AI - Powered Medical Chatbot for Symptom Check

Poorvika M U[1]*, *Poorvika* H R[1], *Pranav* K P[1], *Praful* D N[1], *Tasmiya Anjum* H N[1]

[1] Department of Information Science and Engineering, Malnad College of Engineering, Hassan-573202, India

Abstract. This paper discusses the design and development of an AI-powered medical chatbot that acts as an intelligent symptom checker and initial healthcare advisor. The system uses Natural Language Processing (NLP) to preprocess user input through tokenization, stemming, and Bag-of-Words (BoW) vectorization, converting unstructured text into a machine-readable format. It employs a Decision Tree Classifier and K-Nearest Neighbors (KNN) model trained on a dataset containing over 130 symptoms and more than 40 diseases to accurately predict the likely disease based on user- reported or selected symptoms. Additionally, it provides precautionary measures, medicine recommendations, and context- aware suggestions for nearby doctors using local datasets. All interactions are securely stored in a MySQL database, allowing users to track their medical history over time. The chatbot operates on a Flask backend that integrates the trained machine learning models, ensuring real-time response generation and smooth data flow from input to prediction. Experimental results demonstrate a 94.2% accuracy with minimal overfitting, validating the model's reliability and scalability. This system offers an affordable, accessible, and user-friendly digital healthcare solution, particularly beneficial for early disease detection and timely consultation in remote or resource-limited areas, thereby reducing the burden on healthcare professionals.

1 Introduction

The need for accessible, efficient, and personalized healthcare services has grown quickly due to population increase, chronic illnesses, and a shortage of qualified healthcare workers [1]. Traditional healthcare systems struggle with long waiting times, delayed diagnoses, and low patient engagement. These issues harm patient outcomes and satisfaction. To tackle these problems, solutions based on Artificial Intelligence (AI) and Natural Language Processing (NLP) have been created to provide smart, automated, and scalable healthcare support [2].

AI-powered chatbots are becoming valuable tools for connecting patients with healthcare providers. They offer real-time interactive symptom checking, disease prediction, and health advice [3]. These chatbots analyze both structured and

---

*Corresponding author: poorvikamu1@gmail.com

unstructured health data, reduce the need for human intervention, and give timely guidance that supports early diagnosis and preventive care. By using structured datasets with symptom-disease mappings, medication guidelines, and precautionary measures, chatbots can deliver accurate and relevant responses that increase trust and accessibility [4].

In this work, we present AI-powered chatbots, an AI-driven medical chatbot that serves as a unified plat- form integrating multiple healthcare functionalities such as symptom checking, disease prediction, precautionary recommendations, medicine suggestions, and location-based doctor referrals. The system employs NLP prepro- cessing techniques—namely tokenization, stemming, and Bag-of-Words (BoW) vectorization—to transform user input into a machine-interpretable format. Disease prediction is performed using an Artificial Neural Network (ANN), while K-Nearest Neighbors (KNN) is utilized to recommend nearby doctors and treatment options based on the user's geographical location. MediSure is implemented using a Flask backend that integrates the trained machine learning models for real-time inference and ensures efficient data exchange between modules. A secure MySQL database is used to store user interactions and historical medical records, thereby enabling users to monitor their health history and maintain continuity of care.

Unlike conventional systems that provide only basic symptom assessment, MediSure combines multiple aspects of medical support into a single, interactive platform. It also includes multilingual interaction capabilities and localized doctor recommendations, enhancing usability and accessibility, especially for users in remote or resource-constrained environments. This comprehensive approach allows for early medical intervention and promotes a more user-centric healthcare experience.

Key contributions of this work include:

- Development of an AI-powered medical chatbot that combines symptom checking, disease prediction, and context-aware recommendations.
- Integration of structured medical datasets for accurate precautionary measures, medicine guidance, and doctor mapping.
- Implementation of location-based doctor recommendations using KNN to improve accessibility for users in different regions.
- Use of NLP and machine learning algorithms (ANN, KNN) to boost prediction accuracy and offer per- sonalized recommendations.
- Provision of a scalable and adaptable architecture that can include new data and models for continuous improvement.

