OPEN ACCESS

BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding

Arun Babu [*], Sekhar Babu Boddu

Koneru Lakshmaiah Education Foundation, India

ARTICLE INFO

ABSTRACT

The advent of modern technologies like Artificial Intelligence(AI), Internet of Things(IoT) and Deep Learning(DL) has ushered in a transformative era in healthcare, offering innovative solutions towards personalized healthcare by enhancing the quality of various medical services. Our proposed methodology involves the development of a BERT-based medical chatbot, leveraging cutting-edge deep learning technology to significantly enhance healthcare communication and accessibility. The traditional challenges faced by medical chatbots, such as imprecise understanding of medical conversations, inaccurate responses to jargon, and the inability to offer personalized feedback, are addressed through the utilization of Bidirectional Encoder Representations from Transformers (BERT). The performance metrics of our chatbot underscore its effectiveness. With an accuracy of 98%, the chatbot ensures a high level of precision in handling medical queries. The precision score of 97% attests to the accuracy and reliability of its responses. The AUC-ROC score of 97% indicates the chatbot's exceptional ability to predict specific diseases based on user queries and symptoms, showcasing its robust predictive power. Furthermore, a recall of 96% demonstrates the chatbot's capability to avoid missing cases in medical diagnoses, ensuring comprehensive coverage of potential conditions. The F1 score of 98% showcases the chatbot's proficiency in delivering accurate and personalized healthcare information, striking a harmonious balance between precision and recall. Our BERT-based medical chatbot not only addresses the limitations of traditional approaches but also achieves a remarkable performance with high accuracy, precision, predictive power, and comprehensive coverage, making it a valuable tool for advancing the quality of healthcare services.

1. Introduction

In modern society, most of the realms are digitized and automated in order to offer efficient and increased availability of services through utilization of advanced, sophisticated software technologies. Chatbots are software programs that interact with humans through natural language with the sole aim of simulating a human-like conversation in response to text or voice.[1]

1.1. Overview on AI's role in medical care

Artificial Intelligence (AI) which acts as an alternative to human cognition is progressively being utilized in healthcare. AI-based chatbot systems operate as automated conversational agents to promote health and treatment management by facilitating a recognizable environment for patients and adequate support to physicians and care givers. IoT based revolution along with AI is constantly re-designing contemporary health care systems through its high technical, cost-effective, and communal prospects by presenting a paradigm shift to healthcare.[2] A study by Dr. Eric J. Topol and colleagues, published in the journal "The Lancet Digital Health" demonstrated the significant impact of AI on medical diagnostics. The study highlights how AI algorithms, particularly in the field of radiology, have substantially improved diagnostic accuracy and efficiency. The authors emphasize the potential of AI and its sub components like DL to enhance clinical workflows and ultimately contribute to more effective patient care.[3] Medical BigData powered by mounting accessibility of healthcare data together with speedy progression of numerous data analytics and mining techniques, several

applications in healthcare, ranging from remote monitoring to medical device integration assists in safe keeping of patients and improves the manner in which medical practitioners deliver personalized care. Through successful amalgamation of IoT & AI technologies, chatbots for medical assistance in healthcare can completely transform the face of medical industries world-wide.[1] The Chatbot agent interacts with the users and responds to their queries by making use of its self-learning ability through which they can easily understand the input and offers the precise output to the users. The chatbot creates its own database within the run time from which it recognizes several patterns and through which it gives exact reply to the individual queries.[4] (Fig. 1) Communication with users for the input and corresponding output is possible through necessary algorithms and Prediction Analysis tools.[5]

1.2. AI based Chatbots

Chatbots leverage NLP,[7] dialogue management, and response generation to interact with users in a human-like approach, offering necessary assistance through providing information, answering queries, and performing chores as necessitated. The sophistication of a chatbot's task can vary extensively, ranging from simple rule-based systems to more complex AI-driven models competent enough to comprehend and generate natural language text.[8] Conventional Machine Learning (ML) chatbots do not rely on deep learning or neural networks, often faces several challenges that can impact their effectiveness and capabilities as they struggle to understand and maintain context in complex conversations.[9] Since they mostly rely on predefined rules and patterns, their ability to handle nuanced or multi-turn interactions is limited.[10] ML chatbots may not possess the advanced natural language understanding (NLU) capabilities of deep learning models thus may have difficulty interpreting user queries accurately, particularly when dealing with complex language, idiomatic expressions, or vernacular slangs. This limits their skills to generate dynamic and contextually pertinent responses, resulting in a reduced amount of appealing and recurring interactions.[11] Quality and quantity of training data along with a set of predefined rules in ML requires constant modification of their behavior through manual intervention and retraining, thereby making them less adaptive to changing user needs.[12] While traditional ML chatbots can still be helpful for particular applications and use cases, these confrontations emphasize the fact that several organizations are gradually turning to deep learning and neural network-based chatbots to facilitate towards more sophisticated and efficient conversational experiences.[13]

1.3. BERT (Bidirectional Encoder Representations from transformers)

BERT is a state-of-the-art DL model developed for Natural Language Understanding (NLU) and processing tasks through its transformer-based neural network architecture.[14] Through effective pre-training on adequate corpus of medical data, it can learn to recognize the context and significance of words by reflecting on the words that arrives prior to and following them in sentences and documents. In the medical domain, BERT serves as a powerful natural language processing (NLP) model. Its bidirectional understanding of context allows it to interpret medical jargon, understand nuanced language in healthcare conversations, and provide accurate responses to medical queries. This capability makes BERT well-suited for tasks such as medical text analysis, diagnosis support, and generating contextually relevant information in healthcare-related applications. By leveraging BERT in medical contexts, researchers and developers aim to enhance the precision and effectiveness of natural language understanding within the field of healthcare. Our proposed BERT based medical chatbot is an advanced conversational AI system that employs the Bidirectional Encoder Representations to understand and respond to medical-related queries and connect with necessary medical stakeholders through necessary conversations related to healthcare. The key contributions of our BERT-based medical chatbot include:

- BERT's bidirectional pre-training enables the chatbot to comprehend medical queries with exceptional degree of accuracy, even if they involve complicated medical terms and context.
- Ability to excel in multi-turn conversations, where continuous interactions with users are facilitated making it valuable in telemedicine and patient awareness situations.
- Personalized responses and assistance towards triaging patients by comprehensive evaluation of severity of their symptoms and directing them to the suitable level of care or urgency, leading to enhanced decision support through optimization of resources allocation.

