



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

⊕ www.ijirset.com 🖂 ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405050|

Multiple-Disease Prediction System and Medical Chatbot

Ramesh B E

Associate Professor, Department of Computer Science Department, S.J.M.I.T, Chitradurga, Karnataka, India

Latha SR, Manegara Priyanka, Prajna KL, Varsha Umesh

BE, 7th semester, Department of Computer Science Department, SJMIT, Chitradurga, Karnataka, India

ABSTRACT: The Multi-Disease Prediction System and Medical Chatbot is a comprehensive healthcare solution leveraging Python and machine learning to provide accurate and early disease predictions alongside interactive patient support. The system is designed to predict multiple diseases, including diabetes, heart disease, and more, based on user-provided symptoms and medical history. Advanced machine learning algorithms analyze user input and identify patterns to deliver precise predictions, assisting users in recognizing potential health risks and encouraging timely medical consultations. The application features a user-friendly interface, ensuring accessibility for individuals without technical expertise. This system can be accessed via web or mobile platforms, making it suitable for a wide range of users. Additionally, the prediction results are presented in an easy-to-understand format, accompanied by recommendations for further action.

KEYWORDS: Medical Chatbot, Machine Learning, Symptom Checker, AI in Healthcare.

I. INTRODUCTION

Multi-Disease Prediction System and Medical Chatbot is an innovative healthcare application that combines machine learning algorithms with natural language processing to provide preliminary medical assessments and health-related information to users. The system is designed to analyze patient symptoms and medical history using various ML models, including Random Forest, Stream Lit, Support Vector Machines, and Neural Networks, to predict the likelihood of multiple diseases such as diabetes, heart disease, and various types of cancer. The medical chatbot component utilizes natural language processing and pattern matching techniques to engage in human-like conversations about health concerns and medical queries, trained on a comprehensive medical knowledge base that includes information about symptoms, diseases, medications, and general health guidelines. The system integrates both components seamlessly, allowing users to switch between disease prediction and chatbot interactions as needed, while employing robust data preprocessing techniques and feature selection methods to ensure accurate disease predictions.

The platform maintains user privacy through secure data handling and encryption protocols while storing and chat logs, with regular model updates and continuous learning from user interactions helping to improve the system's accuracy and effectiveness over time. The implementation includes sophisticated data validation techniques to ensure the quality of user inputs and prevent erroneous predictions, along with a comprehensive error handling system that gracefully manages unexpected scenarios. The system also features an intuitive user interface that guides users through the process of entering symptoms and medical history, making it accessible to people with varying levels of technical expertise.

II. LITERATURE SURVEY

The advancements in artificial intelligence (AI) and machine learning (ML) have significantly impacted the healthcare industry, enabling the development of systems for disease prediction and patient care. Numerous studies have demonstrated the potential of ML algorithms, such as Decision Trees, Random Forests, and Neural Networks, in predicting diseases based on clinical datasets. Research has shown that these models, when trained on comprehensive datasets, can achieve high accuracy in diagnosing conditions such as diabetes, heart disease, and cancer. Furthermore, natural language processing (NLP) has played a vital role in the development of chatbots, allowing systems to interact conversationally with users and provide health-related advice effectively. Previous projects like IBM Watson Health





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405050|

and Microsoft's Health Bot Service have paved the way for combining AI with healthcare, focusing on personalized and accessible medical support.

Despite these advancements, there remains a gap in creating integrated systems that combine multi-disease prediction and real-time medical chatbots for widespread accessibility. Most existing solutions either focus on single-disease prediction or lack conversational capabilities for user engagement. Additionally, challenges such as data privacy, lack of scalable infrastructure, and limited datasets for rare diseases continue to hinder the effectiveness of these systems. This project builds upon prior research, aiming to address these gaps by developing a comprehensive platform that uses machine learning for accurate multi- disease prediction and an interactive chatbot for personalized healthcare assistance, while ensuring scalability, affordability, and robust data security.

1. Hamidi. H. and Daraee

In the study by Hamidi and Daraee (2016), titled "Analysis of pre-processing and post- processing methods and using data mining to diagnose heart diseases", the authors explore the role of data mining techniques in enhancing heart disease diagnosis. The research focuses on optimizing pre-processing methods, such as data cleaning and normalization, to improve data quality, and post-processing techniques to refine results. The study highlights the application of algorithms like Decision Trees and SVM, demonstrating the importance of data preparation for reliable medical predictions.

2. Machine Learning for Disease Prediction

Machine learning models have been extensively utilized for disease prediction in various domains. Liang et al. (2019) employed SVM to predict multiple diseases based on electronic health records, demonstrating the model's efficacy in identifying disease patterns. Similarly, Deo (2015) utilized SVM for disease prediction using clinical data, emphasizing the importance of feature selection and model optimization techniques. These studies establish the relevance and effectiveness of machine learning algorithms in disease prediction.

