



Multiple Disease Prediction System Using Machine Learning

¹Yerrapothu Rupa Sri, ²Sodiseti Navya Sree, ³Pilla Jagadeeswar, ⁴Velpula Chiranjeevi, ⁵Dr. S. Sai Kumar

^{1,2,3,4}B.Tech Student, ⁵Sr. Assistant Professor

^{1,2,3,4,5}Information Technology,

^{1,2,3,4,5}Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, India

ABSTRACT: Early disease detection is crucial – more chances of success, less risk. Standard prediction methods require highly skilled personnel and complex tests, which may delay the process. In this study, we introduce a multiple disease prediction system using Machine Learning and providing rapid, precise, and convenient disease detection. XGBoost is chosen as the algorithm for multiple disease prediction due to high performance, scaling capability, and ability to work with various kinds of input data. The database is cleaned, features are selected, and normalization is applied to improve the quality of prediction. Training of XGBoost on symptom data helps the algorithm detect correlations between symptoms and diseases. Instead of making one suggestion, the output of the model consists of the three top-predicted diseases and their probability scores. A probability-based approach provides a more accurate representation of potential diseases. Alongside the prediction feature, our system includes a hospital recommendation service where the system recommends hospitals depending on predicted diseases. Thus, the user will be directed to appropriate facilities after finding out about his potential condition. A web-based system is designed to help users easily and quickly receive predictions and hospital suggestions. The symptom data is entered by the user, and the results are instantly received. According to experiments, XGBoost improves prediction compared to traditional methods. While not being a replacement for professional healthcare experts, our system serves as an intelligent assistant in disease prediction and risk detection.

Keywords: Multiple Disease Prediction, XGBoost, Machine Learning, Symptom-Based Prediction, Probability Analysis, Hospital Recommendation, Healthcare Analytics, Web-Based System.

I. INTRODUCTION

The number of health problems is increasing rapidly, making early disease detection more important than ever. Identifying diseases at an early stage can help reduce serious complications, lower treatment costs, and improve patient recovery outcomes. However, traditional diagnostic methods often depend on doctors, laboratory tests, and continuous monitoring, which can be time-consuming and may delay immediate medical attention. In recent years, advancements in artificial intelligence (AI) and machine learning (ML) have created new opportunities in the healthcare sector. These technologies can process large volumes of medical data and identify patterns that are difficult to detect using conventional approaches. Machine learning, in particular, is highly effective in predicting diseases based on symptoms and patient data, making the diagnostic process faster and more efficient. Most existing disease prediction systems are designed to identify a single disease using algorithms such as logistic regression, decision trees, random forest, and support vector machines. While these models provide reasonable performance, they often face challenges such as limited accuracy, difficulty in handling complex datasets, and lack of flexibility. Additionally, many systems generate only one predicted outcome, which does not accurately reflect real-world scenarios where multiple diseases may share similar symptoms. To address these limitations, this

project introduces a multiple disease prediction system using machine learning. The system is capable of predicting multiple diseases based on user-provided symptoms. It utilizes the XGBoost algorithm, which is known for its speed, high accuracy, and ability to handle complex and structured data effectively. A key feature of this system is that it does not restrict the output to a single prediction. Instead, it provides the top three most probable diseases along with their corresponding probability values. This approach offers a more realistic understanding of potential health conditions and helps users make better-informed decisions. In addition to disease prediction, the system also recommends suitable hospitals based on the predicted conditions. By suggesting appropriate healthcare facilities, it simplifies the process of seeking medical assistance and enhances the practical usability of the system. To ensure ease of access, the system is developed as a web-based application. Users can enter their symptoms through an online interface and receive instant results, including predicted diseases and hospital recommendations. This real-time functionality makes the system convenient for both individuals and healthcare support purposes.

The major contributions of this project include:

- Development of a machine learning-based system for predicting multiple diseases
- Implementation of XGBoost to improve prediction accuracy and performance
- Providing the top three disease predictions with probability scores
- Integration of a hospital recommendation feature
- Creation of a user-friendly web application for real-time interaction

The proposed system aims to support early disease detection, improve healthcare accessibility, and assist users in making informed decisions. Although it is not a replacement for professional medical diagnosis, it serves as a useful tool for preliminary health assessment and guidance.

