SmartDiagnosis: An ML-Powered Chatbot for Proactive Medical Screening

Nandini G R
Computer Science and Engineering,
R L Jalappa Institute of Technology,
Doddaballapur,India-561203

Dr. Manjunatha B N
Associate proffesor & HOD
R L Jalappa Institute of Technology,
Doddaballapur,India-561203

Abstract- *This study discusses the development and deployment of a medical chatbot that can detect infectious diseases and provide medical assistance. To correctly respond to customer inquiries, the chatbot makes use of Machine Learning (ML) techniques such as Support Vector Machine (SVM), Long Short-Term Memory (LSTM) networks, and Natural Language Processing (NLP). Additionally, it talks about that uses machine learning algorithms—specifically logistic regression—to predict a number of diseases, including diabetes, cancer, hypothyroidism, heart disease, and Parkinson's diseases. The purpose of this research is to use machine learning techniques to develop a predictive model that can recognize potential illnesses based on symptoms.. The model is then smoothly incorporated into a chatbot that functions as a platform for communication between users and the system and is a text-based disease prediction aid. The purpose of this integration is to give consumers with a seamless, personalized healthcare experience. The study's goal is to present a unique way for early disease prediction and tailored healthcare advice that is available via an easy-to-use interface.*

Keyword:*Healthcare Chatbot, Disease Prediction, Machine Learning, Natural Language Processing (NLP), Support Vector Machine (SVM), Long Short-Term Memory (LSTM)*

1.Introduction

Artificial intelligence (AI) and machine learning (ML) have significantly advanced technologies, and quick development has led to important breakthroughs in a number of sectors, especially healthcare. Chatbots driven by AI are being used more and more to enhance disease diagnosis, automate answers to patient inquiries, and offer medical aid. These chatbots evaluate user input and produce precise responses by utilizing deep learning and natural language processing (NLP) techniques. The goal of integrating such intelligent technologies into healthcare is to increase accessibility, lessen the workload for medical staff, and provide real-time assistance to people in need of medical information[1]. The COVID-19 pandemic has placed the entire world on high alert for potential public health threats since the end of 2019. Even worse, while the COVID-19 pandemic is still going strong, a second public health concern—monkeypox—was proclaimed a global health emergency in July 2022. These subsequent public healthA community that is still recuperating from the early effects has been severely strained by threats. However, these crises have enhanced public awareness of health issues, particularly the significance of early symptom detection and prompt treatment. This poses a considerable barrier in assessing.

potential illnesses based on specific symptoms[5]. However, it is difficult to provide efficient healthcare solutions because existing healthcare systems might result in lengthier patient wait times, delayed diagnosis, and 38 a general lack of physician availability.

Chatbots with AI capabilities serve as remote healthcare aides, offering automated front-end evaluations, which has improved early disease diagnosis and lessened the burden on healthcare providers[2]. A potential answer is provided by machine learning methods [7]–[8]. These methods can use black-and-white data to forecast diseases with accuracy. Machine learning is free and less susceptible to personal views. The machine learning model will function similarly to a doctor's brain while Communicating with the patient is facilitated by the chatbot [14]–[9]. In order to gather the required information and steer the conversation in the proper direction, a chatbot can be useful [10]–[13]. Thus, by integrating the chatbot and disease prediction model, users can use this disease prediction helper to

get free assistance in determining their potential diseases.Recent developments in natural language processing and deep learning have made it possible to create complex chatbot models that can comprehend and react to medical questions with a great degree of precision. Text-based interactions can benefit from the use of Long Short-Term Memory (LSTM) networks, a subset of recurrent neural networks (RNNs), which have shown exceptional efficacy at processing sequential input. Unlike traditional rule-based chatbots, LSTM-based models may learn from large amounts of textual input, resulting in more understanding responses over time. This ability is essential since medical chatbots must be able to comprehend user queries and provide relevant information without human assistance [1].

This study examines the creation and application of chatbots to offer health-related guidance, data, and initial diagnosis. The study's chatbot could imitate a primary care session by answering user inquiries about symptoms, nutrition, and lifestyle. The chatbot becomes more approachable and user-friendly by accepting text and vocal input, demonstrating how it may assist with health management by providing individualized health information[3]. The rest of the paper is structured as follows: A review of recent research in the same field is provided in Section 2. The various methods utilized for predicting the disease from various datasets were described in Section 3. and a conclusion is drawn in Section 4.