Thus our BERT-based medical chatbot can offer considerable contributions to medical industry by improving the quality of patient care by supporting healthcare professionals. It can also improve information accessibility, and ease healthcare processes ultimately leading to better healthcare outcomes and a more competent healthcare ecosystem. The organization of our study is as follows: Section 2 outlines the prominent research works in relevance to this specific issue. Section 3 presents the proposed medical chatbot using BERT methodology and explains its functioning in detail followed by empirical findings and analysis in Section 4. The final Section 5 concludes this study by summarizing its core points and offers prospective inclinations for future research.

2. Related works

Through successful research works underwent by scholars and application developers, integration of modern technologies like AI,
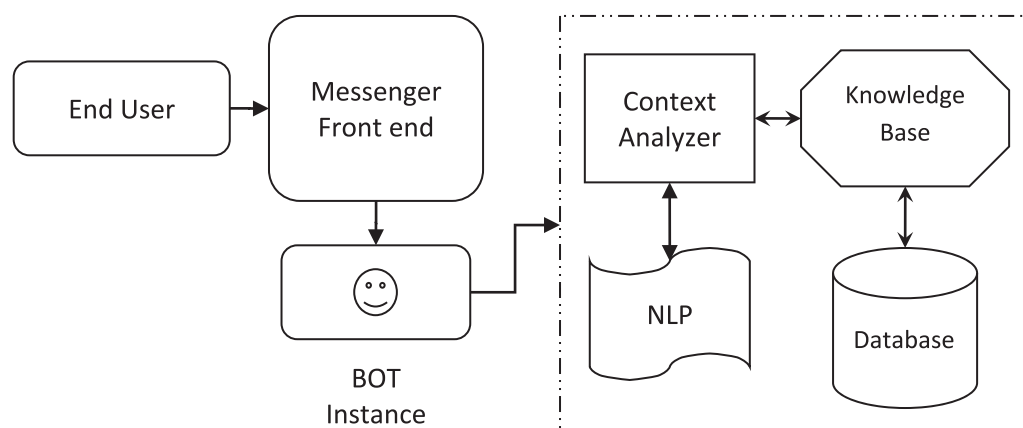

Fig. 1. AI chatbot framework (Source: Battineni et al. [6]).

machine learning and deep learning in medicine, have enhanced the lives of health professionals and patients by offering a high quality care that has the potential to can transform the future of medicine. Health professionals have evolved along with these advanced techniques to perform their jobs better and focus more on delivering excellent personalized medical care. This section is dedicated to recent and related research works in which latest advancements have been utilized to implement Medical Chatbots.

Thwala et al.[15] suggested a novel Chatbot methodology where Recurrent Neural Network (RNN) has been used to imitate the functionality of human brain which can classify the available response classes along with its probability value in COVID prognosis based applications. The system demonstrates minimal response time along with high accuracy and vouches for deep learning approach since it provides better learning in order to make intelligent decisions. Moreover deep learning-based NLP chatbots accomplish greater classification accuracy as they possess the ability to interpret conversations better.[7] Shree et al.[12] highlighted the demand for sophisticated Chatbot systems in health domain where effective communication with consumers is required to provide useful information. Designing healthcare virtual assistant that can understand and respond to complex medical questions in real-time, however, is still a challenging mission. Grasshopper-optimized spiking neural network has been proposed to optimize the weight of synaptic and interconnectivity which enables the chatbot software to process data effectively and arrives at appropriate decisions. Extensive evaluation involving actual health inquiries along with dataset acquired from Kaggle validates the accuracy, speed, and user experience of the suggested model which is found to be satisfactory.

Chakraborty et al.[16] proposed a medical Chatbot which carries out human interaction and prediction operations using Multi Layer Perceptron (MLP). In order to bridge the gap regarding theoretical guidelines and real-time recommendations for developing AI chatbots, this study aims to underline necessary functionalities and possible applications of medical chatbots. The recommended approach explores various challenges posed by practical applications of emerging technologies which will help researchers to get a better understanding of the architecture and utility of modern revolutionary technologies that can help in continuous improvement of medical chatbot operations. Soufyane et al.[7] came forward with an innovative Chatbot model for disease prediction using term and inverse data frequency techniques. This research delves deeper into the merits and demerits of different chatbot design architectures like generative and retrieval types and concludes that chatbots must combine both these approaches to leverage the advantages of both. For instance, the proposed approach made use of retrieval-based methods to identify the most appropriate responses from a knowledge base and then utilized a generative model to fine-tune and customize the response further.

