Trends in Health Informatics

www.thi.reapress.com

THI. Vol. 2, No. 1 (2025) 10–17.

Paper Type: Original Article

Healthchare Chatbot Using SVM & Decision Tree

Atika Chandel*

Kalinga Institute of Industrial Technology, India; atikachandel12371@gmail.com.

Citation:

Received: 7 June 2024 *Revised: 9 September 2024* *Accepted: 21 November 2024*	Chandel, A. (2025). Healthchare chatbot using SVM & decision tree. *Trends in health informatics, 2*(1), 10-17.

Abstract

The increasing demand for efficient healthcare services and the limitations posed by a shortage of medical professionals have accelerated the development of intelligent healthcare solutions. This research presents a healthcare chatbot designed using Support Vector Machines (SVM) and decision tree algorithms to deliver accurate and timely disease predictions based on user-provided symptoms. The chatbot is aimed at bridging the gap between healthcare providers and patients, offering an accessible platform that mimics the interactions of a human medical professional. By leveraging machine learning techniques, the chatbot can classify symptoms, provide medical guidance, and recommend further medical consultation when necessary. The SVM model categorizes symptoms to identify potential health issues, while the decision tree algorithm predicts diseases and suggests treatment paths based on input data. The system was tested using real-world healthcare datasets to ensure accurate disease prediction and effective user interaction. Experimental results demonstrate that the chatbot achieves high accuracy in predicting diseases, significantly outperforming traditional rule-based systems. The system offers improved response times and enhanced user engagement by providing personalized, context-aware recommendations. Despite its advantages, limitations such as model retraining requirements and data biases were observed, paving the way for future enhancements.In conclusion, this chatbot represents a scalable and user-friendly solution to healthcare challenges, especially in regions with limited medical resources. By providing timely assistance, it has the potential to alleviate pressure on healthcare systems, improve patient outcomes, and foster better access to medical guidance. Future work will further integrate advanced NLP capabilities, expand its disease database, and refine user interaction mechanisms to enhance its utility and accuracy.

Keywords: Healthcare chatbot, Support vector machine, Decision tree, Disease prediction, Natural language processing, Medical assistance, Machine learning, Symptom classification, User interaction.

1 | Introduction

In recent years, chatbots have emerged as influential digital tools capable of interacting with users through natural language interfaces. These systems, powered by Artificial Intelligence (AI), can respond to users' queries and provide relevant real-time information [1]. Chatbots are proving their versatility across multiple

Corresponding Author: 21052406@kiit.ac.ine

domains with applications spanning customer support, virtual assistance, online reservations, and general conversation. The healthcare sector, in particular, benefits immensely from chatbot technology due to its potential to alleviate systemic issues such as resource shortages, limited accessibility, and the growing demand for healthcare services [2]. Healthcare chatbots, designed to assist patients by delivering instant access to health information, symptom analysis, and potential disease predictions, represent a promising frontier in digital healthcare innovation.

Despite their potential, many healthcare chatbots face limitations that restrict their effectiveness. Most existing systems provide pre-scripted, monotonous responses that are often incapable of engaging users in nuanced, empathetic interactions. Consequently, these chatbots struggle to simulate the experience of conversing with a medical professional, reducing the likelihood that users will feel comfortable sharing detailed or sensitive health information. Without the capability for adaptive, intelligent communication, such systems may fail to capture critical information from users, which could lead to accurate health assessments or timely interventions. Recognizing these limitations, this research aims to design and develop a healthcare chatbot that transcends these traditional constraints by leveraging machine learning algorithms specifically, Support Vector Machines (SVM) and decision tree models to enable a more interactive, responsive, and human-like experience.

The proposed healthcare chatbot utilizes Natural Language Processing (NLP) to interpret and analyze user inputs in a way that mimics the conversational skills of a human healthcare provider. By integrating machine learning models, the chatbot can dynamically classify symptoms, predict potential health issues, and provide actionable recommendations tailored to each user's unique situation. The SVM algorithm is particularly suited for high-dimensional data classification and can effectively differentiate between various medical symptoms to support diagnostic predictions. Meanwhile, decision tree models introduce a layer of interpretability, allowing the system to factor in multiple variables from past interactions, symptoms, and user history, thus enabling more personalized and contextually relevant responses.

