BERT-MED Chatbot for Healthcare Assistance

Prajwal D R[1], Puneeth H K[2], Pola Manoj Kumar[3], Shreyank T N[4], Champa M S[5]

[1,2,3,4]*Dept. of Information science and Engineering, AMC Engineering college. Bengaluru, Karnataka, India*

[5]*Assistant Professor, Dept. of Information science and Engineering, AMC Engineering college. Bengaluru, Karnataka, India*

Emails: prajwalprajwalgowdar4@gmail.com[1], puneethhk1980@gmail.com[2], manojkumar.mk0965@gmail.com[3], shreyanks351@gmail.com[4], champams123@gmail.com[5]

Abstract

In recent years, the need for digital healthcare assistance has grown because of population increase, limited medical resources, and language differences. Many current medical chatbots are only available in English and offer limited support. This paper introduces BERT-MED, an AI-driven healthcare assistance system that follows templates. It includes a multilingual medical chatbot, a smart hospital finder, and a drug safety module.

The system uses transformer-based sentence embeddings to understand user medical questions and pull disease-related information from structured datasets. The chatbot allows both text and voice interactions and provides answers in multiple Indian languages. To help users understand better, the system offers educational YouTube videos about diseases. Furthermore, the smart hospital finder uses GPS-based location tracking, OpenStreetMap data, and crowd estimation to locate nearby healthcare facilities with risk indicators. The drug safety module gives information on medicine usage, dosage, warnings, and suggestions for similar pills. Experimental results show that this system provides accurate answers, quick information retrieval, and better usability than keyword-based methods. The BERT-MED system offers a scalable and accessible digital healthcare solution that can be deployed in the real world.

Keywords: BERT; Drug safety; Healthcare AI; Hospital finder; Medical chatbot

1. INTRODUCTION

1.1. Overview

Healthcare information access has become a key demand in today's society, especially when people with linguistic diversity face various challenges due to the geographical unavailability of medical infrastructure. People depend on less credible web sources or even delay reaching out for consultations in hospitals due to a lack of awareness and accessibility. Classic rule-based chatbots and keyword-driven search systems fail to understand natural language variations in medical queries, leading to responses that are incorrect.

Recent development in NLP with the help of transformer-based models has given machines the power of truly contextualized understanding. Building upon this, this paper presents BERT-MED, an AI chatbot in health care for disease information, medicine safety details, educational videos, and recommendations on nearby hospitals within a unified framework.

1.2. Objectives

Major goals of this project:

- The objectives of this paper are to present the design of an AI-based medical chatbot that shall be able to understand disease-related queries by using semantic similarity.
- multilingual text and voice-based interaction may support wide accessibility.
- The system should be able to provide medicine usage, dosage, and safety warnings through the provision of a drug safety module.
- Integrate a GPS-based hospital finder to locate nearby healthcare facilities.
- This will help in improving the understanding of the users by utilizing disease-related educational YouTube videos.

1.3. Purpose, Scope, and Applicability

1.3.1. Purpose

This project shall undertake the development of an intelligent healthcare assistance system with the objective of filling in the gap between users and basic medical information. By integrating AI-based language understanding with geolocation services, the system shall be able to offer credible preliminary guidance while referring the issues for professional medical consultation.

1.3.2. Scope

The scope of the proposed system includes:

- Retrieval of information about diseases using the semantic search.
- Multimodal and voice interaction.
- Drug safety information and alternative medicine suggestions.
- Location-based services enable the identification of a place like a hospital.
- It does not diagnose nor perform medical analysis based on images.

1.3.3. Applicability

The BERT-MED AI-based multilingual healthcare assistance system can be very well applied within a wide range of domains where accessible, reliable, and real-time healthcare information is crucial. Its modular design and adoptability make it fitting for the following areas:

- **Rural and Remote Healthcare Services:** Providing basic medical guidance, disease awareness, and hospital location support to users in rural areas with limited access to healthcare professionals.
- **Hospitals and Clinics:** Assisting patients with preliminary disease information, medicine usage awareness, and locating nearby healthcare facilities, thereby reducing the burden on frontline medical staff.
- **Telemedicine Platforms**: Enhancing virtual healthcare services by integrating AI-driven medical chat, drug safety information, and multilingual voice support for remote consultations.
- **Public Health Awareness Systems:** Supporting government and non-government organizations in spreading disease awareness, preventive measures, and emergency signs during health campaigns and outbreaks.
- **Educational Institutions:** Serving as a learning tool for students and healthcare trainees to understand disease symptoms, treatments, and medicine safety through interactive and visual learning content.
- **Mobile and Web Health Applications:** Embedding the system into health apps to provide on-demand medical information, voice-based interaction, and location-based hospital discovery.
- **Emergency and First-Aid Support Systems:** Offering quick access to emergency signs, nearby hospitals, and precautionary measures during medical emergencies before professional help arrives.
-