Tamizharasi et al.[17] proposed a chatbot system that employs SVM to aid medical facilities in assisting patients through voice or text based medical related inquiries. The suggested system receives output based on constructed database that contains medical diagnosis knowledge and provides treatment options to specific disease identified. A multi-lingual chatbot was presented by Badlani et al.[18] which aims to perform disease diagnosis based on user symptoms using estimations done on sentence similarity through TF-IDF (Term Frequency-Inverse document Frquency) and Cosine Similarity techniques and selecting the most relevant reply from its information database which is highly suitable for utilization in rural parts of country as it supports three languages.[17]

El-Zini et al.[19] put forward a deep learning framework based AI assistant to enhance the medical students' conversational skills in completing a clinical assessment of the patient's medical state which are a part of their professional training curriculum. Initially, DL networks learns domain specific word embeddings followed by LSTM derived sentence embeddings prior to CNN model based selection of appropriate response to a particular question from a pre-defined script. Empirical findings carried out on an indigenous dataset validate the efficiency of this framework in comparison to other classic approaches.

Vaira et al.[10] developed a PregBot which is a type of medical virtual support for pregnant women, mothers, and families with young children. This ML and NLP based system offers specialized help and instructions by comprehending the relevant situations by making use of retrieval based architecture using Decision trees and rule base systems. Serban et al.[20] developed a AI based chatbot system that comprises of ensemble model involving natural language generation and retrieval mechanism which makes use of templates, bag-of-words and latent variable neural network techniques. Reinforcement learning is applied on community driven data and real-world user interactions. Training is done to select an appropriate response from its ensemble-based dialogue system. Deep learning and reinforcement learning models are utilized for natural language retrieval and generation which enables the suggested Chatbot system to construct considerable enhancements.

Most of the existing researches highlight the intricacies and challenges involved in developing an effective and secure medical chatbot. The accuracy and safety of medical advice rendered, handling of ambiguity and uncertainty, privacy and ethical concerns, interoperability, multi-modal, multi-modal communicational abilities are some of the concerns that requires appropriate solutions.[18] Addressing these research gaps in medical chatbots often necessitates a holistic approach that combines AI capabilities along with domain-specific data, expert understanding, ethical contemplations, and rigorous testing and validation procedures. Furthermore, the development and deployment of medical chatbots must strictly adhere to regulatory and ethical guidelines to make sure patient's safety and data privacy. Addressing these gaps is essential to ensure a reliable and valuable contribution towards healthcare.

3. Proposed BERT-based medical Chatbot system

Our proposed methodology makes use of BERT (Bidirectional Encoder Representations from Transformers) which is a DL architecture (Fig. 2) that has modernized the Natural Language Understanding (NLU) operations like text classification, query responses, content generation etc.

The following are the key modules on our medical Chatbot system:

- Data collection
- Text Processing module
- BERT model
- Context management
- Entity and Intent Recognition unit
- Dialogue management and response Generation section

3.1. Dataset collection

Medical chatbots are designed to assist users with medical information, advice, or scheduling appointments. To train and evaluate a medical chatbot effectively, its mandatory to access to appropriate datasets that contain examples of user queries, intents, named entities, and possible responses. Our proposed system makes use of MIMIC-III, BioASQ, PubMed, COVID 19 datasets. MIMIC-III is a comprehensive dataset containing clinical notes, diagnoses, medications, lab results, and other medical information. PubMed facilitates access to a vast collection of biomedical literature, including abstracts and full-text articles. COVID-19 includes pandemic related medical information, research articles, and clinical data.[21,22] BioASQ contains biomedical question and answers for information retrieval.

3.2. Text processing module

Text processing is a crucial phase in designing medical chatbot which helps to ensure that users receive accurate and relevant information
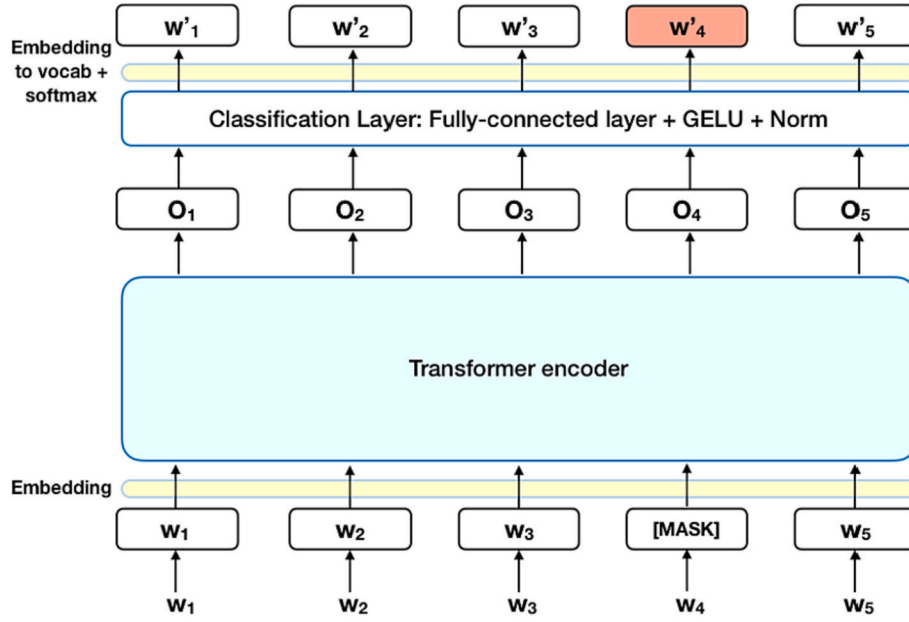
Fig. 2. Generic BERT-architecture.

while maintaining their privacy and adhering to medical regulations.[3] It involves a combination of linguistic analysis steps like tokenization, stop word removal, NLP techniques like stemming and lemmatization, vectorization, N-grams and PoS(Part of Speech) tagging tasks.