2. LINKED WORK

There have been many research on this subject, but in the sections that follow, we will concentrate on the most important ones. Deep learning methods were examined in a study by Rajkomar et al. [18] to forecast patient outcomes for a range of medical disorders. They used data from electronic health records (EHRs) and discovered that their predictive models beat traditional statistical techniques. This study underlines the significance of selecting the correct attributes as well as the necessary for massive datasets in order to construct effective models. However, a study conducted by Kaur et al. [15] examined straightforward algorithms for predicting diseases including diabetes and cardiovascular ailments, such as logistic regression and decision trees. According to their findings, simpler models are frequently easier to use and more interpretable in clinical situations, particularly when data is limited, even when more sophisticated models may attain higher accuracy.

When creating the HELPI chatbot (Karuna et al., 2023), the data collection "The Disease Symptom Prediction Dataset," which is accessible on Kaggle, has 133 columns and roughly 5000 rows (the first 132 reflect thesymptoms, and the final one is the illness), with a total of 42 potential illnesses (diagnoses). Even in this instance, no metrics pertaining to the model's performance are provided. The diagnosis is predicted using decision trees and the CART (Classification and Regression Trees) algorithm.Diabot, a chatbot designed especially for diabetes prediction based on user-input symptoms, is the next model examined (Bali et al., 2019). The RASA framework is utilized in the backend, and an ensemble learning technique is used to solve the classification problem (disease presence/absence.Chatbots are now a vital tool for communicating with and helping patients with their medical needs. A chatbot for mental health help was created by Denecke et al. and provided resources for users, coping mechanisms, and evaluations derived from their contributions. This project showed how chatbots may enhance patient outcomes and offer tailored advice.

According to research by Clark et al., a chatbot that helps patients understand their risk factors and leads them through preventative steps based on anticipated health implications was examined in terms of disease prediction. This study demonstrates how chatbots can support predictive models by improving user engagement and offering useful insights, encouraging a proactive approach to health care. The authors of the November 2018 research by Lokman, A. S., & Ameedeen, M. A. examined subjects related to the chatbot's knowledge domain, answer creation, and textissues like processing, machine learning models, and the use of datasets and evaluation techniques. The authors of the 2019 paper by Kumar, A., Meena, P. K., Panda, D., & Sangeetha, M. detailed how to use Artificial Intelligence Markup Language (AIML) and Latent Semantic Analysis (LSA) with the Python platform to create a chatbot that can respond to any kind of query with an authentic and accurate response. The author discusses many methods for developing chatbots in the paper Raj, S., Raj, & Karkal, 2019, providing important details on chatbot integration. and deployment, and sets up a chatbot

using machine learning. The authors of the study by Tebenkov, E., & Prokhorov, I., 2021, talked about the data analysis method, which enables the analytical system to learn by using comparable techniques in its working process and by addressing difficulties.

3. METHODOLOGY

The three primary steps of this medical chatbot's methodology include data collecting,preprocessing, building the machine learning model, integrating the chatbot, and evaluation. Each stage is critical to ensuring that the chatbot gives accurate illness forecasts and trustworthy medical advice.The steps in the system development process are as follows:

3.1. Data Collection and Preprocessing

A well-structured medical dataset is necessary to build a successful chatbot. This entails compiling data from a variety of sources, such as publicly available datasets (e.g., WHO, CDC, Kaggle), patient medical records, and expert discussions with medical specialists to verify the data. Furthermore, knowledge from current AI-based healthcare Models aid in improving chatbot performance. Before being used, the data is preprocessed to ensure accuracy and enhance quality. Tokenization and lemmatization (dividing text into meaningful components for NLP processing), feature extraction (turning symptoms into numerical data for machine learning), data normalization (standardizing symptoms for improved prediction accuracy), and data cleaning (removing irrelevant and duplicate entries, handling missing values) are all included in this.

3.2. Natural Language Processing (NLP) for Chatbot Interaction

The primary function of an NLP system in a healthcare chatbot is to interact and transact with users in natural language. After registering the user's message, the initial objective of NLP is to clean and process the text, deleting extraneous words or turning them into a standard form. Finally, intent recognition determines what the user was actually asking for: inquiring about symptoms, requests for treatment information, or checking on precautionary details. Then, Named Entity Recognition is utilized to locate important medical terms that relate to the query such as symptom, disease, and medicine. Using this information, the chatbot queries for medical information in its data bank or prediction for an accurate and helpful response; It might also consider the earlier discussion. This allows the chatbot to give relevant health advice while gently advising the user to consult a medical practitioner for serious issues.

