available dataset, which directly

caused the limited amount of known domain knowledge that impacts Q /A performance.

Bayer et al. [18] developed and estimated the text generation approach appropriate to enhance the classifier performance for long and short texts. Initially, the development was enriched by significant fine-tuning as well as prefix extension. Then, the document embedding filter was applied for the instances that were not integrated with the actual classes are excluded and the suggested approach was based on GPT-2 approach. The GPT-2 utilized filter mechanism based on the classifier, trained on class data and this approach has significantly minimized the data augmentation diversity. However, the approach had a lack of factual knowledge as well as tendency to generate biased or offensive text.

Amin and Nadeem [19] developed a medical specialty identification approach from the text-based question of patients according to pre-trained BERT. The database has combined medical text-based questions as well as labeled experts which graze from website for service of medical Q/A. The approach was fine-tuned and identified the needed medical aspect labels among 27 labels from medical question texts. However, the forecast approach performance was crucially depending on quality of a data.

Devlin et al. [20] established a high-quality medical Q/A knowledge graph according to the professional knowledge in medical Q/A research by processes named extraction as well as knowledge fusion. According to that approach, the difficult domain words as well as questioning words were applied for ideal matching rules. The suggested approach was aimed to traverse the enhancement of medical Q/A approaches of accuracy as well as efficiency. The data was stored in Neo4j graph dataset by utilized rule and string-based combined approaches to develop a domain lexicon to classify as well as query questions. However, the suggested approach does not achieve the greater problem coverage because of its limited scale as well as final accuracy of integration approach cannot reach the better results due to close relation to the data source quality.

Luo [21] initially designed and developed a framework to provide an interactive user interface. Then, implemented a Machine Learning (ML) approach according to the intention of classification as well as Natural Language considerate to recognize user plan as well as introduced the SPARQL queries. The suggested method particularly processes the new social network dataset as well as applied it to the traditional knowledge bases to enhance the Chabot capabilities by considering the analytical queries. But the method cannot make the central hidden state to catch the ample textual linguistics.

Kasthuri and Balaji [22] implemented inter- actional education-oriented Chabot, which could answer the queries applied by the learner. A suggested approach was with the Deep Learning (DL) approach for the development of that educational Chabot. This framework was developed to obtain the immediate responses on behalf of wait-to somebody to-response. This approach has greater capability to address the student uncertainty without the requirement of the human support.

The examination of the related works provides valuable perspectives on many techniques and approaches employed in the fields of text categorization and QA. A variety of approaches, including encoder-decoder pipelines and improving pre-trained models like BERT for QA tasks, have been investigated in earlier research. These methods have addressed issues including the expense of data annotation, the constraints of domain-specific knowledge, the improvement of classifier performance, the identification of medical specialties, and the creation of knowledge graphs in the medical domain. These studies are useful, but they have drawbacks as well, including a lack of data, a reliance on poor data quality, and a propensity to produce inflammatory or prejudiced writing. The proposed approach, 'Reciprocating Encoder portrayal from reliable Transformer dependent Bidirectional Long Short-Term Memory for Question and Answering Text Classification' integrates SemGloVe with BERT and BiLSTM, improving word embeddings, semantic relationships, and model performance, offering potential advancements in QA text classification tasks.

III. PROPOSED METHODOLOGY

This paper proposes the use of a BERT – Bi-LSTM based Question Answering text classification approach in NLP. The proposed work comprises of four main stages: collection of datasets, Pre-processing using SemGlove with BERT and Classification. Figure 1. depicts the workflow of the suggested method.

A. DATASET

In this research, the proposed method utilized two-text generation dataset for the sentiment analysis named CR23K and CR100K [23]. These datasets are from the education domain and which consists of three classes such as positive, negative as well as neutral. These datasets are greatly imbalanced with greater number of reviews for the positive label class. The sentimental dispensation depicts greatly the imbalanced nature of the datasets by greater liability towards positive sentiment labels. Table 1 provides the description of both datasets by course reviews and class labels. Due to their availability of a variety of sentiment classes and their applicability to NLP applications, the CR23K and CR1000k datasets were selected. These datasets improve the study's reliability and applicability by providing a reliable testing environment for assessing the suggested model's performance in text classification. The study splits datasets CR23K and CR100K into train, and test sets, allocating 70%, and 30% respectively.

B. DATA PREPROCESSING

Word embedding: In the field of natural language processing, the primary thrust on study has traditionally been on the mathematical representation of text. Discrete representations, conveyed by One Hot Code, and distributional representations, embodied in Global Vectors (GloVe), are the two main

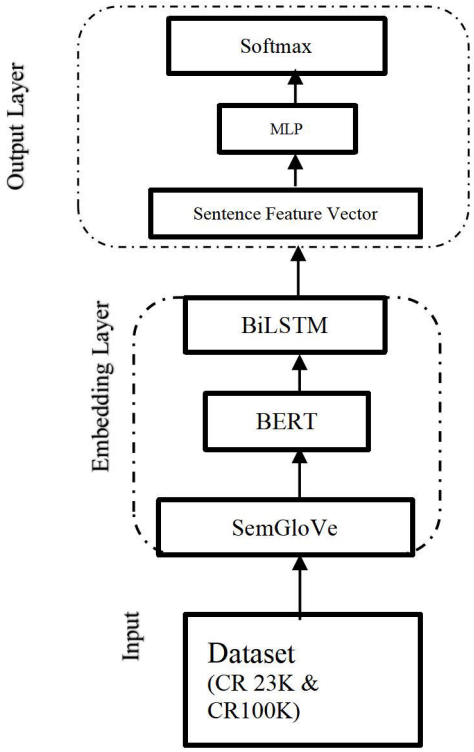

FIGURE 1. Block diagram of the suggested method.

TABLE 1. Dataset description.

Dataset	Course Reviews	Class Labels	Description
CR23K	21940 from Coursera	Positive = 84.2% Negative = 10.6% Neutral = 5.2%	This dataset is manually labelled with three labels.
CR100K	107016 from Kaggle	Positive = 90.9% Negative = 4.7% Neutral = 4.4%	A conversion scheme is utilized to convert ratings into sentiment labels.

word embedding representation techniques used today [6]. The inability of one-hot to distinguish between words is by far its most glaring flaw. Word2vec and latent semantic analysis features are integrated into the SemGloVe model, which has its foundation on statistical theory [7]. Furthermore, it has robust scalability and streamlines the parameter training process, making the model more appropriate for large-scale corpus data sets.