3.3. BERT model

One of the major characteristic of BERT is its bidirectional aspect through which it can read text in both directions. This enables it to capture contextual information from both previous and subsequent words, making it superior in comprehending the precise context and relationships between words. BERT is built upon the Transformer architecture which relies heavily on attention mechanisms through which it can focus on various parts of the input sequence while processing it simultaneously. BERT is pre-trained to enable prediction of missing words /masked words within sentences and this process helps to capture rich semantic and contextual information from the input data. BERT consists of multiple layers of encoders which contains a multi-head self-attention mechanism and a feed-forward neural network.[23] These layers are stacked on top of each other to create a deep architecture. BERT employs word, segment and position embeddings to represent the meaning of individual words, distinguish between different segments of text and to encode the position of words within a sequence. After pre-training, BERT can be fine-tuned on specific downstream NLP tasks, such as text classification, or entity recognition, or query responding.[24] Fine-tuning helps in adapting the pre-trained model to perform well on a particular task. BERT can be extended to multilingual and multimodal data by using variants like VisualBERT.

In a chatbot, user queries are represented as sequences of words or tokens. BERT performs tokenization to break text into sub-word tokens. Let's denote the input query as a sequence of tokens:

$$X = [x1, x2, \ldots, xn] \tag{1}$$

Where x_i represents the ith token.

The attention mechanism in BERT can be mathematically represented as follows: Given a set of input vectors Q (Query), K (Key), and V (Value), the output of attention mechanism for a single head can be calculated as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{dk}}\right)V \tag{2}$$

Where Q,K,V are matrices representing the input queries, keys, and values. QK^T is the dot product of Q and the transpose of K. dk is the dimension of the key vectors. Softmax function is applied along the rows for each query.

Multi-Head Attention layer captures different aspects of the input whose output can be represented as follows:

$$MH(Q, K, V) = Con\ (h1, h2, \ldots, h_h)W^O \tag{3}$$

Where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$ represents the output of the ith attention head. W_i^Q, W_i^K, W_i^V are the weight matrices specific to each head and W^O is the weight matrix of output projection.

BERT contains multiple layers of self-attention-based transformer encoders and the output of a single transformer encoder layer can be represented as.

$$H = BERT(X) \tag{4}$$

Here, H represents the sequence of contextualized embeddings for each token in the input which is characterized in the context of entire input sequence. This aspect is crucial for understanding the meaning of medical queries.The core of the BERT model involving multiple layers of Transformer encoders whose output for a given input sequence X can be denoted as:

$$EncoderLayer(X) = LayerNorm(X + MultiHead(X, X, X)) \tag{5}$$

Where X is the input to encoder layer.$MultiHead(X,X,X)$represents the multi-head self-attention mechanism applied to input. LayerNorm is layer normalization. In a medical chatbot, task-specific output layer needs to be specified that takes the contextualized embeddings H and maps them to specific output. The detailed design of this output layer depends on the task performed by the chatbot, in entity recognition tasks like identifying medical conditions or symptoms in the query, Conditional Random Field (CRF) is used with softmax activation:

$$Entities = OutputLayer(H) \tag{6}$$

Here, *Entities* represents the recognized entities in the input query. For intent classification tasks like determining the user's intent while asking a question about symptoms or treatments, softmax layer is used over the pooled embeddings:

$$Intent = softmax(pool(H)) \tag{7}$$

Where pool(H) denotes the max pooling operation over the sequence of word embeddings. For generating responses in a dialogue-based chatbot, a decoder model is adopted which takes H as input and generates a response sequence token by token. BERT is fine-tuned on medical domain specific data to adapt it to medical language and context which involves training the model on a specific medical dataset, adjusting its weights to improve performance on medical tasks.

Fine-tuning BERT involves training the pre-trained BERT model on a specific downstream task using task-specific labeled data. Let X represents the input sequence for the downstream task which is a sequence of tokens and Y represents the ground truth labels for the task. L_{oss} represents the loss function specific to the downstream task. Cross-entropy loss is represented as.

$$L_{oss} = -\sum_i Y_i log(\hat{Y}_i) \tag{8}$$

Where Y_i and $\hat{Y}_i$ are the true and predicted probabilities for class i. The fine-tuning process involves optimizing the pre-trained BERT model's parameters with respect to the loss:

$$min_{\theta_{BERT}} L_{oss}(X, Y; \theta_{BERT}) \tag{9}$$

This optimization process uses Adam to minimize the task-specific loss. The gradients are calculated with respect to the BERT model's parameters, and the model is fine-tuned for the downstream task. During fine-tuning, BERT layers are updated to avoid over-fitting while still leveraging the pre-trained knowledge from BERT. The fine-tuning process continues for a fixed number of iterations or until convergence is achieved and then used for query answering in medical chatbot. Algorithm for BERT model is provided (Algorithm 1, Fig. 3).