2. LITERATURE REVIEW

2.1. Introduction

A literature survey involves a systematic examination and evaluation of existing scholarly works such as journals, conference papers, and research articles that concern a specific domain. The literature review for this study is about understanding how AI and NLP are changing healthcare assistance systems, mainly on medical chatbots, semantic understandings, multilingual interaction, and accessibility to healthcare.

2.2. Summary of Literature Survey

The main goal of this review paper is to analyze the role of AI and deep learning methods for upgrading healthcare information systems by allowing valid interpretation of medical queries and providing appropriate responses in real time. This survey presented several AI-driven medical chatbot models, transformer-based NLP approaches, and multilingual healthcare systems proposed in the recent literature.

It is shown by existing literature that transformer-based models, like BERT, have much better performance than classic keyword-based approaches in handling medical terminologies and contextual language variabilities. However, most of these are not integrated with drug safety information, hospital discovery, and multi-media educational support. This sets the stage for proposing an integrated healthcare assistance system which integrates semantic NLP, multi-lingual support, and real-world health-care services.

2.3. Key Technologies Reviewed

- **Artificial Intelligence Techniques:** Supervised learning, semantic similarity models, and contextual language representations are used in healthcare dialogue systems.
- **Deep Learning Models:** transformer-based architectures, such as BERT and Sentence-BERT for the comprehension of medical queries and descriptions of diseases.
- **Multilingual NLP Systems:** Translation frameworks and speech-based interfaces allowing access to healthcare across diverse linguistic populations.
- **Details on Health care chatbot**

frameworks: Rule-based, retrieval-based, and AI-driven conversational agents for medical information delivery.

- **Geospatial Healthcare Technologies:** GPS-based systems and OpenStreetMap integration for locating nearby hospitals and healthcare facilities.

2.4. Drawbacks

- Limited contextual understanding in keyword-based chatbots
- Difficulty in multi-lingual and voice-based interaction
- Lack of modules about drug safety and medicine awareness
- Poor integration of the discovery of hospitals along with real-time assistance

3. PROBLEM STATEMENT

The limitations of traditional medical assistance systems have come to the fore with ever-increasing demand for easily accessible healthcare information. While developing healthcare chatbots, most of them still depend on keyword matching and static rules, failing to accurately capture diverse user queries and medical terminology. These generally lack support in multiple languages, drug safety awareness, and real-time location of hospitals.

Insufficient health care access in these regions automatically makes users dependent on the highly unreliable online sources of information, turning misinformation into a reason for delayed medical interventions. Even though AI-based medical chatbots sound very promising, the majority of them find it difficult to master contextual understanding, scalability, and practical integration of real-world healthcare services. There is a significant gap in research on the development of an AI-driven healthcare assistance system that will correctly understand the queries regarding medication, provide reliable information about drugs, support multiple languages, and guide users towards nearby healthcare facilities in real time.

4. RELATED WORKS

Healthcare chatbots have improved from rule-based systems to deep learning models that are capable of learning natural language. The previous systems used rules, which sometimes led to incorrect responses. There have been recent experiments that have shown that models such as BERT are useful in enhancing the level of contextual comprehension in a healthcare chatbot (Reimers & Gurevych, 2019).

The existing research on multilingual NLP in the healthcare sector focuses on the need to support local languages to facilitate widespread acceptance. In addition, geospatial healthcare solutions have also been developed to help users identify nearby healthcare facilities with the use of open mapping solutions. The tendency, though, is for current solutions to address isolated functionality, with an overall lack of a comprehensive solution that incorporates all aspects of healthcare, chatbots, hospital location, and medication safety. The solution fills that void.

5. PROPOSED SOLUTIONS

Beyond the limitations of traditional keyword-based healthcare chatbots, we propose a unified AI-based framework for healthcare assistance, BERT-MED. Unlike the conventional systems that depend on pre-defined rules and responses, this proposed system would use semantic comprehension through transformer-based models to understand the medical queries more accurately. The system is adaptable to variations in medical terminology and user language for better reliability and usability.