SemGloVe proceeds by obtaining the global word-word co-occurrence counts matrix X given a training corpus. Its entries, X_{ij}, indicate the total number of times a word $w_j \in V$ occurs in the context of a word $w_i \in V$, where V is the training corpus's word vocabulary. In terms of the separation between w_i and w_j for the global word-word co-occurrence count, SemGloVe specifies X_{ij} as:

$$X_{ij} = \sum_{w_j \in C(wi)} \text{dis}\left(w_i, w_j\right)$$

$$= \sum_{w_j \in C(wi)} \frac{1}{\left[P_j - P_i\right]} \quad (1)$$

where p_j and p_i are positions in context. Intuitively, words close to w_i get larger weights.

SemGloVe has two upsides over GloVe: first, it can identify semantically meaningful word pairs that GloVe is unable to determine, second, it can create global word-word co-occurrence counts that are far more precise than GloVe.

In accordance with the context, BERT can create distinct embedding for the same word (e.g. the neighboring words). For example, the term "bank" might signify various notions across multiple linguistic situations (such as a financial institution or a piece of land next to a river) or serve distinct purposes (such as a noun or verb). BERT displayed state-of-the-art performance [24], henceforth encoding syntactic and semantic information about the source text [25].

Co-occurrences of Semantics from Multi-Head Self-Attention; The multi-head self-attention weights of BERT evaluates semantic relationships of tokens in contrast to heuristic position-based distance function of GloVe. To be more precise, the BERT self-attention weights determine the word-to-word semantic distance, given a word sequence $W = \{w_1, \ldots, w_K\}$, a window size S, and a pre-trained BERT model. First convert the original BPE-to-BPE attention weights to word-to-word attention weights, since BERT segments, words into BPE tokens using word parts, or byte-pair encodings Gan et al. [26].

$$T = \sum_{i=1}^{N} \sum_{j=1}^{M} AT_{ij} \quad (2)$$

Then, word-to-word attention weight matrix AW € $R^{K \times K}$ is generated by averaging BPE-to-BPE attention weights.

For w_j within local window context of word $wi,_j$€[i – S, i + S] ∩ j ≠ i, attention weight from word w_i to w_j is denoted as AW_{ij} following:

$$AW_{ij} = \frac{1}{mxn} \sum_{k=s_1}^{s_m} \sum_{l=t_1}^{t_n} AT(k, l) \quad (3)$$

where m and n are number of sub-words for w_i and w_j, respectively, and $AT(k, i)$ indicates attention weight from BPE token t_k to t_l. AW_{ij} is sorted descendingly and top-S words as w_i's context words C (w_i) are chosen in order to exclude semantically meaningless terms. Lastly, the Division distance function determines separation between target word w_i and context word w_j:

$$dis\left(w_i, w_j\right) = \frac{AW_{ij}}{AW_{i1}} \quad (4)$$

IV. CLASSIFICATION

After the extracted word features, classification algorithms are used for classification of text data. An extracted feature output is further proceeded to the classification of information retrieval. In this method, the BERT – Bi-LSTM approach is used to classify the text data and it is described in the following section.

A. BERT

Recently, the BERT is utilized at the search engines (like Google) to optimize the explanation of the user's search queries. The BERT exceeds at various functions that make this possible including sequence-to-sequence based language development tasks like Question and Answering. The BERT has been trained on a large corpus, developed to solve the smaller problems and more defined NLP tasks. To utilize the BERT approach, extract the data according to reading conception. This method requires input Question and phrases as text1, text2 for BERT approach accordingly and eventually extracts the respective answers from phrases. Simultaneously, input to BERT needs to extend some definite flags. The depiction vector acquired through BERT could be utilized for following classification task. The [SEP] mark has to be located among sentences to divide two input sentences, and [MASK] is employed to wrap few words in sentences to avail [MASK] vector output on BERT approach that identifies a word.

BERT is majorly classified into three modules such as embedding, transformer and pre-trained fine-tuning approach. The embedding approach consists of three portions such as token, segment and position embedding, as well as output vector of whole embedding approach is a total of those vectors. As a fundamental approach of BERT, transformer controls encoding part of transformer framework. The BERT consists of six encoder blocks, which are stacked with each other to develop an absolute encoder. Every block consists of two layers such as multi-head self-attention and Feed forward Fully Connected Layer (FFCN). The Self-attention designs the set of query vector Q as well as key-value pairs (k, v) to output. An outcome is estimated as weighted_sum values, where weight allocated to every parameter is estimated through correlation among query as well as alike key function. Self-attention implementation authorizes the model to learn long-term dependencies avoiding greater usage of computing resources and which is formulated in equation (4) as follows:

$$Attention = softmax\left(\frac{QK^T}{\sqrt{d}_k}\right)V \tag{5}$$

where, Q – query vector; (K, V) – set of key-value pairs; d_k – dimension of (K, V). A multi-head self-attention layer initially depicts Q, K and V by parameter matrix, after employing self-attention, and eventually connects the outcomes into FCN and which is formulated in equation (5) and (6) as follows:

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{6}$$

$$Multihead\ (Q, K, V)\ Concat(head_1, \ldots, head_h)W^O \tag{7}$$

With them, $W_i^Q \in R^{d_{model} \times d_k}$, $W_i^K \in R^{d_{model} \times d_k}$, $W_i^V \in R^{d_{model} \times d_k}$ and $W^O \in R^{hd_{v \times d_{model}}}$. The individual multi-head attention mechanism is not adequate to extract optimal features. FFCN is extended to every encoder block and is controlled by two linear transformations with ReLU activation function in the middle and is expressed in equation (7) as;

$$FFN\ (x) = max\ (0, xW_1 + b_1)\ W_2 + b_2 \tag{8}$$

The latest pre-fine-tuning method uses the highest expectation to determine a scale for answers and to locate the starting and ending points. It utilizes the linear as well as Softmax approach and hence input is a dependent BERT word embedding. For i^{th} token of phrase, the final layer of BERT encodes it as T_i and after utilizing the vector S and its internal product to estimate the initial position score. Every token of paragraph has arbitrary starting, then this approach utilizes the Softmax activation function to change probably and finally selects the greatest probability as the starting of the answer and which is formulated in equation (8) as:

$$P_i = \frac{e^{S.T_i}}{\sum_j e^{ST_j}} \tag{9}$$

Correspondingly, there is a vector T, which can be utilized to estimate where an answer completes. The pre-trained approach can set up training such that a model is quick, thus obtaining required effect. Simultaneously, pre-trained model plays a significant role in enhancing accuracy as well as reliability of the model and language depiction is greatly dependent on quality together with size of pre-trained aggregation. According to BERT being made to the basis of general models, further training the pre-trained BERT approach generates the rule text databases.