BERT offer solutions to several of the research gaps in existing medical chatbots by handling ambiguity and uncertainty by considering the entire input sentence. Its bidirectional nature allows it to capture dependencies between words, which are crucial for understanding context and handling uncertainty.
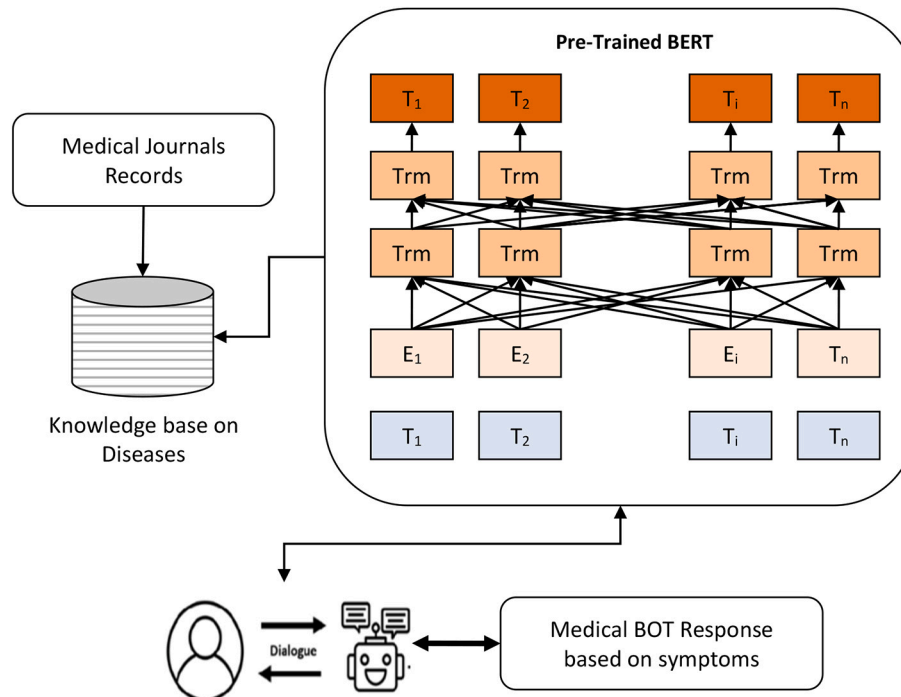

Fig. 3. Proposed design for BERT base medical BOT.

```
================================================================================
Layer (type:depth-idx)                                          Param #
================================================================================
BERT_Arch                                                       --
├─DistilBertModel: 1-1                                          --
│    └─Embeddings: 2-1                                          --
│         └─Embedding: 3-1                                      (23,440,896)
│         └─Embedding: 3-2                                      (393,216)
│         └─LayerNorm: 3-3                                      (1,536)
│         └─Dropout: 3-4                                        --
│    └─Transformer: 2-2                                         --
│         └─ModuleList: 3-5                                     (42,527,232)
├─Dropout: 1-2                                                  --
├─ReLU: 1-3                                                     --
├─Linear: 1-4                                                   393,728
├─Linear: 1-5                                                   131,328
├─Linear: 1-6                                                   1,285
├─LogSoftmax: 1-7                                               --
================================================================================
Total params: 66,889,221
Trainable params: 526,341
Non-trainable params: 66,362,880
================================================================================
```

Fig. 4. Summary of Proposed BERT model.

Algorithm: Pseudocode for a BERT-based Medical Chatbot:

Step 1: Pre-training

\# Load or fine-tune a pre-trained BERT model on a medical text corpus

```
bert_model = load_pretrained_medical_BERT_model()
```

\# Pre-trained weights or fine-tuning

Step 2: User Input Processing

\# Receive user queries or input text

```
user_input = get_user_input()
```

\# Tokenize and preprocess the input

```
input_tokens = tokenize(user_input)
input_ids = convert_tokens_to_ids(input_tokens)
attention_mask = create_attention_mask(input_tokens)
```

Step 3: BERT-based Contextual Understanding

\# Pass the input through the medical BERT model

```
contextual_embeddings = bert_model(input_ids, attention_mask)
```

\# Extract relevant medical entities (e.g., conditions, symptoms)

```
medical_entities = extract_medical_entities(contextual_embeddings)
```

Step 4: Intent Recognition

Classify user intent (e.g., symptom inquiry, treatment information)

```
user_intent = classify_intent(contextual_embeddings)
```

Step 5: Knowledge Base Integration

\# Query a medical knowledge base with extracted entities and user intent

```
knowledge_base_response = query_medical_knowledge_base(medical_entities
user_intent)
```

Step 6: Response Generation

\# Generate a response based on the knowledge base information

```
chatbot_response = generate_response(knowledge_base_response)
```

Step 7: User Interaction

\# Present the response to the user

```
display_response(chatbot_response)
```

Step 8: Dialog Flow

\# Maintain a conversation history to provide context for follow-up queries

```
conversation_history.append((user_input, chatbot_response))
```

\# Allow for natural language interactions and follow-up questions

```
while True:
    user_input = get_user_input()
```

\# Continue with steps 2-7 based on the new user input

Step 9: Exit or Continue

\# Allow the user to exit the chat or continue the conversation as needed

Design of proposed BERT model and its hidden layers is available in Fig. 4, BERT can be fine-tuned for specific medical contexts and individual patient data and can offer more personalized responses based on a patient's medical history and context. BERT can be integrated with other models that handle different modalities to understand and respond to various forms of data, enhancing their utility in medical scenarios where multiple data types are involved. BERT's multilingual capabilities can aid in creating chatbots sensitive to different languages and cultures and can facilitate interoperability between different chatbot platforms and healthcare systems by using a common language representation.

3.4. Context management

This is a critical aspect of building an effective medical chatbot, as it enables the chatbot to maintain awareness of the ongoing conversation, remember previous user inputs, and provide contextually relevant responses. The chatbot maintains a conversation context to keep track of the user's previous queries and responses, enabling coherent and context-aware interactions.

3.5. Entity and intent recognition unit

Determining the user's intent, such as seeking medical advice, scheduling an appointment, or asking for information about a specific condition through extracted entities, and context, the chatbot generates relevant and informative response. Reply generation involves using rule-based templates and retrieval-based methods to provide answers and recommendations.

3.6. Dialogue management and response generation section

To maintain a dialogue state and to keep track of the conversation flow and manage the dialogue turns effectively, chatbot uses BERT to generate responses or suggestions for user queries. Combining BERT-generated responses with pre-defined template responses, coherence and variety along with integration of medical knowledge databases provides accurate and up-to-date information.