3.3.Machine Learning-Based Disease Prediction Model

A chatbot in a machine learning-based disease prediction model is made possible by an intelligent conversational agent that serves as an interface between the user and the disease prediction system. The chatbot uses Natural Language Processing (NLP) techniques to clean and extract pertinent medical terminology like symptoms, duration, and severity from the user's natural language input of his symptoms. The medical parameters are then sent into the machine learning engine, which has been educated with a massive amount of medical data. The model determines the likelihood of a disease by analyzing the symptoms using algorithms like Support Vector Machine (SVM), Decision Trees, or deep learning models like LSTM.After making the prediction, the chatbot creates a succinct, understandable response outlining the potential illnesses and recommending additional actions, namely consulting a physician. The chatbot then keeps learning from feedback and interactions, becoming more accurate and tailored. Convenient and effective health advice is made possible by this blend of chatbots and machine learning.

3.4. LSTM

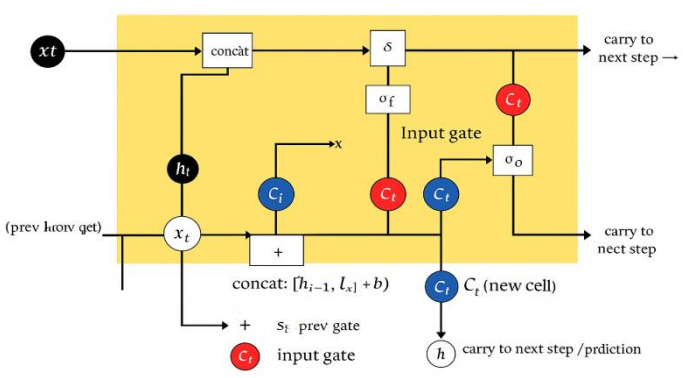

Figure 1: Long short-term memory (LSTM) architecture

Long short-term memory (LSTM) fig(1) is a special type of recurrent neural network designed to process sequential data and keep information over time. In a chatbot focused on health, LSTM is useful because it examines conversation history and understands symptom descriptions in order. For example, "I have a headache for two days, and now fever." Unlike traditional models, LSTMs have a unique structure with memory cells and three gates. What new data the cell can store is determined by the input gate. When data is no longer required, the forget gate decides what should be removed from the cell. What information the cell should transmit on next is decided by the output gate. LSTMs remember context through multiple conversation turns. This allows important symptoms and user inputs to be recalled accurately. As a result, LSTMs improve chatbot capabilities for providing context-aware responses regarding disease prediction and medical advice.

3.5. SVM

Support Vector Machine (SVM) is used to classify user-reported symptoms into possible disease categories. When a user enters symptoms, the chatbot first applies Natural Language Processing (NLP) to extract relevant medical terms and convert them into a numerical feature vector. The system then passes these attributes to the SVM model, which has been trained on a labeled medical dataset. The SVM method identifies the optimum hyperplane that divides distinct illness classes based on symptom patterns. For example, if a user reports fever, cough, and weariness, the SVM model can determine if the symptoms are due to influenza, COVID-19, or another condition. Since SVM works well with high-dimensional data and small datasets, it is a good fit for healthcare situations where correct classification is vital. The chatbot then uses the predicted class to offer guidance, such as possible conditions and next steps, while reminding users to consult a medical professional for confirmation.

SVM was used to categorize instances according to pertinent medical characteristics.

- Heart Disease Prediction: Making predictions about heart disease based on characteristics such as age, BMI, blood pressure, and cholesterol levels [1].
- Diabetes Prediction: Making predictions of diabetes based on age, BMI, insulin, and glucose levels.
- Lung Cancer Prediction: Making predictions about lung cancer by using characteristics such as age, smoking history, tumor markers, and symptoms.
- Hypothyroid Prediction: Using TSH, T3, T4 levels, and other metabolic markers, hypothyroidism can be predicted.
- Parkinson's Disease Prediction: Using vocal indicators including frequency, jitter, and shimmer to predict Parkinson's disease