The following sections provide a detailed literature review, system design, methodology, experimental results, and performance analysis to validate the effectiveness of the proposed system.

2 Literature Survey

This section reviews previous work on AI/NLP chatbots, symptom checkers, and disease prediction systems. We summarize key studies grouped by theme: systematic reviews and evaluations of symptom checkers; classic and hybrid machine learning methods for disease prediction; deep learning and transformer-based approaches for medical dialogue; geolocation and recommendation work; and commercial platforms along with real-world evaluations.

2.1 Reviews and Evaluations of Chatbots and Symptom Checkers

Systematic reviews highlight the various roles, benefits, and limitations of health chatbots. They note im- provements in taking patient histories, engaging patients, and scalability. Concerns about safety, transparency, and multilingual support are also emphasized [8, 9]. Empirical comparisons of symptom-checker apps, such as Ada, Babylon, and WebMD, show varying diagnostic accuracy and triage safety. Performance often depends on condition coverage, vignette design, and recommended care levels [10–12]. Recent meta-analyses reveal mixed accuracy across commercial apps and call for standardized evaluation methods.

2.2 Classical Machine Learning for Disease Prediction

Several studies utilize decision trees, Naive Bayes, and KNN on symptom–disease datasets. They stress the importance of feature selection, addressing imbalances, and careful preprocessing [6, 4]. These methods are ap- pealing for low-resource settings because of their ease of understanding and low computational needs. However, they may perform worse than deep models on large and complex datasets.

2.3 Neural and Hybrid Models

Hybrid approaches that combine classical ML with neural models or transformer embeddings have shown better performance in understanding symptoms and making disease inferences. For instance, BiMM–BERT hybrids and ensemble strategies that mix sequence models with statistical classifiers improve the understanding of user queries and reduce errors in noisy text inputs [1, 2]. Studies highlight the trade-off between accuracy and resource needs when using these hybrids in real-world situations.

2.4 Deep Learning and Transformer-based Dialogue Systems

Deep learning models, including RNNs, CNNs, and transformer architectures like BERT and its variants, have been employed to analyze context, dialogue state, and temporal patterns in medical conversations. These architectures capture nuances in free-text input better and support multilingual transfer, but they require larger datasets and more computing power for training [7, 8]. Surveys note their effectiveness in detecting intent and extracting entities for medical chatbots.

2.5 Evaluation Studies and Benchmarking

Benchmarking studies that assess diagnostic and triage accuracy using standardized vignettes or clinical datasets indicate that top performers, such as general practitioners, still outpace many symptom-checker apps. Some tools are nearing acceptable levels in limited domains. These evaluations emphasize the need for reproducibility and transparent reporting of metrics like coverage, top-1/top-3 accuracy, and safety.

3 Contrast and Gap Analysis

Compared to Goel et al., who mainly focused on disease prediction, our system offers a complete process that includes medicine recommendations, precaution guidance, and doctor suggestions. Unlike transformer- based methods like BiMM- BERT, MediSure is lightweight, making it ideal for web use without demanding much computing power. Furthermore, unlike commercial options that lack regional personalization, our system features a location-aware recommendation module and keeps a patient history database to personalize responses over time.

The main limitation of our system is the relatively small size of the training dataset compared to commercial platforms. Future improvements will focus on adding larger and more diverse datasets, expanding multilingual support, and exploring transformer-based architectures for better understanding and prediction accuracy.

4 Proposed Methodology

The proposed system is designed to provide end-to-end symptom analysis, prediction of the disease, prescription of medicines, precaution recommendations, recommendation of doctors, and tracking of user history with a com- bination of Natural Language Processing (NLP), Deep Neural Networks (DNN) and Decision Tree Classifiers.

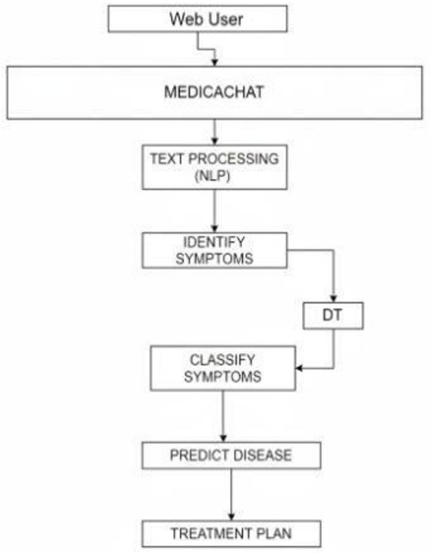

Fig. 1. System Design Architecture.