The proposed framework comprises three interconnected phases that together provide comprehensive healthcare assistance:

- **AI-Based Medical Query Understanding:** This phase uses SentenceTransformer-based semantic embeddings to contextualize disease-related user queries in real time. By understanding context rather than keywords, this system pins relevant diseases, symptoms, treatments, precautions, and emergency signs with greater accuracy. This makes the system capable of handling informal queries and out-of-sight language variations that conventional systems usually fail to interpret.
- **Drug Safety and Visual Aids:** Once a disease or medication is identified, it retrieves information on medicine usage, dosage, and warning from structured datasets and external data sources like the OpenFDA API. Additionally, educational disease-based YouTube videos are provided to give users better clarity through visual learning. This helps users gain better awareness while avoiding unsafe medication practices.
- **Hospital Finder and Location-Based Support:** In this phase, the system detects nearby hospitals, clinics, and pharmacies

using GPS-based geolocation and OpenStreetMap data. Also, it computes crowd index, risk levels, and estimated waiting times to help the user with the best choice of healthcare facilities, especially when there is an emergency. This proactive guidance reduces delays and improves healthcare access.

6. STSTEM ARCHITECTURE

Table 1 is a complete architecture design for the BERT-MED system. The user interacts with the application using a web application interface, which is capable of handling text as well as voice inputs. The back-end component of the application uses a semantic embedding model to look for corresponding information from various databases, including the medical, pill, and hospital databases. Even services like Open Street Map are used.

Table 1 presents the overall System Architecture of the BERT-MED Application The BERT-MED system follows a modular client–server architecture implemented as a web-based application. The system is organized into three primary modules: Medical Chatbot, Hospital Finder, and Drug Safety. The frontend provides an interactive user interface for text and voice input, while the backend manages semantic processing, data retrieval, and response generation.

Table 1. Overall System Architecture of the BERT-MED Application

Module	Description	Key Technologies
Medical Chatbot	Disease query handling	Sentence Transformer
Multilingual & Voice	Multi-language & Speech	Google TranslatorgTTs
Visual Assistance	Educational Videos	YouTube S
Hospital Finder	Nearby hospitals &risks	GPS, OpenStre
Drug Safety	Medicine info & Safety	CSV, OpenFDA

6.1. Medical Chatbot Module

The medical chatbot module uses a pre-trained SentenceTransformer model to generate semantic embeddings for user queries. Disease names stored in the medical dataset are encoded offline. During runtime, cosine similarity is computed between the query embedding and disease embeddings to identify the most relevant disease. Once a match is found, the system retrieves comprehensive disease-related information, including symptoms, treatment, precautions, home remedies, diet recommendations, doctor advice, emergency signs, and an AI-generated summary.

6.2. Multilingual and Voice support

In this respect, the system will provide multilingual translation services for a number of Indian languages. Users may communicate with the chatbot by either text or voice input. Text-to-speech functionality uses the responses to create audible messages, therefore allowing hands-free use and being more convenient for the visually impaired and those with low literacy levels.

6.3. Visual Aid using YouTube Videos

The chatbot also has visual assistance, which provides educational YouTube videos related to the disease. The videos load dynamically with the name of the disease and only then can be viewed by the users. The visual content is for awareness and educational purposes only, not to substitute professional medical advice.

6.4. Smart Hospital Finder Module

It helps the user to find nearby hospitals, clinics, and pharmacies using GPS-based location detection or manually entered addresses. Healthcare facility data is downloaded from OpenStreetMap using Overpass API queries. Distances are calculated by geodesic formulas. The system makes an estimation of crowd level and risk categories by heuristic scoring depending on the type of facility, location, and time of day. The results are visualized on an interactive map with color-coded markers and heatmaps.

6.5. Drug Safety Module

The module of drug safety will allow users to search for medicines and retrieve information including usage, dosage, warnings, and precaution. This system will first search a local pill dataset and fall back to publicly available APIs about drug information if needed. It uses semantic similarity and fuzzy matching techniques to handle spelling variation and partial queries. This module also suggests similar pills based on semantic similarities among medicine names.

7. METHODOLOGY

The methodology of the proposed BERT-MED system is structured to ensure the delivery of accurate medical information, effective awareness in drug safety, and dependable discovery in hospitals.

