Y.Prasanna, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

CareBot: A Medical Chatbot using NLP

Y.Prasanna, Kasula Vijay Vardhan , Kuchangari Sahit, Gaddameedi Pranay Kumar
Department of Information Technology Gokaraju Rangaraju Institute of Engineering and Technology Hyderabad, India

Abstract- A chatbot, or conversational agent, is an AI-powered software designed to communicate with users using natural language. In the healthcare domain, medical chatbots play a crucial role in providing preliminary health guidance, answering medical queries, and assisting patients with symptom analysis. One of the major challenges in developing a medical chatbot is designing an effective dialogue system that can accurately understand user input and provide relevant responses. Early chatbot models relied on rule-based systems and statistical methods, which had limited flexibility in handling diverse conversations. However, with advancements in artificial intelligence and natural language processing (NLP), deep learning techniques, particularly end-to-end neural networks, have revolutionized chatbot development. Since 2015, encoder-decoder recurrent models, originally designed for neural machine translation, have become dominant in conversational AI due to their ability to generate meaningful and context-aware responses. This project aims to develop a medical chatbot using NLP techniques to enhance healthcare accessibility. The chatbot will leverage deep learning models to analyze user queries, understand symptoms, and provide preliminary advice based on medical knowledge. By integrating state-of-the-art advancements in NLP, this system seeks to offer accurate, reliable, and interactive healthcare assistance, improving user experience and bridging the gap between patients and medical professionals.

Keywords: ML Algorithms, cross-validation, feature selection, tokenization, stemming, F1-score, Accuracy.

I. INTRODUCTION

CareBot is an intelligent virtual health assistant that ensures the users can access timely, easy to access, and customized healthcare information. With Python as its language and with Streamlit playing a role in its deployment, CareBot has been adapted with natural language processing (NLP) to be able to comprehend symptoms described by users and offer accurate suggestions about likely diseases and suggested treatments. It also provides wellness advice fitting the user's particular health concerns, thus promoting a holistic approach to well being.

The CareBot is supposed to help the users by identifying potential needs for health care and give practical advice that would help them to improve their well-being early. Although it cannot act as a substitute for professional medical advice, it is the first point of reference users can go to get health related information while at home.

CareBot aims at reducing the gap between individuals and healthcare knowledge through the use of AI symptom detection, context-aware advice, multilingual support especially in regions that do not have access to medical experts.

The general objective of this project is to develop the perception of medically unrelated issues using the complex AI technologies. Overcoming misconceptions and implementable holistic care among varied groups accrues one's health literacy.

II. LITERATURE REFERENCES

[1] There is a webMD symptom checker online, free people can use it by entering certain symptoms and get a list of possible conditions according to the predefined questions. Keen as they are WebMD's rules-based systems limit flexibility and accuracy of responses perhaps as users can only select from symptoms list that may relegate responses to ill-defined or complicated health conditions.

[2] Like with WebMD, Mayo Clinic's symptom checker first gives them a list of symptoms to select from, after which it returns to them potential diagnoses. Although, it has medically reviewed information, the format remains structured with limitation to users in describing symptoms and issues without a set answer. The system is more people focused on medical and it does not offer personalized wellness/lifestyle guidance.

[3] Buoy Health has a chatty AI model in the shape of a virtual health assistant. It is set up to ask users questions concerning their symptoms, and offer them probable diagnoses. The system also gives advice on whether users require medical attention. Even though conversational approach from the part of Buoy Health is ebbing the industry to the more personalized interactions, the company focuses on diagnosis and does not have a support across the broader aspects of lifestyle management and wellness advice.

[4] Babylon Health uses AI to offer remote diagnostics, symptom checking and health advice through a bot. It is multilingual and that makes it easier to reach a global audience. Babylon's system is a personalized health insight system, but it focuses relatively on symptom triage and fails to give personalized wellness tips such as CareBot.

III. METHODOLOGIES

The CareBot project adopts a systematic approach to offering customized health support. The user first enters a query related to health in a text box that may be in any language. This input is converted to English with the help of the Deep Translator library to make processing uniform. With the aid of a pre-trained SentenceTransformer model, the system translates the input into a semantic embedding and matches it against symptom information from a health database in order to find the most similar record.

System Architecture

The CareBot project architecture is a modular, scalable, and user-focused system that combines various important components to provide intelligent health support. At the heart of the system lies a Natural Language Processing (NLP) engine fueled by transformer-based models, which allows the application to interpret and comprehend user requests in natural, conversational language. When a user submits a symptom or health concern through the Streamlit interface, the input is processed and inserted into vector form, then matched against a pre-encoded list of symptoms based on semantic similarity to find the closest match. This list, in CSV format, is a repository of disease names, connected symptoms and is the knowledge base for the system's medical replies.

Data Preprocessing

In the project for CareBot, preprocessing the data is essential in obtaining accurate and meaningful results through analysis. The health data, which includes disease names, symptoms, and remedies, is initially imported using Pandas with assigned column names for uniformity. To clean the text for semantic comparison, both user input and symptom data are normalized through conversion to lowercase, which assists in circumventing case-sensitive mismatches while detecting keywords. The query entered by the user is also translated into English via Deep Translator for compatibility with the pre-trained SentenceTransformer model. The model transforms both the symptoms of the dataset and the user input translated into English into semantic embeddings, facilitating similarity comparison effectively. A similarity threshold is also used to exclude weak matches and provide accurate disease identification. Lastly, the results, such as the

diagnosis and customized health advice, are rendered back into the user's chosen language, keeping accessibility multi-lingual.