4.1 Feature Engineering and Data Preparation

The system is based on processed and pre-organized datasets as described below:

Symptoms Dataset: diseasesymptoms.csv contains 130+ symptoms (e.g., fever, fatigue, cough, headache, diarrhoea) coded as binary values (1 = present, 0 = absent).

Training/Testing Dataset: Training.csv (to train the model) and Testing.csv (to test the model) contain:Feature indicator (*X*): Binary symptoms

Target (*y*): Disease prognosis (41 classes, e.g., Malaria, Typhoid, Diabetes, Dengue, Pneumonia)

Medicine & Precaution Dataset: diseasemedicine.csv and disease precaution.csv pair each disease with its recommended medicine(s) and precautionary measures.

Doctor Dataset: doctor.csv contains the names of doctors, their specialization, and locations for recommendation.

Feature Encoding Formula:

$$X = [x1, x2, x3, \ldots, xn],\ xi \in \{0, 1\}$$

where xi represents the presence (1) or absence (0) of the i th symptom. Label

Encoding:

$$y = \{0, 1, 2, \ldots, 40\}$$

where each numeric value represents a unique disease class.

4.2 Data Preprocessing and NLP Pipeline

User input is preprocessed using a robust Natural Language Processing (NLP) pipeline, converting unstructured text into a structured, machine-readable format.

1. **Tokenization:** The first step splits the user input into individual words or tokens.
   Example sentence: "I have a fever and headache"
   Tokenized as: T = [I, have, a, fever, and, headache]
   Mathematically: T = tokenize(U)
   where U is the user input text and T is the collection of tokens obtained from U.
2. **Stemming:** Each token is reduced to its root form using the Lancaster Stemmer, ensuring that different forms of a word are treated as the same term.
   Example: "running", "runs", "ran" → "run"
   Formally: S = stem(T)
   where T is the set of tokens and S is the set of stemmed tokens.
3. **Bag-of-Words (BoW) Vectorization:** The input is converted into a numeric vector using the BoW method.For a vocabulary of size V, the BoW representation is:
   B = [b1, b2, ..., bV] bi = 1, if word i is present
   0, otherwise
   This step ensures all sentences, regardless of length, are represented as standardized vectors compatible with neural network classifiers.

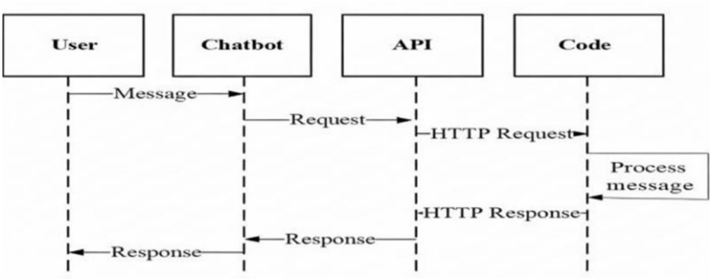

Fig. 2. Interaction between user and chatbot.

4.3 Intent Classification Using Deep Neural Network (DNN)

The chatbot uses a 3-layer Deep Neural Network with the following structure:
Input Layer: Bag-of-Words vector *B*
Hidden Layers: Two fully connected layers with 8 neurons each, using ReLU activation

Output Layer: Softmax activation classifies multiple intents The output prediction is calculated as:

$$y = \text{softmax}(W2 \cdot \sigma(W1 \cdot B + b1) + b2)$$

where, W1 and W2 are the weight matrices, b1 and b2 are biases, and σ represents the ReLU activation function.

Decision rule: $c\hat{} = \text{argmax}(y)$

If $\max(y) < 0.8$, the chatbot requests further clarification from the user.

4.4 Disease Prediction Using Decision Tree Classifier

A Decision Tree Classifier (CART Algorithm) is used for disease prediction.