This study is motivated by the ongoing challenges in global healthcare access, particularly in regions with significant resource limitations, such as rural or underserved areas. In developing countries like India, where healthcare infrastructure is often stretched thin due to high population density and limited professional availability, healthcare chatbots have the potential to fill essential gaps by offering basic medical support and guidance. By acting as a virtual assistant, the proposed chatbot can extend healthcare access to individuals who may otherwise face long waiting times or geographic barriers to traditional healthcare services. As such, this research aims to create a scalable solution that enhances user engagement and satisfaction and provides timely medical assistance, thereby alleviating some of the strain on healthcare systems in resource-constrained settings.

2|Literature Review

The integration of AI and NLP has revolutionized healthcare, particularly through the development of chatbots. These digital assistants are increasingly deployed to provide medical information, triage patients, and offer personalized health advice.

Early chatbot systems, such as ELIZA and A.L.I.C.E., relied on simple pattern recognition and scripted responses, demonstrating the potential for human-computer interaction in healthcare. However, their limitations in understanding complex queries and providing contextually relevant information hindered their widespread adoption.

Recent advancements in AI and NLP have enabled more sophisticated chatbot systems to be developed. Flora Amato et al. explored using AI-driven chatbots to engage patients in personalized health conversations, utilizing platforms like IBM Watson. This research highlighted the potential of chatbots to simulate human-like interactions and deliver tailored health advice.

Divya et al. [3] proposed a linear approach to symptom extraction and disease diagnosis, emphasizing the importance of accurate symptom mapping for effective chatbot-based healthcare. This approach can help identify potential health conditions and guide users to appropriate medical attention.

Kowatsch et al. [4] focused on developing text-based healthcare chatbots for mobile applications, emphasizing patient interaction and daily health monitoring. These systems can enhance patient adherence to treatment plans by providing consistent communication and reminders. However, their reliance on predefined scripts and pattern-matching techniques limits their scalability and adaptability.

Our research proposes a chatbot that integrates SVM and decision tree algorithms for disease prediction to address these limitations. This approach aims to improve the accuracy and adaptability of chatbot interactions by leveraging predictive modeling and real-time data analysis. The chatbot can predict potential diseases by analyzing user-provided symptoms and providing relevant medical advice.

In conclusion, integrating AI and NLP techniques has significantly advanced the capabilities of healthcare chatbots. However, more sophisticated systems that can adapt to diverse user inputs, provide accurate diagnoses, and offer personalized health recommendations are still needed. Our proposed chatbot addresses these challenges by combining advanced machine learning algorithms with NLP techniques.

Table 1. Comparative analysis.

Survey Papers	Techniques	Functionality	Mathematical Model	Result	Conclusion	Advantages	Disadvantages
A self diagnosis medical chatbot using artificial Intelligence	It gives linear design that containes aartificial intelligence, pattern matching, disease, query processing techniques. It gives chatbot dialogues finite stae graph and fuctional architecture.	Identifies the corresponding symptoms and minor or major disease and gives advice according to that	Not given	Provides personalized diagnoses based on symptoms	It is concluded that the system in user friendly and it can be used by who knows how to type in their language.	It is user friendly and the implementation of chatbot can provide better treatment in a short amount of time.	Chatbot identifies only minor and major disease and gives premission whether to go to a doctor or not.
Disease prognosis with symptoms endorsement using combinational classifier	It gives a model for disease prediction from symptoms clustering and combinational classufucation by using various algorithms like K-Nearst Neighbor, Decision Tree, Nzive Bayes.	It uses association rules for symptoms and predicts the disease well.	Given with decision tree algorithm and Bayes theorem	Better than the rule based systems and accuary of predictions can be ameli iorated due to combinational classifiers.	It is concluded that on successful impelementation, the system will be able to predict the disease effectively to achieve better result.	It is far better than a rule based system by increasing accuary of identifying diseases by combinational classifiers.	If the number of symptoms will be less given then accuary of idetifying diseases may degrade.

3 | Variables and Equations

Various variables and mathematical equations are essential for effective data processing, decision-making, and user interaction in developing an AI-based health chatbot. Here is an overview of key components that define the underlying operations and behavior of such a system:

3.1 | Key Variables

User input variables (U)

Represent user-provided data such as symptoms, queries, or contextual information. Example: $U = \{s_1, s_2, ..., s_n\}$ }, where s_i denotes each symptom or input item.