3.7. Inclusion and exclusion criteria for selecting medical chatbot questions

The chatbot has been designed to operate seamlessly in multiple languages, addressing linguistic diversity. Cultural sensitivity is crucial, necessitating an understanding and respect for local cultural norms in its responses. To facilitate accessibility, the chatbot should have low resource requirements, making it practical for areas with limited technological infrastructure. Adaptability to connectivity issues, such as intermittent or low internet access, ensures widespread usability. Medical accuracy and Ethical considerations has been prioritized, encompassing user privacy, confidentiality, and the avoidance of biases and discrimination. Language limitations where the chatbot is restricted to a single language, has been avoided to ensure inclusivity. Cultural insensitivity in responses and High resource requirements need to be pondered upon in resource-constrained areas. Providing inaccurate or unreliable medical information poses risks to users' health and is a significant exclusion factor. Incompatibility with existing healthcare systems hinders effective collaboration and information sharing, making it a critical exclusion criterion.

4. Results analysis and discussion

Analyzing and discussing the results of a BERT-based medical chatbot is a crucial step in assessing its performance and understanding its strengths and weaknesses. The Dataset size is around 11,000 medical questions and answers extracted from Sources MIMIC-III, BioASQ, PubMed, COVID 19 datasets. Clinical experts from diverse medical domains, healthcare practitioners, including nurses and general practitioners, provided valuable insights into patient interactions and common health-related inquiries. To ensure the linguistic and cultural appropriateness of the questions, language experts and cultural sensitivity consultants were actively engaged in the process. Ethical considerations were addressed through collaboration with bioethicists and

legal experts, ensuring that the methodology adhered to the highest standards of patient privacy and confidentiality. Furthermore, input from patients and patient advocacy groups was sought to incorporate the patient perspective and validate the relevance of the selected questions. This multidisciplinary collaboration ensures that the proposed methodology is robust, inclusive, and reflective of the diverse perspectives and expertise required for a comprehensive understanding of medical chatbot interactions. We collected sentences containing information on symptoms with labels in relation to medical specialty aspect, conducted data preprocessing, and ultimately constructed a pipeline of sentences for this study. The hardware and software requirements are NVIDIA GeForce RTX 3090 (16GB VRAM) with 32GB DDR4 RAM and Python 3.8 has been used for simulation. The BERT model parameters are provided in Table 1.

Hugging Face's Transformers library is utilized for working with BERT as it provides pre-trained BERT models and tools for fine-tuning them on specific jobs like medical tasks. Deep learning framework is TensorFlow and NLP library spaCy is used for text preprocessing, tokenization, and other NLP-related tasks.

We have employed five performance evaluation metrics i.e. Accuracy, Precision AUC-ROC analysis, Recall and F1-Score in our experiment. The metrics can be defined as follows:

$$Accuracy = (TN + TP)/(TN + TP + FN + FP) \tag{10}$$

$$Precision = TP/(TP + FP) \tag{11}$$

$$Recall = TP/(TP + FN) \tag{12}$$

$$F1\ Score = 2^* ((Precision^* Recall)/(Precision + Recall)) \tag{13}$$

Where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative. The performance of the BERT model is evaluated by comparing it with other baseline models like LSTM, SVM, and BI-LSTM in terms of afore-mentioned performance metrics for the same task and the results are presented in Table 2 and confusion matrix is presented in Fig. 5.

It could be observed that our BERT based Medical Chabot has achieved the highest accuracy of 94%, demonstrating its superior performance. It had a precision of 0.92, indicating high accuracy in query responses. The AUC-ROC score of 0.97 suggests excellent power to predict specific diseases based on user queries and symptoms. Recall at 0.95 indicates its ability to ensure that the chatbot doesn't miss cases where the condition is present in medical diagnosis. F1 score of 0.93 provides a balanced measure of precision and recall.

LSTM demonstrated competitive performance compared to BERT, with an accuracy of 88%, precision of 0.86, AUC-ROC score of 0.92, recall at 0.89 and F1 score of 0.87. SVM achieved an accuracy of 84%, which is lower than both BERT and LSTM, precision of 0.82, AUC-ROC score of 0.88, recall of 0.80 and F1 score of 0.81. Bi-LSTM performed well with an accuracy of 91% but was outperformed by BERT (Fig. 6).

It had a precision of 0.90, indicating high accuracy and AUC-ROC score of 0.94 suggests good discriminatory ability, recall at 0.92 and

Table 1
Simulation parameters.

Hyper parameters	Values
Learning rate	0.001
Batch Size	16
Weight Decay	0.01
Dropout rate	0.5
Attention masking	15%
Optimizer	Adam
Loss function	Cross Entropy
Attention Head Count	8
Transformer layer count	12
Tokenizer	Words

Table 2
Performance Analysis of proposed model with other baseline models.

Model	Accuracy	Precision	AUC-ROC	Recall	F1 Score
LSTM	0.88	0.86	0.92	0.89	0.87
SVM	0.84	0.82	0.88	0.8	0.81
BI-LSTM	0.91	0.90	0.94	0.92	0.91
Proposed Model	0.98	0.97	0.97	0.96	0.98

F1 score of 0.91 provides a balanced measure of precision and recall. Comparing our findings with previous research in medical chatbots, our study advances the field by leveraging the bidirectional context understanding of BERT. This approach overcomes challenges faced by traditional models in accurately interpreting medical nuances, contributing to enhanced precision and context relevance.

Thus our proposed BERT outperformed all baseline models (LSTM, SVM, and BI-LSTM) across all metrics, demonstrating its effectiveness in contextual understanding and pre-trained embeddings give it a significant advantage in capturing nuanced linguistic patterns.