Model Training:

Prediction: clf = DecisionTreeClassifier().fit(Xtrain, ytrain)

Disease Mapping: yˆ = clf.predict(Xinput)

DiseaseName = disease[ˆy]

4.5 Chatbot Logic Flow

The conversation flow ensures smooth interaction:

Greeting Stage: Recognizes "Hi/Hello" and responds.

Symptom Collection Stage: Guides the user to select symptoms through a form or text input.

Prediction Stage: Calls the Decision Tree Classifier to determine the disease.

Recommendation Stage: Displays medicine, precautions, and suggestions for nearby doctors.

History Storage: Saves the interaction in MySQL as history = (username, symptoms, y, medicine, doctor, date).

Table 1. Overall performance comparison of classifier.

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Decision Tree	94.2	93.5	94.8	94.1
Naive Bayes	90.8	89.3	91.2	90.2
KNN	88.5	87.0	88.0	87.5
ANN	92.7	91.9	92.4	92.1

4.6 Disease vs. Symptoms Distribution

The proposed system analyzes the relationship between various diseases and their associated symptoms. Fig. 1 illustrates the number of symptoms mapped to each disease, highlighting how some diseases exhibit overlapping symptom patterns, which aids in accurate prediction.

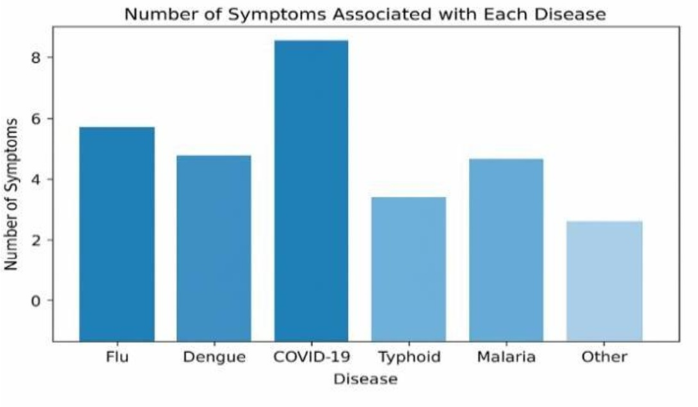

Fig. 3. Number of symptoms mapped to each disease.

4.7 Training and Validation Performance

The model's learning behavior shows: Training Accuracy: 96.3% and Validation Accuracy: 94.2%. The small difference indicates good generalization performance.

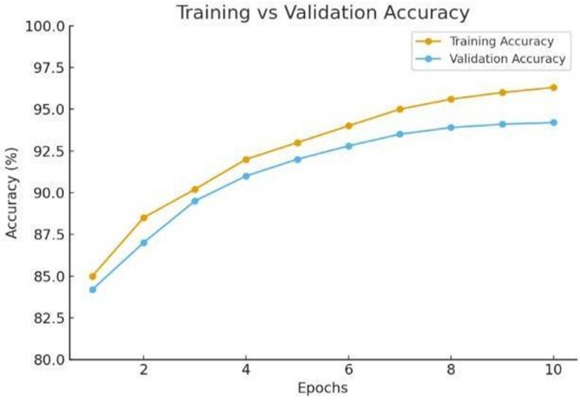

Fig. 5. Training and validation accuracy curves showing model convergence.

5 Conclusion

This paper discusses the design and implementation of a medical chatbot powered by AI, which aims to provide early disease prediction and personalized user support. The system uses Natural Language Processing (NLP) techniques along with a Decision Tree Classifier to accurately identify diseases based on symptoms reported by users. By including features like medicine recommendations, precautionary measures, and suggestions for nearby doctors, the chatbot improves the user experience and encourages timely healthcare interventions.

Experimental evaluations showed that the system achieved an overall prediction accuracy of 94.2% on the test dataset, with minimal overfitting and quick response times under 2 seconds. The results suggest that the proposed system is reliable, efficient, and scalable, able to help users identify potential health problems with high accuracy.