Training data variables (D)

Represents datasets used to train the chatbot's machine learning models, including symptom descriptions, severity levels, and treatment recommendations, $D = \{(x_1, y1), (x_2, y_2), ..., (x_m, y_m)$.

Here, xi is an input instance (e.g., symptoms), and y_i is the expected output (e.g., diagnosis or recommendation).

Response variables (R)

Represent the chatbot's outputs, such as disease predictions or user advice. Example: $R = f(U)$, where f is a function that processes user input and generates a response.

Machine learning model parameters (θ)

Include weights, biases, and other parameters adjusted during the training phase to optimize predictions. These parameters are adjusted using algorithms such as gradient descent to minimize error between predicted and actual outputs. Equations used in AI health chatbot.

Symptom classification (decision function)

$$F(x) = \text{argmax}_y, P(y \mid x,\theta).$$

I. F(x) predicts the most probable disease y given input features x (user symptoms) and parameters θ.

II. The decision function is typically used with classifiers like SVM or decision trees.

Similarity matching (NLP for user queries)

$$\text{Similarity}\left(Q_i, Q_j\right) = \frac{\text{Count}\left(Q_i \cap Q_j\right)}{\sqrt{\text{Count}\ (Q_i} \times \sqrt{\text{Count}\ (Q_j\ )}}.$$

Measures similarity between user queries Q_i and known responses using cosine similarity.

Confidence score calculation

$$C = \frac{\sum_{i=1}^{n} w_i . s_i}{\sum_{i=1}^{n} w_i}.$$

C denotes the confidence score for a given response based on weights assigned to symptoms or user inputs.

Decision tree for diagnosis

A decision tree uses recursive splitting based on conditions to reach a diagnosis: f $s_i < t$, then branch A else branch B. Here, s_i is the symptom score, and t is a threshold value determined by training data.

This mathematical framework helps the AI-based health chatbot offer accurate responses, make informed decisions, and enhance user interactions through continuous learning and adaptation.

4 | Proposed Framework

The proposed framework for the healthcare chatbot leverages a combination of NLP and machine learning algorithms to deliver accurate disease prediction and user-specific medical advice. The system is designed to provide a user-friendly, interactive experience, allowing patients to input symptoms via a text-based interface. The chatbot then processes this input to diagnose potential health issues and offer tailored responses, including treatment recommendations or guidance to seek further medical consultation.

The framework comprises four main components: a user input module, an NLP module, a machine learning module, and a response generation module. The user input module captures and preprocesses user queries, removing noise and extracting relevant keywords. The NLP module analyzes these keywords, applies tokenization, and identifies critical phrases to understand the user's intent and symptoms better. This processed data is then passed to the machine learning module, which employs SVM to classify symptoms and decision tree algorithms to predict diseases.

The decision tree component of the system determines possible diagnoses by evaluating the hierarchy of symptoms, taking into account their severity and any related contextual factors. This hierarchical decision-making process helps narrow potential diseases, ensuring the system offers relevant and accurate responses. The SVM model complements this by categorizing symptoms into distinct classes, enabling more precise diagnosis and reducing prediction ambiguity.

Once the diagnosis is made, the response generation module crafts a customized response, including symptom-specific advice, recommended precautions, or information about nearby healthcare professionals if further treatment is necessary. This modular design ensures the chatbot can engage in dynamic, context-aware conversations and adapt to varied user inputs.

The proposed framework facilitates intelligent disease prediction and emphasizes accessibility, scalability, and user engagement, offering a powerful tool to enhance healthcare delivery and accessibility.