B. Bi-LSTM

The Bi-LSTM approach is constituted of forward and backward LSTM in which the data can be worked in both the direction. The backward direction procedure collects the hidden features with data pattern, which is basically eliminated by LSTM. The forward hidden layer L_f, backward hidden layer L_b and output sequence $GHI_o(t)$ is utilized to update the network. The network updates iteratively backward from T to 1 and forward as 1 to T. The updated network parameters can be expressed in equation (9) to (11) as follows:

$$L_f = \sigma\left(W_1 GHI_i(t) + W_2 L_{f-1} + b_{L_f}\right) \tag{10}$$

$$L_b = \sigma\left(W_3 GHI_i(t) + W_5 L_{b-1} + b_{L_b}\right) \tag{11}$$

$$GHI_o = W_4 L_f + W_6 L + b_{GHI_o} \tag{12}$$

where, L_f, L_b and $GHI_o(t)$ – forward pass, backward pass and last output layer. W – Weight coefficients as well as b_{L_f}, b_{L_b} and b_{GHI_o} – biases.

BiLSTM involves replicating the network's first recurrent layer such that two layers are now side by side. The input sequence is then supplied to the first layer in its original form, and a reversed copy is supplied to the second layer. Every time step, the outputs from the two LSTMs are typically concatenated [16].

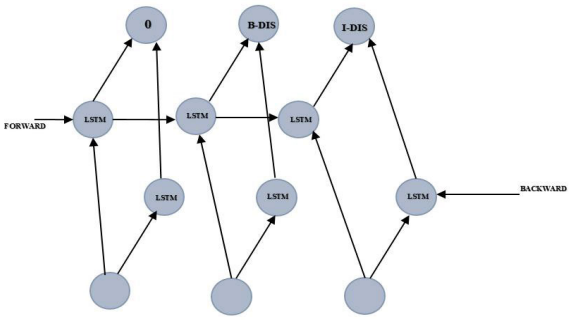

FIGURE 2. BiLSTM structure.

C. BERT – Bi-LSTM

Alternatively, the simple-weight integration approach is among BERT and Bi-LSTM. BERT- Bi-LSTM maintains BERT and Bi-LSTM as upstream and downstream. The BERT has a capability to learn statistical features of an adjacent word as well as Bi-LSTM is efficient in learning the environmental data then coin it with the augmentation of the non-computer language approach that has general desires on statistical features as well as certain meanings depending on the context. Hence, the BERT-Bi-LSTM has the potential to analyze the user query based on text classification.

The model carries an output $C \in R^n$ of final layer trained in BERT, and extend the weight $W_a \in R^{d_a \times n}$ as the Bi-LSTM input approach, which is mathematically expressed in equation (12) as follows;

$$a_i = g_1 \left(W_a C_i + b_a \right) \tag{13}$$

where, $1 \leq i \leq n$, n is feature vector dimension of sentence after the training of the BERT; b_a – vector of the bias; g_1 – activation function acquires the sigmoid function.

The standard LSTM estimates a hidden layer h from one direction, whereas Bi-LSTM estimates two hidden layer in various directions, and eventually integrates the outcomes from various directions to output. These output vectors can be formulated in equation (13) as follows;

$$v_i = \vec{h}_i + \overleftarrow{h}_i \tag{14}$$

where, the forward and backward hidden layer vectors as $\vec{h}$ and $\overleftarrow{h}$. Furthermore, the approach utilizes the *tanh* as activation function g_2 to estimate hidden layer, where, h is computed as expressed in equation (14);

$$h_i^d = g_2 \left(W_h^d a_i + U h_{i-1}^d + b_h^d \right) \tag{15}$$

where, $W_h^d \in R^{d_h \times d_a}$ – weight matrix of a_i, U – approximate hidden layer weight matrix output h^d at time $i-1$; $b_h^d \in R^{d_h}$ - bias vector approximate to d^{th} index.

Hidden layers h_i^d are integrated into the vector H, which is a last level feature vector and utilizes the ReLU as the activation function. An arbitrarily estimation of last text classification is expressed in equation (15) as follows;

$$P\left(y \mid H, W_s, b_s\right) = softmax(W_s H + b_s) \tag{16}$$

where, $W_s \in R^{|s| \times |l|}$ as well as $b_s \in R^{|l|}$ represents output layer parameters; $|l|$ – category numbers.

D. EXPERIMENTAL RESULTS

In this section, outcome and effectiveness of proposed method is estimated by various performance metrics like accuracy, precision, recall and F1-score that estimates model performance. The mathematical representation of these performance metrics are expressed in Equation (17 - 20),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

where, TP – True Positive;
TN – True Negative;
FP – False Positive;
FN – False Negative.

E. PERFORMANCE ANALYSIS

The study will evaluate the model's performance on various examination of diverse question types present in the dataset to be essential in improving the model's efficacy in diverse settings. Including feedback analysis questions in the proposed study facilitates a more profound comprehension of the model's capacity to interpret and tackle particular problems expressed by reviewers, including issues related to course structure, evaluation systems, or missing learning resources. The study focuses on feedback analysis to improve the model's overall efficacy in adapting user preferences and correcting any flaws in course offerings. This section shows the quantitative and qualitative analysis of the proposed BERT – Bi-LSTM using accuracy, precision, recall and F1-score are presented in Tables 2, 3 and 4. Table 3 illustrates

TABLE 2. Dataset review labels.