4.1. Advantages and disadvantages of our proposed model

The proposed medical chatbot based on BERT showcases notable advantages in its ability to deliver highly accurate and precise responses to medical queries, boasting a 98% accuracy and 97% precision. Its exceptional disease prediction capabilities, reflected in a 97% AUC-ROC score, enhance its utility for foreseeing specific health conditions based on user inputs. Furthermore, the chatbot ensures comprehensive coverage with a high recall of 96%, minimizing the risk of overlooking potential medical diagnoses. The model strikes a balance between precision and recall, achieving a well-rounded F1 score of 98%. Leveraging BERT's bidirectional context understanding, the chatbot excels in interpreting nuanced medical language, facilitating natural and contextual conversations.

However, challenges exist, including the computational demands of BERT-based models, potential biases in training data influencing performance, and the interpretability of complex decision-making processes. Continuous learning for adaptation to evolving healthcare scenarios and addressing data privacy concerns also pose operational considerations. Additionally, the model's effectiveness may vary in handling uncommon medical cases with limited training data. Despite these challenges, the proposed chatbot stands as a promising tool in revolutionizing healthcare communication and information accessibility.

5. Conclusion

This study presents the development of a medical chatbot powered by BERT, a cutting-edge NLP and demonstrates a significant innovation in healthcare information dissemination and patient engagement. The achieved accuracy of 98% demonstrates the chatbot's precision and reliability in addressing a wide array of medical queries. Notably, a precision score of 97% underscores the model's accuracy, providing trustworthy responses in interpreting intricate medical language. The exceptional AUC-ROC score of 97% signifies the chatbot's robust predictive power, particularly in foreseeing specific diseases based on user queries and symptoms. Additionally, the chatbot exhibits a high recall of 96%, assuring that it minimizes the risk of overlooking potential medical diagnoses, thereby ensuring comprehensive coverage. The F1 score of 98% further emphasizes the chatbot's proficiency in achieving a harmonious blend of precision and recall. Future improvements include integration of multilingual support for broader user reach, and continuous learning from user interactions along with strengthening data privacy and security measures. In conclusion, the medical chatbot developed using BERT offers a transformative approach to healthcare information dissemination by addressing future enhancements, it has
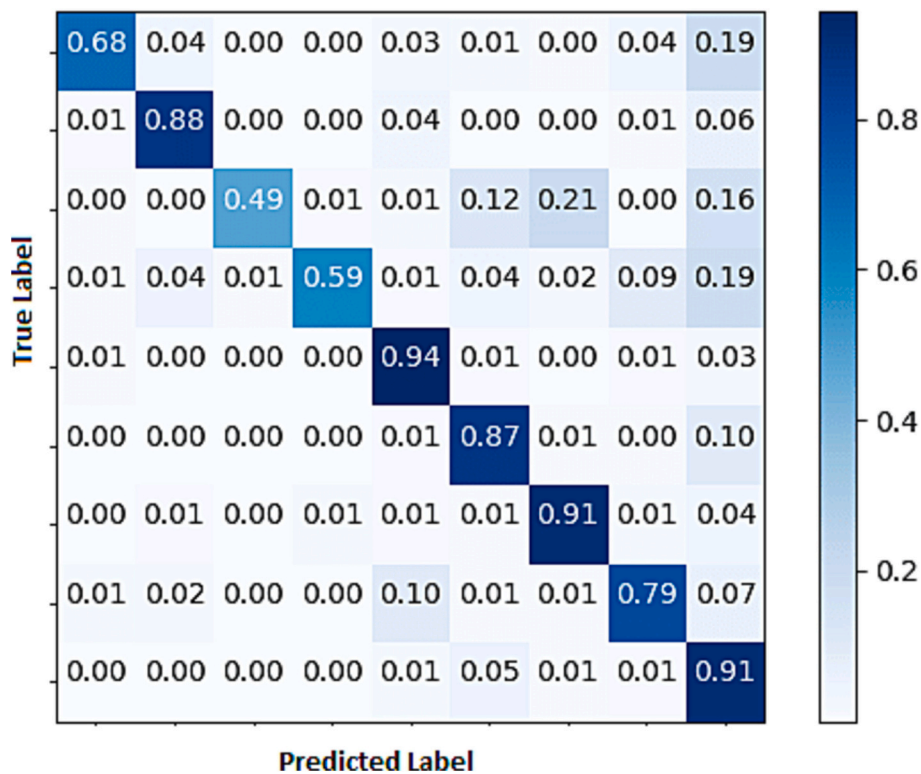
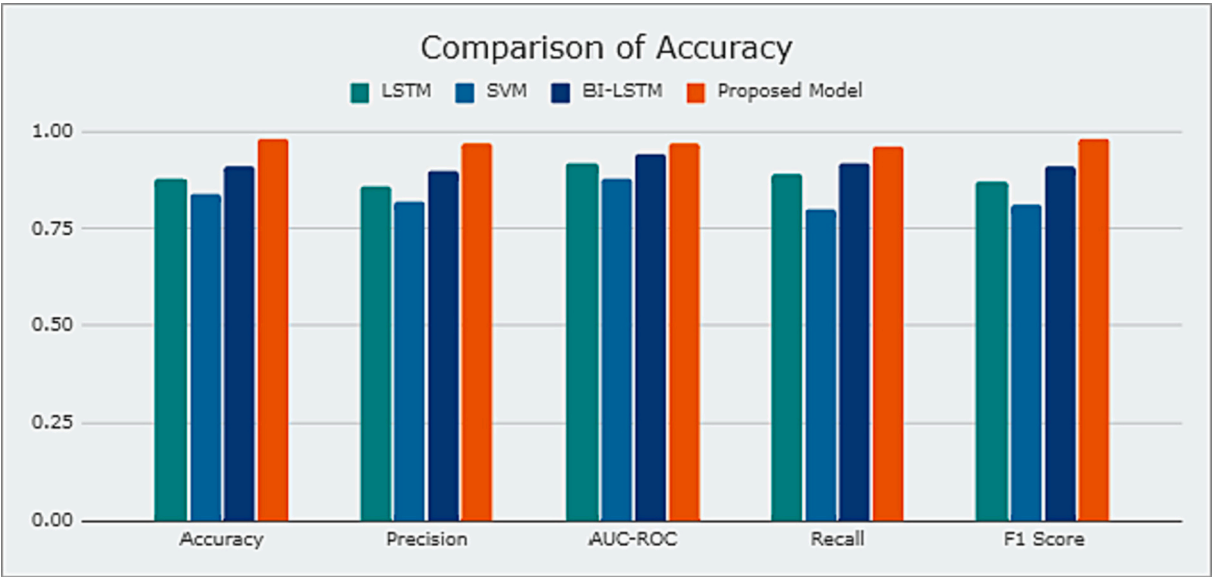**Fig. 5.** Confusion matrix of proposed model.