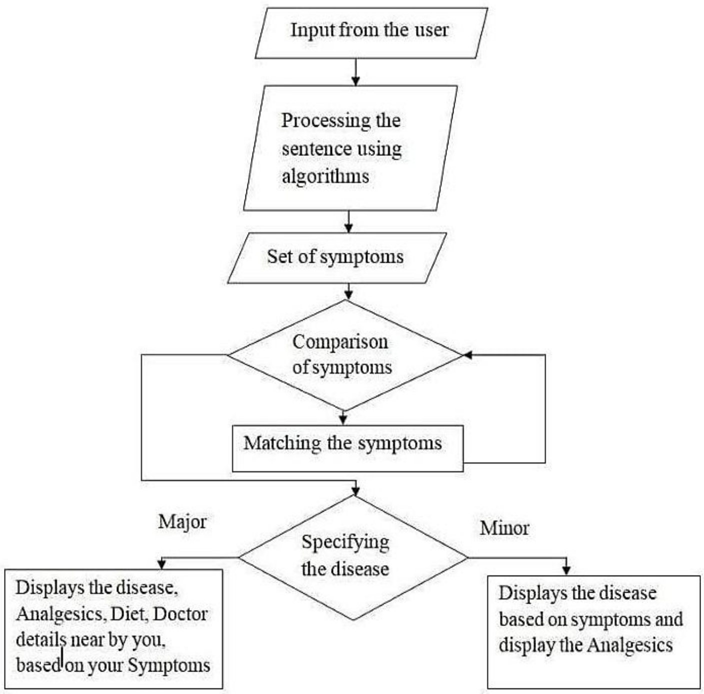

Fig. 1. Data-flow diagram.

5 | System Architecture

The above figure proceeds with the user can start their conversation with the chatbot that is user-friendly, and it will be stored in the database for future reference. The chatbot will clarify the user's symptoms with a series of questions, and the symptom confirmation will be done. The disease will be categorized as minor and major disease. The chatbot will reply whether it's a major or minor disease. If it's significant, the user will be suggested with the doctor's details for further treatment and display the analgesics. It also provides food suggestions, which means which food you have to take more to recover from the disease. The chatbot user interface can be user-friendly by using chatbot not to go to hospitals for even small problems.

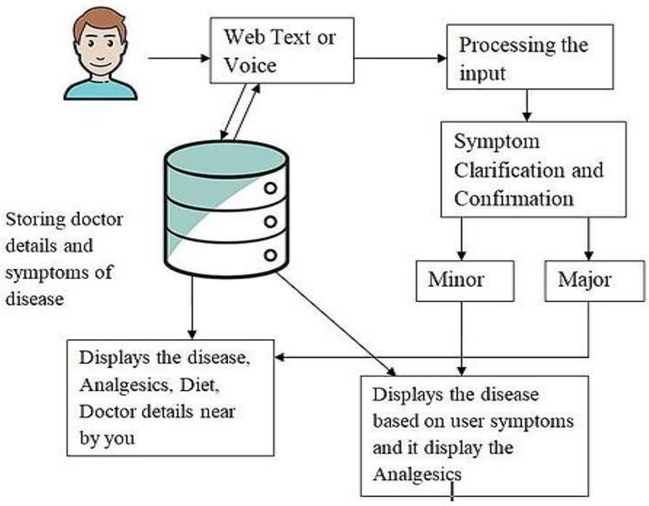

Fig. 2. Operational overview.

6 | Experimental Setup

Before we began our set-up, we needed the relevant datasets, which we browsed and referenced from Kaggle [5]. The datasets are training, testing, symptom description, symptom severity, and sympotom precaution.

To start the set-up, we imported the necessary libraries, namely pandas, sklearn, numpy, csv, and warnings. Preprocessing, decision tree classifier, tree, train test split, and cross val score were imported from sklearn's different sub-libraries.

Afterward, we read the training and testing datasets (X held the symptoms, and Y held the prognosis) with pandas and visualized them with functions such as head and info. When data reduction was made using a group, we found 41 unique diseases and 133 symptoms. From preprocessing, we used 'LabelEncoder' to transform our prognosis from Y so that every prognosis had a unique numeric value we could use. We then split the training dataset into a 1:2 ratio of test and train.

Using the Decision Tree Classifier, we ran the train test split from the experimental set-up to get a cross-validation score of 0.979. Using an SVC model, we got a model score of 1.

Using feature importances from decision tree classifier, we assigned different importance to the symptoms and visualized them using print.

We then set up four distinct dictionaries, severity dictionary, description_list, precaution dictionary, and symptoms-dict, to map severity, descriptions, precautions, and symptoms from the datasets. We then calculated the severity of the condition by using the sum of days, severity, and length of symptoms.

Finally, we take user input on the symptoms they're facing, encode it, and pass it through a function that provides an interactive diagnostic experience. It collects user input on symptoms and symptom duration, traverses the tree to predict disease, cross-checks the prediction with additional symptoms, and provides explanations and precautions.