Review	Positive	Negative	Neutral
Easy to follow and includes a lot...	1	0	0
Really nice teacher!	1	0	0
I was disappointed because the...	0	1	0
Good content, but the course...	0	1	0
Very structured approach. Thank...	1	0	0
Program demystifies the evolving...	0	1	0
Very Basic course for learners...	0	0	1

TABLE 3. Performance analysis using pre-trained model.

Methods	Accuracy	Precision	Recall	F1-score
Text2GraphQL	0.76	0.70	0.63	0.61
SPARQL	0.81	0.72	0.65	0.63
XLNet	0.84	0.75	0.69	0.64
TextGCN	0.88	0.78	0.71	0.65
BERT	0.89	0.80	0.73	0.67

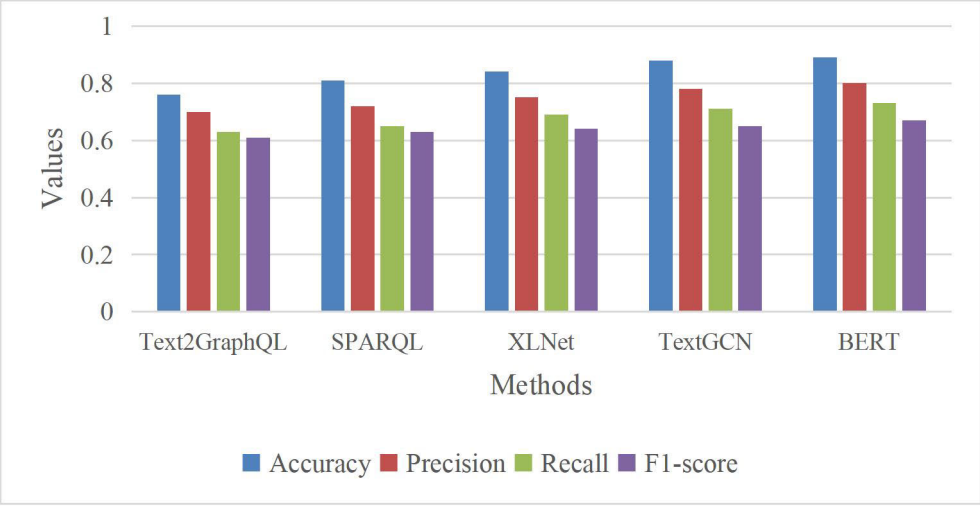

FIGURE 3. Graphical representation of proposed method using pre-trained models.

the performance of a proposed method utilizing pre-trained models. Table 4 depicts the performance of proposed method using Deep Learning (DL) algorithms. Table 5 depicts the performance of proposed method in classification.

An overview of the text distribution throughout a collection of reviews is shown in this table 2. Each row represents a review, and the columns show the number of reviews that are classified as neutral, positive, or negative depending on how they feel.

Table 3 and fig 3 represents the performance of proposed method with various pre-trained models. The performance of a proposed method is evaluated and matched with existing methods like Text2GraphQL, SPARQL, XLNet and TextGCN. The obtained results shows that the BERT model attains the accuracy of 0.89, precision of 0.80, recall of 0.73 and F1-score of 0.67 respectively which is better when compared to the existing methods.

Table 4 and fig 4 represents the performance of a proposed method with various DL approaches. The performance of the proposed method is evaluated and matched with the existing methods like Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Recurrent Neural Network (RNN) and LSTM. The acquired results shows that the Bi-LSTM model attains the accuracy of 0.90, precision of 0.79, recall of 0.74 and F1-score of 0.68 respectively which is better when compared to the existing methods.

Table 5 and fig 5 depicts the performance of proposed BERT-Bi-LSTM with various methods. The performance of a proposed method is estimated and matched with existing methods like BERT-CNN, BERT-MLP, BERT-RNN and BERT-LSTM. The obtained results shows that the BERT-Bi-LSTM model attains accuracy of 0.92, precision of 0.85, recall of 0.79 and F1-score of 0.73 respectively which is better when compared to the existing methods.

The dataset balancing procedure is demonstrated in the fig 6, which shows a comparison of the numbers of positive, negative, and neutral data before and after a balancing approach is used.

An accuracy comparison between two datasets, CR23K and CR100K, using various models is shown in the fig 7. Decision Tree, BLSTM with proposed BERT-BiLSTM are the three models that are contrasted. Every model's accuracy % is shown on the vertical axis. It is clear that when compared to the other models, the proposed BERT-BiLSTM model obtains a significantly greater accuracy, perhaps about 92%, on both datasets. This suggests that when it comes to effectively categorizing or predicting outcomes on these datasets, the proposed BERT-BiLSTM model performs better than the Decision Tree and BLSTM models.

The figure 8 provided is a histogram illustrating the distribution of text lengths across five labels. The x-axis

TABLE 4. Performance analysis of the proposed method using DL algorithms.

Methods	Accuracy	Precision	Recall	F1-score
CNN	0.77	0.69	0.68	0.61
MLP	0.82	0.72	0.69	0.63
RNN	0.85	0.76	0.71	0.65
LSTM	0.89	0.78	0.72	0.68
Bi-LSTM	0.90	0.79	0.74	0.68

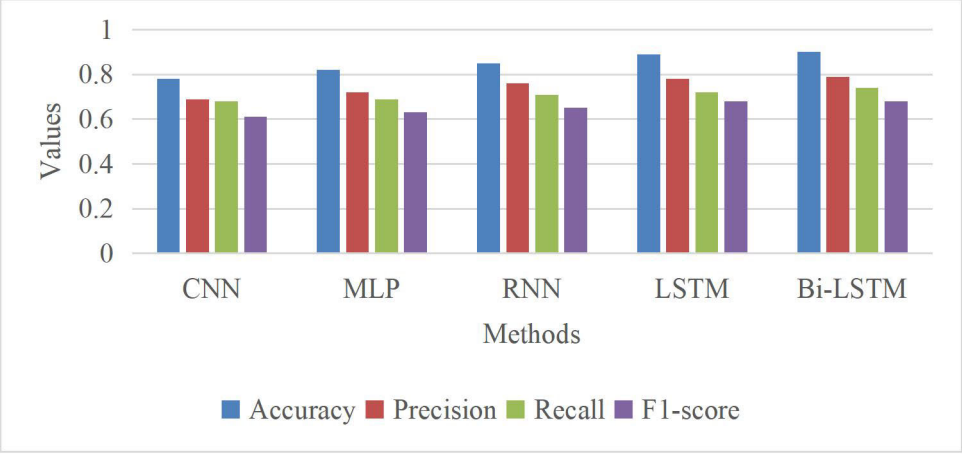

FIGURE 4. Graphical representation of proposed method using DL methods.