- **Data Collection:** It collects medical disease data, pill information, and hospital location data from various CSV datasets. The datasets include disease names, symptoms, treatments, precautions, medicine usage, dosage, and warning information.
- **Preprocessing of Data:** The collected data is cleaned and structured by removing missing values, normalizing text fields, and standardizing column formats to ensure that it's consistent and improves retrieval accuracy.
- **Semantic Model Development:** First, disease names and medicine names are embedded contextually using SentenceTransformer. These representations enable semantic similarity matching, instead of keyword comparison.
- **Real-time Query Processing**: User queries are processed in runtime and matched against the stored embeddings using cosine similarity to retrieve the most relevant disease or medicine.
- **The finest Multilingual and Voice Processing:** Google Translator currently translates the responses into the chosen language, and gTTS allows the text-to-speech mechanism for voice output in cases of readability inconvenience.
- **Discovery and Risk Estimation:** In a Hospital Healthcare facilities within proximity are identified using GPS-based user location and OpenStreetMap data. Heuristic-based crowd index and risk levels are computed to estimate waiting time and hospital load.
- **Testing and Deployment:** The system is tested through real-time user interactions and multiple healthcare scenarios to evaluate the response accuracy, usability, and performance before the deployment.

Figure 1: Flow chart describing the processing of healthcare data to generate a response.

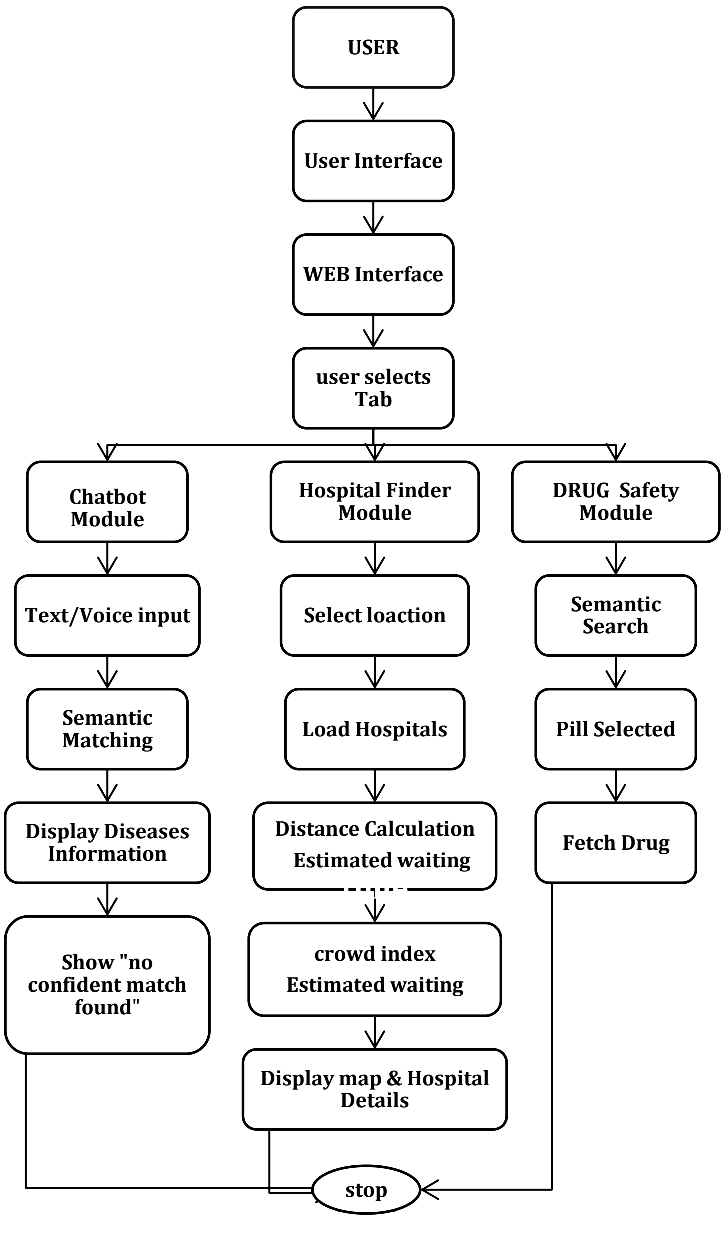

Figure 1. Workflow diagram of the proposed BERT-MED AI-based healthcare assistance system

8. WORKING

The functioning of BERT-MED, the healthcare support system, starts with the submission of a query by a user through text or audio input using the web interface. This is followed by the processing of the query into a semantic form using SentenceTransformer embeddings, allowing for context understanding of queries about diseases, symptoms, or medications.