7 | Experimental Results and Discussion

The final result of the project looks something akin to.

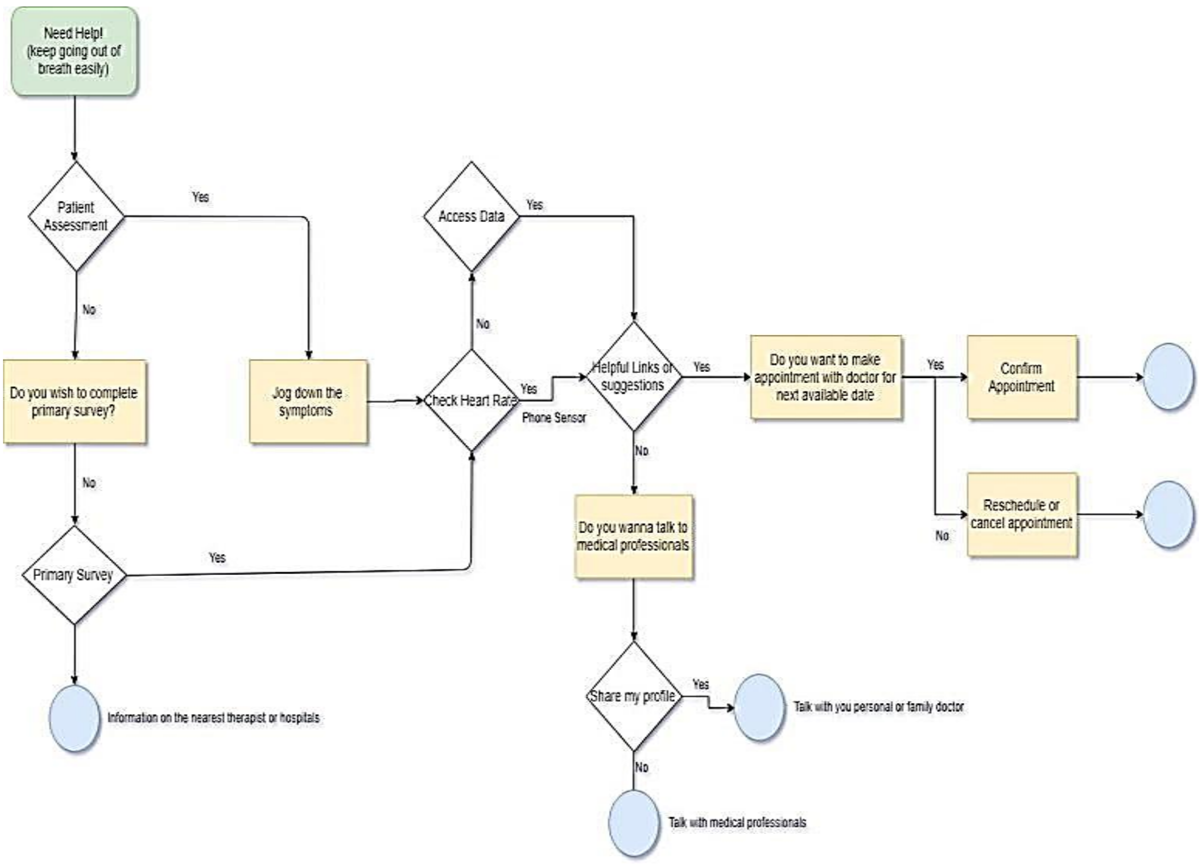

Fig. 3. Example output.

To evaluate the proposed healthcare chatbot's performance, we utilized a publicly available medical dataset containing a diverse range of symptoms, diagnoses, and treatment recommendations. The dataset was preprocessed to ensure data quality and consistency.

7.1 | Model Implementation

The chatbot was implemented using Python, leveraging popular libraries such as NLTK, Scikit-learn, and Tensor Flow. The SVM and Decision tree models were trained on the preprocessed dataset, with hyperparameters tuned using grid search and cross-validation

techniques.