Fig. 6. Accuracy scores of proposed Model.

the potential to revolutionize healthcare access and engagement in the modern digital age.

CRediT authorship contribution statement

Arun Babu: Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Sekhar Babu Boddu:** Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Chowdhury MNUR, Haque A, Soliman H. Chatbots: A game changer in mHealth. In: *2023 sixth international Symposium on Computer, Consumer and Control (IS3C)*. IEEE; 2023, June:362–366.

2. Safi Z, Abd-Alrazaq A, Khalifa M, Househ M. Technical aspects of developing chatbots for medical applications: scoping review. *J Med Internet Res*. 2020;22(12), e19127.

3. Topol E. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK; 2019.

4. Bao Q, Ni L, Liu J. HHH: an online medical chatbot system based on knowledge graph and hierarchical bi-directional attention. In: *Proceedings of the Australasian computer science week multiconference*. 2020, February:1–10.

5. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med*. 2023;388(13):1201–1208.

6. Battineni Gopi, Chintalapudi Nalini, Amenta Francesco. AI Chatbot design during an epidemic like the novel coronavirus. *Healthcare*. 2020;8(2):154.

7. Soufyane A, Abdelhakim BA, Ahmed MB. An intelligent chatbot using NLP and TF-IDF algorithm for text understanding applied to the medical field. In: *Emerging trends in ICT for sustainable development: The proceedings of NICE2020 international conference*. Cham: Springer International Publishing; 2021, January:3–10.

8. Athota L, Shukla VK, Pandey N, Rana A. Chatbot for healthcare system using artificial intelligence. In: *2020 8th International conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO)*. IEEE; 2020, June:619–622.

9. Ayanouz S, Abdelhakim BA, Benhmed M. A smart chatbot architecture based NLP and machine learning for health care assistance. In: *Proceedings of the 3rd international conference on networking, information systems & security*. 2020, March:1–6.

10. Vaira L, Bochicchio MA, Conte M, Casaluci FM, Melpignano A. MamaBot: a System based on ML and NLP for supporting Women and Families during Pregnancy. In: *Proceedings of the 22nd international database engineering & applications symposium*. 2018, June:273–277.

11. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *Jama*. 2016;315(6):551–552.

12. Shree R, Rastogi A, Kalaiarasan C. Machine learning-driven cutting-edge approach for designing a healthcare Chatbot. *Int J Intell Syst Appl Eng*. 2023;11(8s):198–205.

13. Fonna MR, Widyantoro DH. Tutorial system in learning activities through machine learning-based Chatbot applications in pharmacology education. In: *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*. IEEE; 2021, September:1–6.

14. Kim Y, Kim JH, Kim YM, Song S, Joo HJ. Predicting medical specialty from text based on a domain-specific pre-trained BERT. *Int J Med Inform*. 2023;170, 104956.

15. Thwala EKI, Adegun AA, Adigun MO. Self-assessment Chatbot for COVID-19 prognosis using deep learning-based natural language processing (NLP). In: *In 2023 international conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)*Vol. 1.

16. Chakraborty S, Paul H, Ghatak S, et al. An AI-based medical Chatbot model for infectious disease prediction. *Ieee Access*. 2022;10:128469–128483.

17. Tamizharasi B, Livingston LJ, Rajkumar S. Building a medical chatbot using support vector machine learning algorithm. *J Phys Conf Ser*. 2020, December;1716(1). p. 012059). IOP Publishing.

18. Badlani S, Aditya T, Dave M, Chaudhari S. Multilingual healthcare chatbot using machine learning. In: *2021 2nd International Conference For Emerging Technology (INCET)*. IEEE; 2021, May:1–6.

19. El Zini J, Rizk Y, Awad M, Antoun J. Towards a deep learning question-answering specialized chatbot for objective structured clinical examinations. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE; 2019, July:1–9.

20. Serban IV, Sankar C, Germain M, et al. A deep reinforcement learning chatbot. *arXiv Prepr*; 2017. arXiv:1709.02349.

21. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv Prepr*; 2020. arXiv:2005.14165.

22. Dinan E, Urbanek J, Szlam A, Kiela D, Weston J. TransferTransfo: a transfer learning approach for neural network based conversational agents. *arXiv Prepr*; 2019. arXiv:1901.08149.

23. Nie Y, Williams J, Dinan E, Weston J. Dialogue natural language inference. *arXiv Prepr*; 2020. arXiv:2005.07421.

24. Chen X, Qian J, Lu H, Zhu H. BERT for joint intent classification and slot filling. *arXiv Prepr*; 2019. arXiv:1902.10909.