TABLE 5. Performance analysis of the proposed method.

Methods	Accuracy	Precision	Recall	F1-score
BERT-CNN	0.79	0.69	0.69	0.62
BERT-MLP	0.83	0.74	0.70	0.65
BERT-RNN	0.86	0.77	0.72	0.69
BERT-LSTM	0.90	0.81	0.75	0.70
BERT-Bi-LSTM	0.92	0.85	0.79	0.73

represents text length, while the y-axis shows frequency. Labels 1-5 depict different categories. Label 1 exhibits a higher frequency of shorter texts compared to the other labels, suggesting that texts categorized under Label 1 tend to be briefer.

The fig 9 "Positive Reviews" visualizes the density of positive reviews relative to "s_pos" scores, ranging from 0 to 1 on the x-axis. The y-axis represents the density of reviews within each score range. These variations help discern prevalent score ranges among positive reviews, providing insights into the distribution of text classification within the dataset.

The fig 10 provided depicts the distribution of reviews across different "s_pos" scores, ranging from 0 to 1 on the x-axis. Density, representing the number of reviews per score range, is depicted on the y-axis. Peaks in the histogram suggest concentrations of negative reviews at particular "s_pos" values, offering insights into the prevalence of negative review within the dataset.

The distribution of neutrality ratings for reviews, which range from 0 to 1, is shown in the fig 11. Scores mostly lie in the range of 0.8 to 0.9 and around 0.3. This suggests that a sizable percentage of evaluations convey neutral or balanced opinions as opposed to very positive or negative ones. Greater concentrations of evaluations with matching neutrality ratings are indicated by higher bars, highlighting the dataset's high prevalence of neutral comments.

The fig 12 illustrates the progression of accuracy metrics across nine experiments. The x-axis denotes the number of experiments, while the y-axis represents accuracy as a percentage. Two lines are depicted training accuracy and a testing accuracy. Both accuracies demonstrate an upward trend after the initial experiment, signifying improved performance

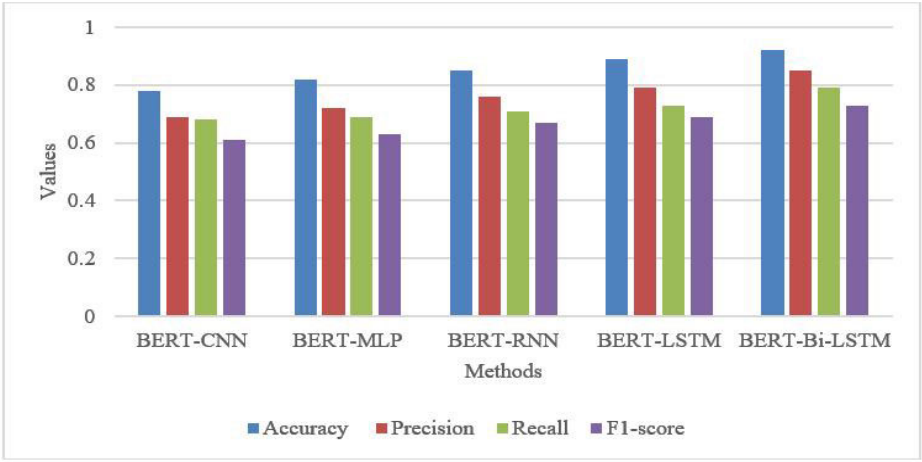

FIGURE 5. Graphical representation of proposed method in of classification performance.

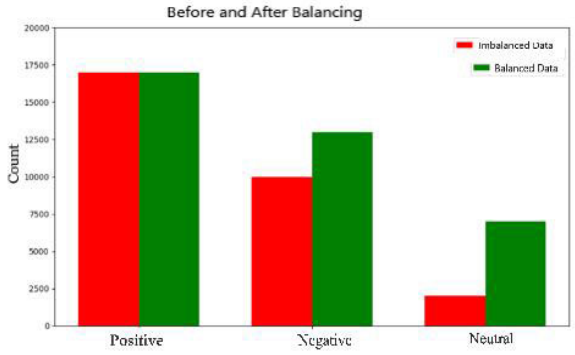

FIGURE 6. Before and after balancing dataset.

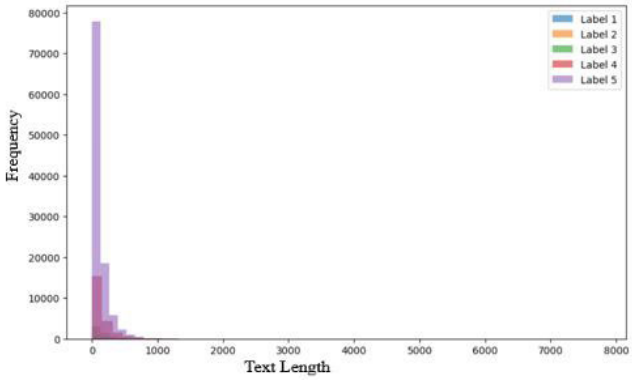

FIGURE 8. Distribution of text lengths by labels.