3.6. Implementing Chatbots and Generating Responses

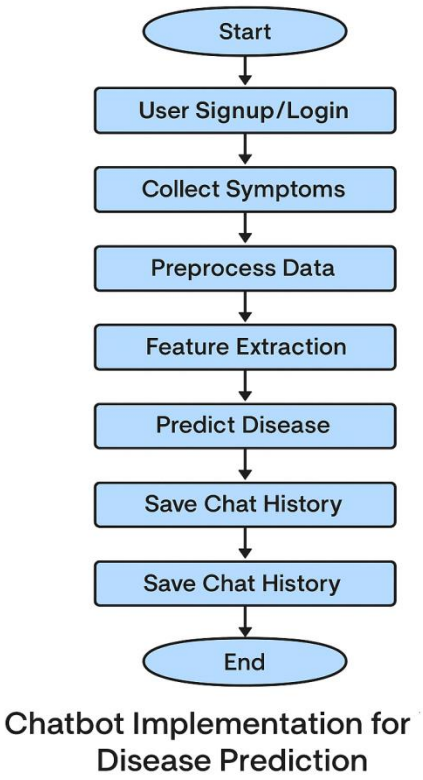

Figure 2: Flowchart for chatbot implementation

After training, the model is included into the chatbot to offer real-time medical advice and disease predictions. The chatbot operates according to a defined procedure:

1. Start
2. User Signup/Login – Register or log in to the system.
3. Collect Symptoms – User inputs symptoms via chatbot interface.
4. Preprocess Data – Apply NLP techniques (tokenization, stop-word removal, lemmatization).
5. Feature Extraction – Convert symptoms to numerical features.
6. Predict Disease – Use machine learning model (e.g., SVM, LSTM) to classify symptoms.

7. Display Results – Show predicted disease and recommendations.
8. Save Chat History – Store interactions for future reference.
9. End

4. ALGORITHM

START

load training and testing CSV files

preprocess data

train decision tree model

train SVM model

load symptom description

// ask user name

print “user name”;

do

//ask user for symptom

print “ enter symptom name”

valid = isValidSymptom(symptom);

if (!valid)

print("Invalid symptom. Re enter.\n");

end if

while symptom is not valid;

// user decision to predict possible disease

if (severity > 13) // more symptoms

print"You should consult a doctor";

else

print"Take precautions and monitor your symptoms” ;

end if

showDescription(disease);

showPrecautions(disease);

END

The above algorithm is designed to predict the possible diseases based on the user input symptoms using machine learning techniques.It start by loading training and testing data from CSV files and preprocessing them through a labelled encoding and train test splitting.The system then trains two models:Decision Tree Classifier and SVM model additional information such as symptoms descriptions, severity levels and recommended precautions are also loaded from the externalCSV files. The user is asked to enter their name, greeted, and then prompted to provide a symptom.The input symptom is validate against a know the list; if its invalid the system continues prompting until the valid symptoms is entered.The user is then asked how many days they have experienced the symptom. Based on the provided input, the systems uses the trained decision tree model to predict the possible diseases. if more information is needed to improve the accuracy, the system asks about the additional related symptoms and refine the productions, finally the system display the predicted diseases and brief description and suggested precautions.

5. TEST CASE

The Healthcare chatbot system consists of a few important parts. First, it collects the user’s name and show it back to create a friendly interaction. then it takes a symptoms like “headache” or “nausea” and predict the possible diseases, along with the descriptions and safety tips. next it ask about more symptoms to refine the diagnosis and checks if the second production matches the first.Lastly, it assess how severe the symptoms are based on their type and how long they’ve lasted, advising whether to see a doctor or take a precautions.All these parts have been tested successfully and work as expected.

The table below shows that there are some test cases:

Table 1: Test Case for Healthcare Chatbot System

Module	Input	Output	Status
User Info Collection	Input user name	Display user name	Tested OK
Diagnosis Prediction	Symptoms entered, e.g., ["headache", "nausea"]	Predicted disease(s), description and precautions displayed	Tested OK
Secondary Prediction	Confirmed symptoms from user	Additional prediction (agreement check with first prediction)	Tested OK
Severity Assessment	Symptoms, Days: (["fever", "cough"], 7)	Advice (doctor consultation recommendation or precaution as per severity formula)	Tested OK

6. CONCLUSION

One of the most significant areas of daily life is medicine, and in this study, we attempted to show how machine learning and other artificial intelligence approaches can be used to better this crucial area. And it shows how a chatbot combined with machine learning and natural language processing can offer a simple and efficient method of making disease predictions based on user symptoms. In order to anticipate potential illnesses including diabetes, heart disease, hypothyroidism, Parkinson's disease, and more, the system gathers the user's symptoms, interprets them using natural language processing (NLP), and applies models like logistic regression, SVM, and LSTM. After that, the chatbot saves the conversation history for later use and provides straightforward, understandable answers and guidance. This method facilitates early disease detection, lessens the workload for physicians, and increases access to healthcare information. The system can assist users in taking prompt action and enhance overall healthcare assistance, even though it cannot take the position of a medical practitioner.