II. LITERATURE SURVEY

Machine learning has become an important tool in the healthcare domain for predicting diseases based on patient data and symptoms. It enables faster analysis and supports early diagnosis by identifying patterns in large datasets. Several studies and project-based implementations have explored different approaches for building disease prediction systems, focusing on improving accuracy, usability, and real-world applicability. JIEEE – Journal of Informatics, Electrical, and Electronics Engineering [1] discussed the use of machine learning techniques for disease prediction systems. Their work demonstrates how various algorithms can be applied to medical datasets to identify diseases at an early stage. The study highlights the feasibility and effectiveness of using machine learning models for multi-disease prediction tasks. The project presented by nevonProjects.com [2] provides a practical implementation of a multiple disease prediction system using machine learning. It outlines the system architecture, including data preprocessing, model training, and user interface design. This work serves as a reference for developing similar systems and understanding the overall workflow of integrating prediction models into real-world applications. Tutorialspoint [3] explains the concept of Data Flow Diagrams (DFD), which are essential for representing the flow of data within a system. Understanding DFDs helps in designing and visualizing the structure of the proposed system, including input processing, prediction modules, and output generation. Datasets from Kaggle have been widely used in disease prediction research due to their accessibility and structured format. The Heart Disease Dataset [4] provides clinical parameters that help in identifying cardiovascular-related conditions. It is commonly used to train and evaluate machine learning models for heart disease prediction. The Diabetes Prediction Dataset [5] includes features such as glucose levels, BMI, and other health indicators, enabling the development of models that can predict diabetes effectively. This dataset supports the creation of reliable prediction systems by offering real-world health data. Similarly, the Liver Disease Dataset [6] contains medical attributes that assist in identifying liver-related disorders. It helps in training models to recognize patterns associated with liver diseases and contributes to building multi-disease prediction systems. Overall, these studies and resources highlight the growing importance of machine learning in healthcare. They also emphasize the need for systems that can handle multiple diseases, provide accurate predictions, and offer user-friendly interfaces. The proposed system builds upon these ideas by using advanced algorithms and integrating features such as probability-based predictions and hospital recommendations.

III. PROPOSED SYSTEM

The proposed system introduces a multiple disease prediction model that boosts accuracy and gives useful results based on the user's symptoms. Unlike older systems that only give one possible disease, this model uses a probability method to show the top three predicted diseases along with their scores, making it easier to understand potential health issues. The system uses the XGBoost algorithm, which is good at learning patterns from symptom data and gives dependable predictions. In addition to predicting diseases, the system also recommends hospitals that are suitable for the identified conditions, helping users take the right steps next. The process involves collecting data, cleaning and preparing it, selecting important features, training the model, making predictions with probability scores, and finally deploying the system on a web platform. Overall, the system improves accuracy, provides instant results, and supports better decisions in healthcare.

IV. METHODOLOGY

The methodology outlines the steps involved in developing a system for predicting multiple diseases using machine learning. The approach consists of several stages, including data collection, data preprocessing, model development, and evaluation of the model's performance.

4.1 Dataset Description

The dataset used in this study helps predict multiple diseases based on common symptoms and basic patient information.

It includes both personal details and symptoms in a structured way.

Main Features in the Dataset are the following:

- Age
- Gender
- Fever
- Cough
- Headache
- Fatigue
- Chest Pain
- Breathlessness
- Nausea
- Vomiting
- Body Pain
- Joint Pain
- Sore Throat
- Dizziness
- Weight Loss
- Blood Pressure Level
- Sugar Level

Each symptom is either present or not, shown as 1 for present and 0 for not present.

The goal is to predict which disease a person might have.

This dataset is used to train the XGBoost model to find patterns between symptoms and diseases, making mult-disease prediction more accurate and efficient.

4.2 Data Preprocessing

Data preprocessing is a key part of creating the disease prediction system. The dataset might have problems like extra spaces, wrong formatting, or categories that need to be turned into numbers.

In this project, the following steps are performed to prepare the data:

- Removing unnecessary spaces and fixing the column names
- Changing categorical data like gender into numerical values
- Using label encoding for the disease column
- Ensuring all symptom values are either 0 or 1
- Checking the dataset for missing or incorrect information

These steps help make the data clean and ready for training the XGBoost model. After preprocessing, the dataset is split into training and testing sets for model development and evaluation.

4.3 Model Description

The proposed system uses the XGBoost algorithm to predict diseases based on user inputs like symptoms, age, and gender. XGBoost is a powerful machine learning method that combines multiple decision trees to improve prediction accuracy and handle structured data effectively. Instead of providing just one result, the model displays the top three possible diseases along with their probability scores, giving a clearer understanding of potential conditions. The system workflow is also represented using a data flow diagram (DFD), which explains how data moves from user input to the final prediction output.