After accomplishing the semantic analysis, the system takes various processing routes in accordance with the type of query:

- **Medical Chatbot Processing:** If it is related to diseases, the system tries to match the given query with the stored embeddings of diseases using cosine similarities. Once the closest disease is found, it tries to fetch information like symptoms, treatment, cautions, home remedies, diets, advice from doctors, emergency signs, and an artificial intelligence summary related to that particular disease.
- Drug/Pill Safety Processing: When searching for information related to medicine, the system calls upon the pills dataset to gather instructions for using medicines, their dosage levels, or cautions about their use. If such information is not found, it is gathered through verified sources such as the OpenFDA API. Moreover, semantic similarities enable the system to suggest other pills that could work similarly.
- Visual Assistance Module: To make it easier for users to understand their data, the system offers educational YouTube videos on the respective diseases. These videos provide information about the symptoms, causes, and preventive measures for diseases and are retrieved dynamically using the name of the respective disease. This helps in understanding the data in a better manner, which would otherwise be difficult for laymen or people who lack medical literacy.
- Multilingualism & Voice Interaction: All information retrieved can be translated into the language chosen by the user through translation services. This system also produces voice outputs by using text-to-speech technology so that users can hear information about their health in their preferred language.
- Hospital Finder Operation: If the user allows the application to access their location, it allows the system to search for nearby hospitals, clinics, or pharmacies using geolocation through GPS. This is followed by the calculation of crowd-index, waiting times, or risk levels by using heuristics, which is then displayed on the map to help the user choose their destinations.

Finally, after completing the response, the system is back in the idle state, allowing for endless interactions. Such an integrated workflow is one that would showcase the merging of semantic AI, drug safety consciousness, visual learning, or geospatial support for healthcare into one healthcare support service.

9. RESULT AND DISCUSSIONS

9.1. Results

The proposed system was evaluated through functional testing and real-time interaction scenarios. The medical chatbot demonstrated appropriate semantic matching for common disease queries and handled a few variations in user input well. Embedding-based methods can produce more reliable and relevant results compared to keyword-based methods.

9.2. Discussions

This greatly improved accessibility with the multilingual and voice-enabled interface. Integration of YouTube-based educational videos improved user understanding of medical conditions. The drug safety module provided consistent medicine information and useful recommendations for similar pills. The hospital finder module identified nearby facilities and presented meaningful crowd and risk indicators to support informed decision-making.

CONCLUSION

This paper introduced BERT-MED, a template-compliant AI-assisted healthcare support system which comprises a multilingual healthcare chatbot, a smart hospital locator, and a medication safety component. The system uses transformer-based semantic embeddings, multilingual translation, and geospatial functionalities to provide precise, accessible, and usable healthcare support. The addition of educational videos on various diseases improves user awareness. The proposed system has proven that NLP, geolocation, and multimedia assistance are highly useful in designing scalable healthcare support systems

ACKNOWLEDGEMENT

The authors are thankful to AMC Engineering College, Bengaluru, for providing them with the necessary infrastructure. The support from the faculty members of the Department of Information Science & Engineering is also thankfully acknowledged.

REFERENCES

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 4171–4186.

[2]. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP

[3]. Pap, S., & Oniga, F. (2024). Multilingual medical dialogue systems using transformer-based models. IEEE Transactions on Affective Computing, in press.

[4] Sharma, P. & Mehta, K. (2024). Geospatial intelligence for healthcare facility discovery and accessibility analysis. *IEEE Access*, 12, 155982–155994

[5] OpenStreetMap Contributors, (2024). OpenStreetMap: The free wiki world map. Retrieved from https://

[6]. U.S. Food and Drug Administration. (2024). OpenFDA drug label API documentation. Retrieved from https://open.fda.gov

[7]. Google LLC. (2024). Google Text-to-Speech (gTTS) documentation.

[8.] Birari, H. P., Lohar, G. V., & Joshi, S. L. (2023). Advancements in machine vision for automated inspection of assembly parts: A comprehensive review. International Research Journal on Advanced Science Hub, 5(10), 365–371. https://doi.org/10

[9] Rajan, P., Devi, A., B, A., Dusthackeer, A., & Iyer, P. (2023). A green perspective on the ability of nanomedicine to inhibit tuberculosis and lung cancer. International Research Journal on Advanced Science Hub, 5(11), 389–396. https://doi.org/10

[10] Keerthivasan, S. P., & Saranya, N. (2023). Acute leukemia detection using deep learning techniques. International Research Journal on Advanced Science Hub, 5(10), 372–381.