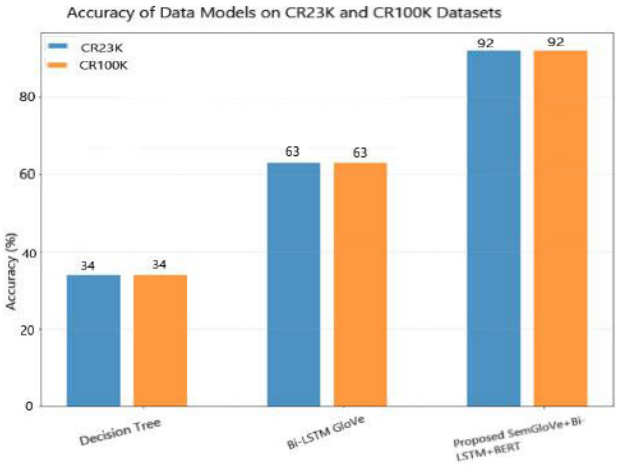

FIGURE 7. Accuracy of different models on CR23K and CR100K dataset.

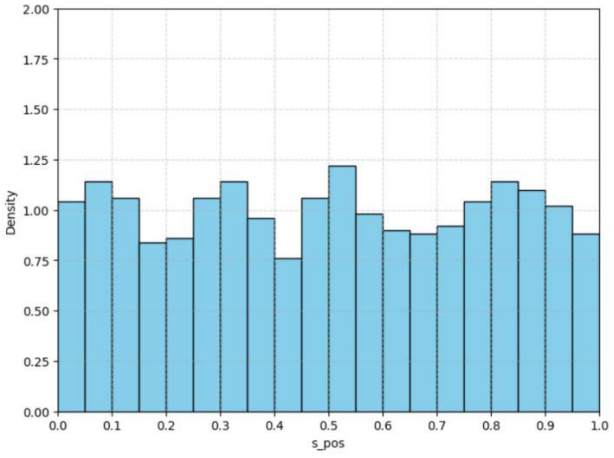

FIGURE 9. Positive reviews.

over successive trials. This suggests that the model effectively learns from the training data and performs adequately on unseen data.

The fig 13 illustrates the progression of accuracy metrics across nine experiments. The x-axis denotes the number of experiments, while the y-axis represents accuracy as a percentage. Two lines are depicted training accuracy and a range

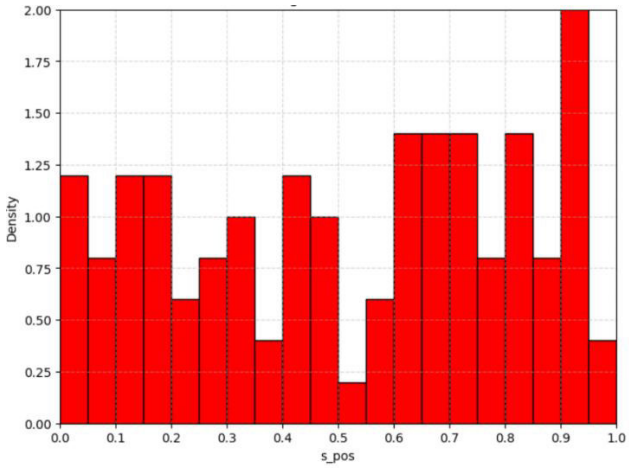

FIGURE 10. Negative reviews.

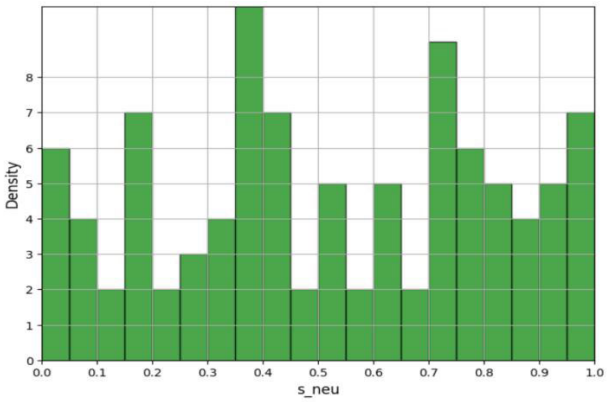

FIGURE 11. Neutral reviews.

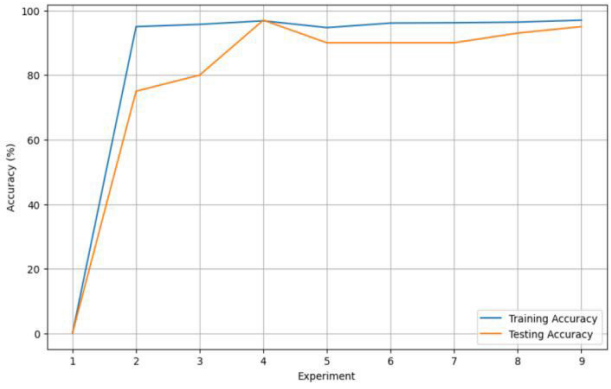

FIGURE 12. Training and testing accuracy of Bi-LSTM.

line representing testing accuracy. Both accuracies demonstrate an upward trend after the initial experiment, signifying improved performance over successive trials. This suggests that the model effectively learns from the training data and performs adequately on unseen data.

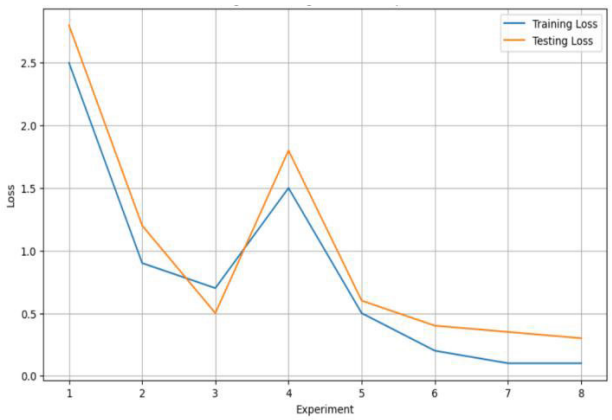

FIGURE 13. Training and testing loss.

F. COMPARATIVE ANALYSIS

Table 6 illustrates the comparison of the performance of the proposed method with existing methods. The existing metrics as in equations [17], [18], [19], [20] are used for evaluating the ability of the performance.

TABLE 6. Comparison of proposed method with existing methods.