4.4 Prediction Approach

The system uses a probability-based method with the XGBoost model. Rather than using several models together, it relies on one trained model to make predictions based on information like symptoms, age, and gender. The model gives a probability score for each disease and picks the top three most likely ones. This way, the system offers several possible results instead of just one, which makes it easier to use and more realistic, as it better reflects what happens in real life.

4.5 System Flow

User Input -> Data Processing -> Feature Extraction -> Machine Learning Model -> Diagnosis Prediction -> Hospital Recommendation -> Result.

4.6 System Architecture

This architecture ensures smooth collaboration between the frontend, backend, and machine learning elements. The patient provides their symptoms; these are further processed and fed into the machine learning algorithm.

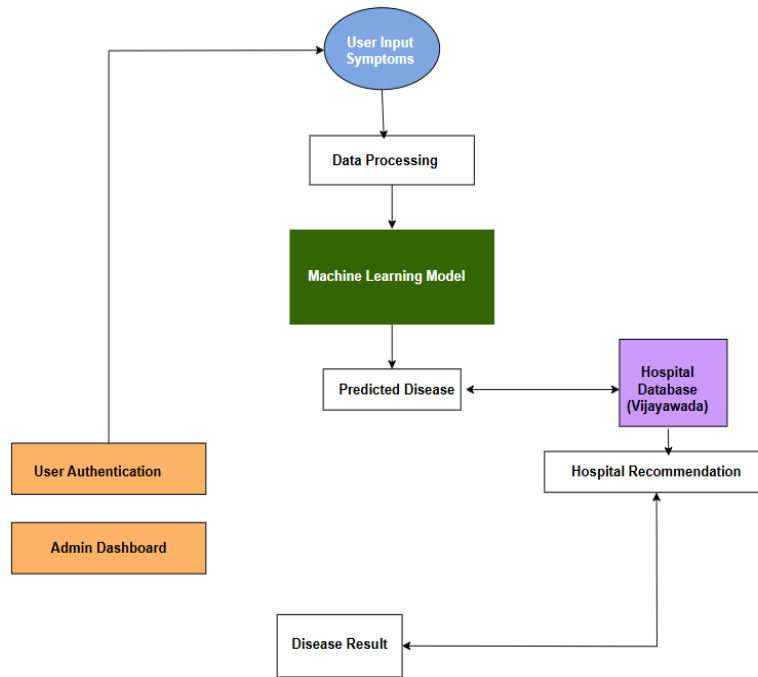


Fig. 1: System Architecture of Disease Prediction System with Hospital Recommendation.

V. RESULTS

The proposed system was evaluated using a symptom-based dataset, where the XGBoost model achieved an accuracy of 77% in predicting diseases. The system provides the top three predicted diseases with probability scores, offering more informative results than single predictions. The web application generates fast and reliable outputs and also suggests suitable hospitals based on the predictions. Overall, the system is efficient, user-friendly, and useful for early disease detection and healthcare guidance.