Model Selection and Training

Care was taken in selecting models in the CareBot project to ensure the system would have the ability to properly understand and answer a vast number of user health questions. The all-MiniLM-L6-v2 model from SentenceTransformers was selected because of its great trade-off between speed, performance, and computational complexity. This is a pre-trained transformer-based architecture specifically tailored to producing high-quality sentence embeddings capable of capturing semantic content in text. It was chosen over heavier versions such as BERT or RoBERTa in order to keep the model lightweight and responsive, particularly critical for real-time use such as in a health assistant. Because all-MiniLM-L6-v2 is already fine-tuned for semantic similarity tasks on large-scale natural language datasets, there was no further training or fine-tuning required in the project. In this case, the model is applied directly to encode both the user's translated input and the symptom data from the health dataset in dense vector representations. These embeddings are compared with cosine similarity in order to determine the most semantically relevant symptom match, which serves as the foundation for providing a potential disease suggestion and recommended treatment. This approach not only guarantees precision in comprehension of diverse health issues but also drastically minimizes development time as well as computational overhead.

User Interface and Interaction

All interactions—including dataset uploads, preprocessing, model selection, and results display—are managed through a Graphical User Interface (GUI) built with Streamlit. Key features include:

- The user interface (UI) of the CareBot project is thoughtfully designed using Streamlit to provide a clean, responsive, and user-friendly experience.
- It features a dark theme with custom CSS styling for elements such as chat bubbles, buttons, and health tips, giving it a modern and professional appearance.
- Additionally, the interface includes a section for personalized health tips.

IV. RESULTS

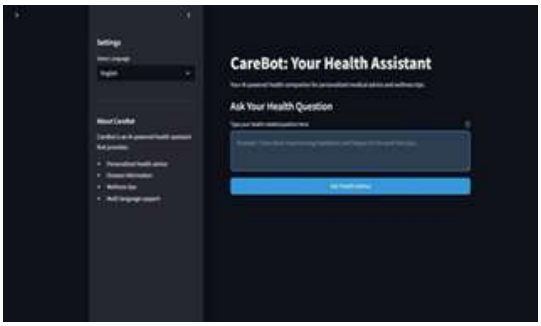

FIG-1. User Interface

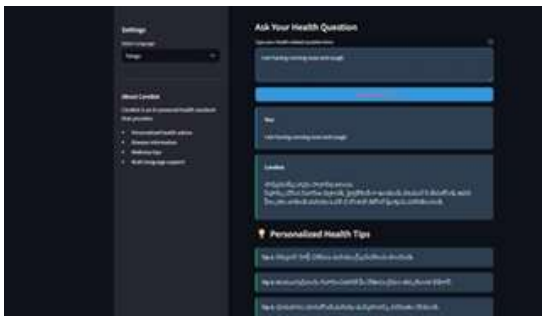

FIG-2. Response in another language

V. CONCLUSION

The CareBot project effectively illustrates the possibility of using artificial intelligence, natural language processing, and semantic similarity methods to create a friendly, multilingual health assistant. By enabling users to explain their symptoms in natural language and return relevant disease recommendations along with customized wellness advice, CareBot bridges the gap between conventional symptom checkers and conversational AI interfaces. The system supports multiple languages, targeting users with a variety of linguistic backgrounds, and focuses not only on disease identification but also on lifestyle enhancement

through health advice. Transformer-based models improve symptom interpretation accuracy and versatility, and Streamlit offers a visually friendly and easy-to-use interface. CareBot is generally a comprehensive and expandable solution for daily health queries and wellness assistance

VI. FUTURE PROSPECTS

Voice-Based Interaction:
Adding speech recognition and natural language processing to enable users to interact with CareBot by voice command. This would enhance the usability and convenience of the system, particularly for users who find it hard to type, like seniors or disabled people. Voice interaction can also facilitate hands-free operation, which can improve user experience in different environments.

Deep Learning Model Enhancement:
Subsequent releases may take advantage of advanced machine learning and deep learning algorithms driven by large-scale, heterogeneous medical data sets. This would enhance the system's capability to interpret intricate patterns of symptoms, distinguish between sets of similar diseases, and make more precise, tailored diagnostic recommendations and treatment plans..

Mobile Application Development:
Developing native mobile apps for iOS and Android devices to expand accessibility and ease of use. Mobile apps might leverage device capabilities like push notifications, location services, and integration with mobile health sensors to provide a more engaging user experience

Integration with Wearable Devices and IoT:
Integrating CareBot with wearable health sensors (such as smartwatches, fitness trackers, and biosensors) to gather real-time physiological measures such as heart rate, sleep quality, activity level, and blood oxygen. This stream of data potentially could lead to more accurate health monitoring and personalized advice based on real user health trends.

REFERENCES

1. BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. (2024). ScienceDirect.
2. MediBERT: A Medical Chatbot Built Using KeyBERT, BioBERT, and GPT-2 Hossain, S., Sharar, R., Bahadur, M. I., Sufian, A., & Nabil, R. H. (2023). International Journal of Intelligent Systems and Applications.
3. HealthCare Chatbot Using NLP Techniques and Deep Learning Algorithm. Authors not specified (2023x)
4. KG-MTT-BERT: Knowledge Graph Enhanced BERT for Multi-Type Medical Text Classification He, Y., Wang, C., Zhang, S., Li, N., Li, Z., & Zeng, Z. (2022). arXiv preprint arXiv:2210.03970
5. Med-Bot: An AI-Powered Assistant to Provide Accurate and Reliable Medical Information Ahan Bhatt, Nandan Vaghela (2024)
6. Contributions to the study of SMS spam filtering: New collection and results. -Almeida, T., Hidalgo, J. M & Yamakami, A. (2011)
7. SMS Spam Detection Using Non-Content Features – HKUST https://cse.hkust.edu.hk › ~qyang › Docs › ieee_i