3. Heart Disease Prediction

Several studies have explored the use of machine learning, including SVM, for heart disease prediction. Rajendra Acharya et al. (2017) developed an SVM-based model to predict heart disease using a combination of demographic, clinical, and electrocardiogram (ECG) features. Their study achieved high accuracy in the detecting heart disease, underscoring the potential of SVM in this domain. Additionally, Paniagua et al. (2019) utilized SVM to predict heart disease based on features such as blood pressure, cholesterol levels, and medical history. These studies highlight the applicability and effectiveness of SVM in heart disease prediction.

4. Diabetes Prediction

The prediction of diabetes using machine learning models, including SVM, has garnered significant attention. Poudel et al. (2018) employed SVM to predict diabetes based on clinical and genetic features, demonstrating the model's potential for accurate diabetes risk assessment. Similarly, Al-Mallah et al. (2014) utilized SVM to predict diabetes using features such as glucose levels, body mass index, and blood pressure. These studies underscore the effectiveness of SVM in diabetes prediction and emphasize the importance of incorporating relevant features.

5. Parkinson's Disease Prediction

Machine learning techniques, including SVM, have been explored for the prediction of Parkinson's disease. Tsanas et al. (2012) employed SVM to predict the severity of Parkinson's disease based on voice features, achieving promising results. Additionally, Arora et al. (2017) utilized SVM to predict Parkinson's disease using voice recordings, highlighting the potential of SVM in non-invasive and accessible prediction methods. These studies demonstrate the feasibility of SVM in Parkinson's disease prediction and its potential for early detection.

III. METHODOLOGIES

Problem Identification: Define the key healthcare challenges, such as the need for multi-disease prediction and access to medical advice, and outline project objectives to address these issues.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405050|

Data Collection and Preprocessing: Gather medical datasets containing symptoms and disease correlations from reliable sources. Preprocess the data to remove inconsistencies, handle missing values, and normalize it for model training.

Machine Learning Model Development: Select appropriate machine learning algorithms (e.g., Decision Trees, Random Forest,) to train on the processed dataset. Perform feature selection to identify critical symptoms for accurate disease prediction.

Chatbot Development: Design a conversational chatbot using natural language processing (NLP) frameworks like Dialog flow or Rasa. Train the chatbot to understand and respond to health-related queries and provide guidance.

Integration of Components: Integrate the machine learning-based prediction system with the chatbot to provide seamless user interaction and personalized advice based on predicted diseases.

System Testing and Validation: Perform extensive testing to evaluate the accuracy of disease predictions and the responsiveness of the chatbot. Validate the system using real-world data and gather feedback for improvement.

Deployment: Deploy the application on a cloud platform or local server, ensuring scalability, reliability, and accessibility for users.

Monitoring and Updates: Continuously monitor system performance, update the disease prediction model with new datasets, and improve chatbot capabilities to enhance user experience and accuracy.

IV. ML MODELS

1. Classification Models:

Logistic Regression – Good for binary/multi-class disease prediction (e.g., diabetes: yes/no).
Decision Tree – Simple and interpretable; used for condition-based predictions.
Random Forest – An ensemble method that reduces overfitting and improves accuracy.
Support Vector Machine (SVM) – Effective in high-dimensional spaces, good for small datasets.
K-Nearest Neighbors (KNN) – Works well with well-labeled, small-to-medium datasets.
Naive Bayes – Good for probabilistic prediction with textual or categorical data.
Gradient Boosting Machines – High accuracy and often used in competitions. (e.g., XGBoost, LightGBM)

2. Deep Learning Models:

Artificial Neural Networks (ANN) – Suitable for complex patterns across multiple features. Convolutional Neural Networks (CNN) – Best if working with medical images (X-rays, MRIs). Recurrent Neural Networks (RNN), LSTM – Useful when dealing with time-series patient data.

3. ML/NLP Models for Medical Chatbots:

Natural Language Processing (NLP):

TF-IDF + Logistic Regression – Basic intent detection using keyword-based text features.

Rasa - Open-source NLP framework for chatbots; uses pipelines with DIET, BERT, or CRF entities.

Dialogflow / Microsoft Bot Framework – Prebuilt NLP systems with ML-powered intent handling.

BERT / **BioBERT** / **ClinicalBERT** – Transformer-based models fine-tuned for medical question answering.

Seq2Seq Models (Encoder-Decoder) – Useful for generating context-aware responses. GPT-based Models (e.g., ChatGPT, GPT-3) – For advanced, conversational, human-like interactions.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405050|

V. DATA FLOW DIAGRAM



Fig 5.1: Data flow diagram

VI. PROPOSED SYSTEM

The system integrates two main components:

1. Multiple Disease Prediction Module

2. Medical Chatbot Module

These components work together to help users identify potential health issues based on symptoms and provide preliminary guidance.