References

[1]. Lavanya, D., Dudyala Bhavya Sree, Katam Vamsi Krishna, Challagolusula Venkatarao, and Kilaru Mohan Brahmendra. April 2025 , "AI-Driven Healthcare Chatbot with Multilingual Support."

[2].Mandepudi, P., Sindhu, G., Rohan, S. S., Reddy, L. C., & M, P. (2025). AI-Driven medical chatbot for early disease detection. International Journal for Research in Applied Science and Engineering Technology, 13(3), 2324–2329. https://doi.org/10.22214/ijraset.2025.67802.

[3].Godikat, Saniya and Pawar, Siddhi and Mhatre, Yash and Malve, Dnyaneshwari, “Multiple Disease Classification and a Personalized AI Chatbot Assistant” (May 24, 2025). Available at SSRN: https://ssrn.com/abstract=5268301 or http://dx.doi.org/10.2139/ssrn.5268301.

[4].Veronica, Oltean Anisia, Ioan Daniel Pop, and Adriana Mihaela Coroiu. "Medical Chatbot for Disease Prediction Using Machine Learning and Symptom Analysis." International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2025).

[5] Sia, M., Ng, K., Haw, S., & Jayaram, J. (2024). Chronic disease prediction chatbot using deep learning and machine learning algorithms. Bulletin of Electrical Engineering and Informatics, 14(1), 742–751. https://doi.org/10.11591/eei.v14i1.8462.

[6]. M. Clark and S. Bailey, "Chatbots in Health Care: ConnectingPatients to Information," Emerging Health Technologies,CADTH Horizon Scans, Canadian Agency for Drugs andTechnologies in Health, Ottawa, ON, Rep. no. EH0122, Jan.2024.

[7]. B. Smith, A. Khojandi, and R. Vasudevan, “Bias in Reinforcement Learning: A Review in Healthcare Applications,”

ACMComputing Surveys, vol. 56, no. 2, pp. 1–17, Feb. 2023, doi: 10.1145/3609502.

[8]. S. Dubey, G. Tiwari, S. Singh, S. Goldberg, and E. Pinsky, “Using machine learning for healthcare treatment planning,” Frontiers in Artificial Intelligence, vol. 6, Apr. 2023, doi: 10.3389/frai.2023.1124182.

[9]. D. M. V. V. Saradhi, K. Tejaswini, A. Kumar, and K. Rajkumar, “Chatbot system for healthcare,” International Research Journalof Modernization in Engineering Technology and Science, Jan. 2023, doi: 10.56726/IRJMETS-NCASCTE202225.

[10]. C. MacKay, W. Klement, P. Vanberkel, N. Lamond, R. Urquhart, and M. Rigby, “A framework for implementing machinelearning in healthcare based on the concepts of preconditions and postconditions,” Healthcare Analytics, vol. 3, pp. 1-9, Nov.2023, doi: 10.1016/j.health.2023.100155.

[11]. Karuna, G., Reddy, G. G., Sushmitha, J., Gayathri, B.,Sharma, S. D., and Khatua, D. (2023). Helpi–an automated healthcare chatbot. In E3S Web of Conferences,volume 430, page 01040. EDP Sciences.

[12]. S. Chakraborty et al., "An AI-Based Medical Chatbot Model for Infectious Disease Prediction," in IEEE Access, vol. 10, pp. 128469-128483, 2022, doi: 10.1109/ACCESS.2022.3227208.

[13]. A. Prasad, B. S. S. S. Jennifer, D. Ghsoh, H. Busshetty, and D. T. J.T, “Chatbot in Healthcare,” International Journal of Engineering Research in Computer Science and Engineering, vol. 9, no. 11, pp. 26–28, Nov. 2022, doi: 10.36647/ijercse/09.11.art008.

[14]. S. Badlani, T. Aditya, M. Dave, and S. Chaudhari, “Multilingual healthcare chatbot using machine learning,” in 2021 2nd International Conference for Emerging Technology, INCET 2021, IEEE, May 2021, pp. 1–6, doi:10.1109/INCET51464.2021.9456304.