Author	Method	Accuracy	Precision	Recall	F1-score
Pin Ni et al., [18]	Text2GraphQL	0.76	75	72	72
Amin et al., [19]	CNN	0.80	0.75	0.55	0.59
Devlin et al., [20]	BERT+ BiGRU	0.83	0.82	0.80	0.71
Xuo et al.,[21]	SVM	0.90	0.83	0.75	0.72
Proposed BERT-Bi-LSTM	BERT-Bi-LSTM	0.92	0.85	0.79	0.73

G. ERROR CASE STUDIES

Error case studies are essential for comprehending the shortcomings and potential areas of development in the suggested approach for text categorization with question answers. Several significant problems were found once the dataset was analyzed. First, ambiguity-related misclassification was noted, where a review that included the term ‘‘Boring’’ was incorrectly categorized as positive. This suggests that it can be difficult to discern between positive and negative attitudes when presented with ambiguous wording. Language translation problems were seen in reviews including non-English material, which might result in incorrect categorization or mistakes in the model. Analyzing these mistake case studies, improvements may be made to the system to improve its resilience and accuracy in categorizing course reviews, and feedback from learners.

H. CASE STUDY

An innovative use case in the medical field is the incorporation of the Q/A approach into a Chabot for healthcare. The Chabot, which is intended to provide medical advice and respond to patient inquiries, functions as a handy and easily navigable resource for those looking for information on symptoms, drug information, and general health advice. The BERT-Bi-LSTM architecture is employed by the Q/A model, which is the central element of the Chabot's comprehension of natural languages module, to analyze user inquiries, extract relevant information, and produce precise replies. Through user-friendly interfaces, including online applications or mobile apps, patients communicate with the Chabot by typing in their medical questions in normal language. In exchange, the Chabot quickly provides insightful responses and suggestions. The healthcare Chabot greatly increases the effectiveness of providing medical support by utilizing the Q/A methodology. Instant answers to their questions help patients, decreasing the need for manual intervention and shortening the time they must wait for medical consultations. This implementation has implications for several related fields. It improves patient education by supplying dependable health information, increases healthcare accessibility by offering prompt medical advice, especially in underserved or remote areas, and maximizes healthcare assets by automating regular inquiries and freeing up healthcare providers to concentrate on complex cases. The Q/A model's incorporation into the healthcare Chabot shows how it may enhance telemedicine, encourage patient involvement, and boost the effectiveness of healthcare delivery.

I. QUALITATIVE EXAMPLES AND SENSITIVITY ANALYSIS

In qualitative instances, the model shows its ability in a review expressing positivity towards a music class. The sentiment is evident with phrases like ''very helpful'' and ''basic music knowledge.'' This positive sentiment is accurately captured by the model, reflected in its high rating of 5. Sensitivity analysis modifies input variables, hyper parameters, dataset size, and model architecture to evaluate the resilience of the model. These assessments offer insightful information about how well the model performs in various contexts, which helps optimize and fine-tune the model for improved efficacy in practical applications. Qualitative examples demonstrate how accurately the model comprehends and reacts to a range of issues, and sensitivity analysis to guarantee the model's flexibility and dependability under different circumstances, which adds to its general effectiveness and usefulness.

J. DISCUSSION

In comparison to previous techniques, the suggested BERT-Bi-LSTM model offers improved performance metrics and represents a potential development in question-answering text categorization. Through the utilization of SemGloVe embeddings in conjunction with BERT and Bi-LSTM systems, the model exhibits enhanced F1-score values, accuracy, precision, and recall. Qualitative examples demonstrate the model's ability to correctly answer a range of questions, especially in the medical field. Sensitivity analysis also shows how resilient the model is to changes in input data, architectural options, hyper parameters, and dataset sizes. These results highlight the model's flexibility and dependability in practical settings. Subsequent investigations may examine supplementary optimization methodologies and assess the model's efficacy in diverse fields and linguistic frameworks. The suggested method has great potential in advanced text classification tasks for question-answering.

V. CONCLUSION

In this paper, BERT- Bi-LSTM model is proposed for the text classification of Q/A. This method utilized the CR23K and CR1000k datasets for the effective text classification of the NLP. The pre-processed data from SemGloVe is then utilizing by the pre-trained model of BERT. BERT BiLSTM-Attention Similarity Model increases the precision of question-to-question similarity calculations by employing BiLSTM-Attention for feature extraction, which assigns greater weight to significant portions of the embedding, and BERT as an embedding layer to transform questions into embedding. The exponential function is multiplied by the similarity function to produce the semantic similarity score. The proposed BERT – Bi-LSTM model has provided better results by utilizing evaluation metrics like accuracy, precision, recall and F1-score of values about 0.92, 0.85, 0.79 and 0.73 when compared to the existing methods of Text2GraphQL, GPT-2, BERT and SPARQL. To further improve accuracy of the similarity model, alternative pre-trained language models like RoBERTa or XLNet, could be suggested as embedding layer in supplant to BERT.

Subsequent investigations may pursue several avenues to tackle the constraints and enhance the conclusions of this research. To further improve model performance, technological advancements might entail experimenting with other pre-trained language models, such as RoBERTa or XLNet. The BiLSTM-Attention method might be improved, or other designs could be investigated for more effective feature extraction and embedding, as part of model improvement efforts. Expanding the use of the concept might involve implementing it in different fields outside of question and answering, including document categorization or sentiment analysis. Extending the model to novel domains, fine-tuning hyper parameters for various datasets, and ensuring efficacy and scalability in practical applications might present difficulties. Robust experimentation, working with domain experts to curate datasets, and ongoing model monitoring and improvement based on user input and changing requirements are some possible solutions.

REFERENCES

[1] C. Miura, S. Chen, S. Saiki, M. Nakamura, and K. Yasuda, ''Assisting personalized healthcare of elderly people: Developing a rule-based virtual caregiver system using mobile chatbot,'' *Sensors*, vol. 22, no. 10, p. 3829, May 2022.