1. Multiple Disease Prediction Module

Input: User inputs symptoms via chatbot or web/mobile interface.

Data Preprocessing: Cleaning, encoding symptoms, handling missing values.

Feature Extraction: Mapping symptoms to relevant features.

ML Model(s): Trained using labeled health datasets (e.g., symptom-disease mappings).

Multi-label classification models like: Random Forest, XGBoost, Logistic Regression, ANN (for complex feature relationships).

IJIRSET©2025

| An ISO 9001:2008 Certified Journal |

11735





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405050|

Output: Top probable diseases ranked by prediction probability. Risk score or confidence level.

2. Medical Chatbot Module

Input: User's natural language queries (e.g., "I have a headache and fever").

Processing NLP Pipeline: Tokenization, Lemmatization Intent recognition (e.g., symptom report, treatment query)

Named Entity Recognition (NER) to extract symptoms, duration, intensity.

Dialogue Management: Rule-based for simple queries (e.g., greetings, FAQ), ML-based (e.g., Rasa, Dialogflow, or BERT) for symptom analysis and follow-up.

Integration: Sends structured symptom data to the disease prediction module. Fetches prediction results and explains them conversationally.

Output: Predicted diseases

Basic recommendations (e.g., hydration, rest)

Follow-up questions or redirection to a real doctor to serve.

3. System Modules

A. User Interface Module

A web or mobile interface for users to Input symptoms manually or via checklist. View predicted diseases and suggestions.

B. Data Preprocessing

Module Functions Clean and normalize input symptom data. Encode categorical data for ML models. Handle missing values or vague symptom descriptions.

C. Symptom-Disease Mapping

Mapping Module Uses a structured dataset of Symptoms (features), Diseases (labels) Sources Open health datasets or collected clinical data.

D. Disease Prediction Engine Machine Learning Models

Random Forest – Handles multiple classes, avoids overfitting. Logistic Regression – Fast and interpretable baseline. Naive Bayes – Useful for text/symptom probability. XGBoost/Gradient Boosting – For high-performance resul

E. Results Interpretation Module

This Converts raw predictions into user-friendly output :Disease names, Risk level or urgency, General recommendations (optional)

- 4. Workflow
- 1. User inputs symptoms
- 2. System preprocesses input
- 3. Prediction model processes input
- 4. System outputs most probable diseases

VII. CONCLUSION

The Multi-Disease Prodiction. System and Medical Chatbot represent a significant step toward integrating artificial intelligence into healthcare. By leveraging machine learning models, the system provides accurate predictions for diseases like Diabetes, Heart Disease, and Parkinson's based on user inputs. The integration of a medical chatbot enhances the accessibility of bealthcare by offering personalized guidance and preliminary advice, bridging the gap for individuals who may lack immediate access to medical professionals. This project demonstrates the potential of Al to empower users with early diagnosis and support, enabling timely medical intervention.

The use of Python and machine learning ensures scalability, flexibility, and ease of implementation, while the Streamlit-based user interface provides a seamless and intuitive experience. Testing has confirmed the system's accuracy, performance, and user-friendliness, though improvements in input validation and error handling can further enhance its reliability. Security and privacy have also been prioritized to ensure users feel confident in sharing sensitive health data, adhering to data protection standards.

IJIRSET©2025

| An ISO 9001:2008 Certified Journal |





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405050|

In the future, this project can evolve to include more diseases, deeper medical insights, and multilingual support to cater to a global audience. By continuously integrating advanced algorithms and larger datasets, the system can improve its prediction capabilities and user interactions. The Multi-Disease Prediction System and Medical Chatbot exemplify how Al-driven solutions can address real-world challenges, promoting affordable, accessible, and efficient healthcare for everyone.

REFERENCES

[1] R. Katarya and P. Srinivas, "Predicting heart disease at early stages using machine learning: A survey," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 302–305

[2] Parashar A, Gupta A, Gupta A. Machine learning techniques for diabetes prediction. Int J Emerg Technol Adv Eng. 2023;4(3):672-675.

[3] Arora S, Aggarwal P, Sivaswamy J. Automated diagnosis of Parkinson's disease using ensemble machine learning. IEEE Trans Inf Technol Biomed. 2022;21(1):289-299.

[4] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J.Mach Learn Res. 2011;12:825-830.

[5] Marco Polignano, F. Narducci, Andrea Iovine, Cataldo Musto, Marco De Gemmis, Giovanni Semmeraro, "HealthAssistantBot: A Personal Health Assistant for the Italian Language" in 2020, IEEE based, pp. (99):1-1









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com