Using the decision tree classifier, we ran the train test split from the experimental set-up to get a cross-validation score of 0.979, and upon using an SVC model, we got a model score of 1. By further manual testing, the chatbot ensured that the predictions were in line with the dataset

8 | Conclusion

Our medical chatbot is designed to provide essential medical assistance to patients suffering from common ailments such as fever, cold, typhoid, malaria, and jaundice. Given our country's increasing population and limited number of healthcare professionals, this chatbot aims to address the growing healthcare needs. While similar systems exist in foreign countries, they are not readily accessible in our nation. By deploying this chatbot, we aim to bridge this gap and improve the efficiency and performance of the medical industry, ultimately reducing mortality rates [6].

Our chatbot incorporates a wide range of crucial features for providing effective medical assistance. Unlike previous models, which were often expensive and inaccessible to the general public, our system is designed to be affordable and user-friendly. The integration of AI-powered chatbots into healthcare has the potential to revolutionize patient care. These chatbots can significantly reduce healthcare costs and enhance overall efficiency by automating routine tasks, improving communication, and facilitating faster decision-making. The user-friendly interface of our chatbot allows individuals to easily interact with the system and receive timely medical advice, regardless of their technical expertise [7].

For future scope, there can be wider data sets used; introducing neural networks along with decision trees can help tackle more complicated cases; as decision trees are limited to the data provided, deep neural networks will help with extrapolating and judging the service even if it the prognosis isn't present in our current dataset. In addition, LLMs could be introduced for a more de-centralized execution if the model is hosted, and it'll be self-improving as more and more users log on to the platform.

Acknowledgments

We sincerely thank Dr. Shubadip Pramanik for his steadfast support and invaluable guidance throughout this project. His expertise and dedication were pivotal in shaping every aspect, from the initial undertaking to curating the report and research paper. This endeavor would not have been possible without his critical insights and unwavering encouragement.

Author Contribution33

Data collection, cleaning, and analysis, Dewansh Saboo, Saswati Padhy. Proposed Framework, Atika Chandel. Experimental Set-up, Nekhil K. Agarwal.

References

[1] Wu, L. (2024). Research on the application of mechanical automation feedback control in vehicle stability control under fatigue driving state. *2024 3rd international conference on artificial intelligence and autonomous robot systems (AIARS)* (pp. 356–361). IEEE. https://doi.org/10.1109/AIARS63200.2024.00073

[2] Kumar, A., Bag, A., Anand, A., Saha, S., Mohapatra, H., & Kolhar, M. (2025). Examining healthcare services utilizing cloud technology in intelligent urban environments. In *Revolutionizing healthcare systems through cloud computing and iot* (pp. 77–98). IGI Global. https://doi.org/10.4018/979-8-3693-7225-8.ch004

[3] Divya, S., Indumathi, V., Ishwarya, S., Priyasankari, M., & Devi, S. K. (2018). A self-diagnosis medical chatbot using artificial intelligence. *Journal of web development and web designing*, *3*(1), 1–7. https://core.ac.uk/download/pdf/230494941.pdf

[4] Kowatsch, T., Nißen, M., Shih, C. H. I., Rüegger, D., Volland, D., Filler, A., ... & Farpour-Lambert, N. (2017). Text-based healthcare chatbots supporting patient and health professional teams: preliminary results of a randomized controlled trial on childhood obesity. *Persuasive embodied agents for behavior change (PEACH2017)*. ETH Zurich, Department of Management, Technology and Economics. https://www.researchgate.net/publication/320161507_Text-based_Healthcare_Chatbots_Supporting_Patient_and_Health_Professional_Teams_Preliminary_Results_of_a_Randomized_Controlled_Trial_on_Childhood_Obesity

[5] Kaggle. (2020). *What are some of the best ways to find datasets?* https://www.kaggle.com/questions-and-answers/280618

[6] Mohapatra, H., Mishra, S. R., Rath, A. K., & Kolhar, M. (2025). Sustainable cities and communities: role of network sensing system in action. *Networked sensing systems*, 173–198. https://doi.org/10.1002/9781394310890.ch7

[7] Priyadarshi, S., Subudhi, S., Kumar, S., Bhardwaj, D., & Mohapatra, H. (2025). Analysis on enhancing urban mobility with IoT-integrated parking solutions. In *Interdisciplinary approaches to transportation and urban planning* (pp. 143–172). IGI Global. https://doi.org/10.4018/979-8-3693-6695-0.ch006