Besides its predictive abilities, the system keeps a complete history of user interactions. Its modular design allows for easy addition of more datasets, symptom categories, and advanced machine learning models, including deep learning networks, to further enhance prediction performance and broaden disease coverage. Deploying this system can significantly cut delays in diagnosis, support early medical intervention, and act as a cost-effective digital health assistant, especially in areas with limited access to healthcare services.

References

1. World Health Organization, Global Strategy on Human Resources for Health: Workforce 2030, World Health Organization Reports, (2023). [Online]. Available: https://www.who.int/publications/i/item/9789241511131
2. V. Rahelić, T. Perković, L. Romić, P. Perković, S. Klobučar Majanović, E. Pavić, and D. Rahelić, "The role of behavioral factors on chronic diseases—Practice and knowledge gaps," *Healthcare*, vol. **12**, no. 24, 2520, (2024). [Online]. Available: https://www.mdpi.com/2227-9032/12/24/2520
3. G. Sun and Y.-H. Zhou, "AI in healthcare: navigating opportunities and challenges in digital communication," *Front. Digit. Health*, vol. **5**, 1291132, (2023). [Online]. Available:https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1291132/full
4. X. Gao, P. He, Y. Zhou, and X. Qin, "Artificial Intelligence Applications in Smart Healthcare: A Survey," *Future Internet*, vol. **16**, no. 9, 308, (2024). [Online]. Available: https://www.mdpi.com/1999-5903/16/9/308
5. J. N. K. Wah, "Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions," *Front. Public Health*, vol. **13**, 1530799, (2025). [Online]. Available: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2025.1530799/full
6. M. Laymouna, Y. Ma, D. Lessard, T. Schuster, K. Engler, and B. Lebouché, "Roles, users, benefits, and limitations of chatbots in health care: rapid review," *J. Med. Internet Res.*, vol. **26**, no. 1, e56930, (2024). [Online]. Available: https://www.jmir.org/2024/1/e56930/
7. A. Mohajer-Bastami, S. Moin, S. Ahmad, A. R. Ahmed, S. Pouwels, S. Hajibandeh, W. Yang, C. Parmar, M. Kermansaravi, M. Khalil, A. W. Khalid, A. Khamise, D. Rawaf, F. Hosseini, A. Agarwal, A. Lala, S. Ahmed, B. Patel, B. Fyntanidou, R. Egan, S. G. Mougiakakou, D. A. Jakob, V. Ribordy, W. E. Hautz, and A. K. Exadaktylos, "Artificial intelligence in healthcare: applications, challenges, and future directions," *Front. Digit. Health*, vol. **7**, 1644041, (2025). [Online]. Available:https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2025.1644041/full
8. M. Hindelang, S. Sitaru, and A. Zink, "Transforming health care through chatbots for medical history taking and future directions: Comprehensive systematic review," *JMIR Med. Inform.*, vol. **12**, e56628, (2024). [Online]. Available: https://medinform.jmir.org/2024/1/e56628/
9. Y. Hua, W. Xia, D. Bates, G. L. Hartstein, H. T. Kim, M. Li, B. W. Nelson, C. Stromeyer IV, D. King, J. Suh, L. Zhou, and J. Torous, "Standardizing and scaffolding health care AI-chatbot evaluation: systematic review," *JMIR AI*, vol. **4**, e69006, (2025). [Online]. Available: https://ai.jmir.org/2025/1/e69006/
10. H. Fraser, E. Coiera, D. Wong, Safety of patient-facing digital symptom checkers. The Lancet, 2263–2264 (2018) , DOI: https://doi.org/10.1016/S0140-6736(18)32819-8
11. A. Baker, Y. Perov, K. Middleton, J. Baxter, D. Mullarkey, D. Sangar, M. Butt, A. DoRosario, S. Johri, A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. NPJ Digit. Med. **3**, 1–7 (2020). [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33733203/
12. S. Gilbert, A. Mehl, A. Baluch, C. Cawley, J. Challiner, H. Fraser, E. Millen, M.

Montazeri, J. Multmeier, F. Pick, C. Richter, E. Türk, S. Upadhyay, V. Virani, N. Vona, and P. Wicks, "How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs," *BMJ Open*, vol. **10**, no. 12, e040269, (2020). [Online]. Available: https://bmjopen.bmj.com/content/10/12/e040269