[2] K. Nassiri and M. Akhloufi, "Transformer models used for text-based question answering systems," *Int. J. Speech Technol.*, vol. 53, no. 9, pp. 10602–10635, May 2023.
[3] J. Park, Y. Cho, H. Lee, J. Choo, and E. Choi, "Knowledge graph-based question answering with electronic health records," in *Proc. Mach. Learn. Healthcare Conf.*, Oct. 2021, pp. 36–53.
[4] Z. H. Syed, A. Trabelsi, E. Helbert, V. Bailleau, and C. Muths, "Question answering chatbot for troubleshooting queries based on transfer learning," *Proc. Comput. Sci.*, vol. 192, pp. 941–950, Jan. 2021.
[5] K. Shuang, Z. Zhang, J. Loo, and S. Su, "Convolution–deconvolution word embedding: An end-to-end multi-prototype fusion embedding method for natural language processing," *Inf. Fusion*, vol. 53, pp. 112–122, Jan. 2020.
[6] T. T. Nguyen, A. D. Le, H. T. Hoang, and T. Nguyen, "NEU-chatbot: Chatbot for admission of national economics university," *Comput. Educ., Artif. Intell.*, vol. 2, Jan. 2021, Art. no. 100036.
[7] M. A. Kuhail, J. Thomas, S. Alramlawi, S. J. H. Shah, and E. Thornquist, "Interacting with a chatbot-based advising system: Understanding the effect of chatbot personality and user gender on behavior," *Informatics*, vol. 9, no. 4, p. 81, Oct. 2022.
[8] R. Matic, M. Kabiljo, M. Zivkovic, and M. Cabarkapa, "Extensible chatbot architecture using metamodels of natural language understanding," *Electronics*, vol. 10, no. 18, p. 2300, Sep. 2021.
[9] J. J. Bird, A. Ekárt, and D. R. Faria, "Chatbot interaction with artificial intelligence: Human data augmentation with T5 and language transformer ensemble for text classification," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 4, pp. 3129–3144, Apr. 2023.
[10] F. de Arriba-Pérez, S. García-Méndez, F. J. González-Castaño, and E. Costa-Montenegro, "Automatic detection of cognitive impairment in elderly people using an entertainment chatbot with natural language processing capabilities," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 12, pp. 16283–16298, Dec. 2023.
[11] A. Nath, R. Sarkar, S. Mitra, and R. Pradhan, "Designing and implementing conversational intelligent chat-bot using natural language processing," *Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol.*, vol. 7, no. 3, pp. 262–266, May 2021.
[12] H. Xiong, S. Wang, M. Tang, L. Wang, and X. Lin, "Knowledge graph question answering with semantic oriented fusion model," *Knowl.-Based Syst.*, vol. 221, Jun. 2021, Art. no. 106954.
[13] B. Xu, R. Cai, Z. Zhang, X. Yang, Z. Hao, Z. Li, and Z. Liang, "NADAQ: Natural language database querying based on deep learning," *IEEE Access*, vol. 7, pp. 35012–35017, 2019.
[14] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. A. Almazroi, "A fine-tuned BERT-based transfer learning approach for text classification," *J. Healthcare Eng.*, vol. 2022, Jan. 2022, Art. no. 3498123.
[15] G. R. S. Silva and E. D. Canedo, "Towards user-centric guidelines for chatbot conversational design," *Int. J. Human–Comput. Interact.*, vol. 40, no. 2, pp. 98–120, Jan. 2024.
[16] P. Ni, R. Okhrati, S. Guan, and V. Chang, "Knowledge graph and deep learning-based text-to-GraphQL model for intelligent medical consultation chatbot," *Inf. Syst. Frontiers*, vol. 26, no. 1, pp. 137–156, Feb. 2024.
[17] Q. Qiu, M. Tian, K. Ma, Y. J. Tan, L. Tao, and Z. Xie, "A question answering system based on mineral exploration ontology generation: A deep learning methodology," *Ore Geol. Rev.*, vol. 153, Feb. 2023, Art. no. 105294.
[18] M. Bayer, M.-A. Kaufhold, B. Buchhold, M. Keller, J. Dallmeyer, and C. Reuter, "Data augmentation in natural language processing: A novel text generation approach for long and short text classifiers," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 1, pp. 135–150, Jan. 2023.
[19] M. Z. Amin and N. Nadeem, "Convolutional neural network: Text classification model for open domain question answering system," 2018, *arXiv:1809.02479*.
[20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[21] X. Luo, "Efficient English text classification using selected machine learning techniques," *Alexandria Eng. J.*, vol. 60, no. 3, pp. 3401–3409, Jun. 2021.
[22] E. Kasthuri and S. Balaji, "Natural language processing and deep learning chatbot using long short term memory algorithm," *Mater. Today, Proc.*, vol. 81, no. 2, pp. 690–693, 2021.
[23] A. S. Imran, R. Yang, Z. Kastrati, S. M. Daudpota, and S. Shaikh, "The impact of synthetic text generation for sentiment analysis using GAN based models," *Egyptian Informat. J.*, vol. 23, no. 3, pp. 547–557, Sep. 2022, doi: 10.1016/j.eij.2022.05.006.
[24] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J. Big Data*, vol. 8, no. 1, pp. 1–16, Dec. 2021.
[25] J. C. Costa, T. Roxo, J. B. F. Sequeiros, H. Proenca, and P. R. M. Inácio, "Predicting CVSS metric via description interpretation," *IEEE Access*, vol. 10, pp. 59125–59134, 2022.
[26] L. Gan, Z. Teng, Y. Zhang, L. Zhu, F. Wu, and Y. Yang, "SemGloVe: Semantic co-occurrences for GloVe from BERT," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 2696–2704, 2022.

M. SUGUNA received the B.E. degree in CSE from the Kumaraguru College of Technology, Coimbatore, the master's degree in CSE from the Government College of Technology, Coimbatore, and the Ph.D. degree in information and communication engineering from Anna University, Chennai. She is currently working as an Assistant Professor Senior Grade-II with the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai. She has published more than 20 research articles in international journals and 15 international conferences. Her research interests include data analytics, health care analytics, cloud computing, and agile project management. She is a member of ISTE and IAENG.

K. S. SAKUNTHALA PRABHA is currently working as an Ambitious Research Scholar in computer science and engineering program with Vellore Institute of Technology, Chennai Campus. She has contributed to academic research by publishing a conference paper, showcasing her dedication in advanced knowledge in her chosen field. Her research interests include deep learning and data analytics.