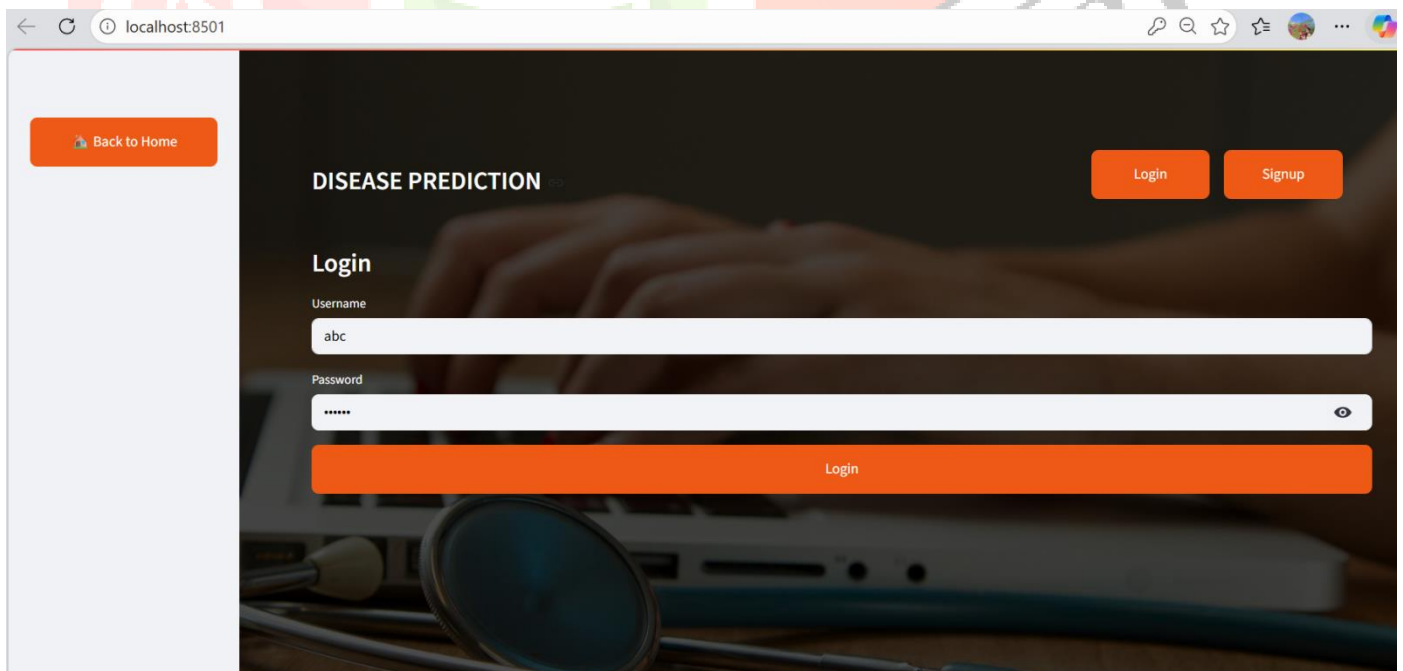


Fig. 2: User Login Interface for the Multiple Disease Prediction System

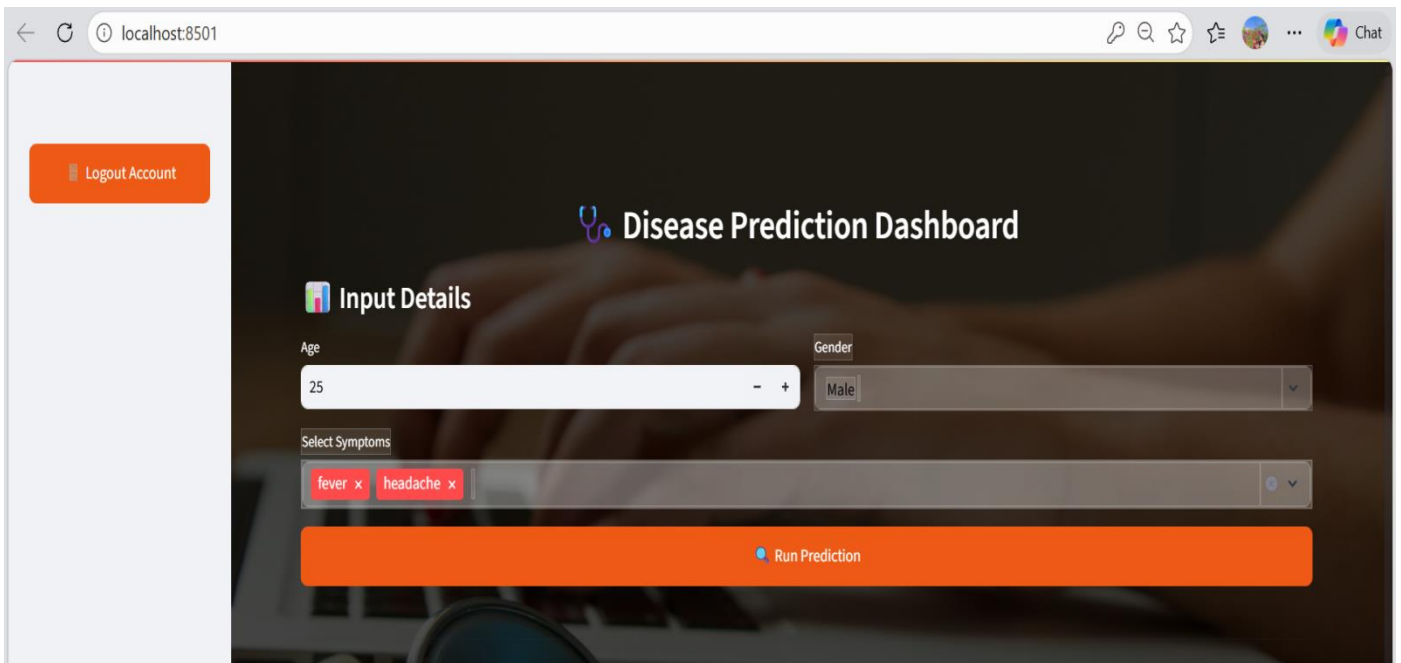


Fig. 3: Disease Prediction Dashboard with User Input Interface

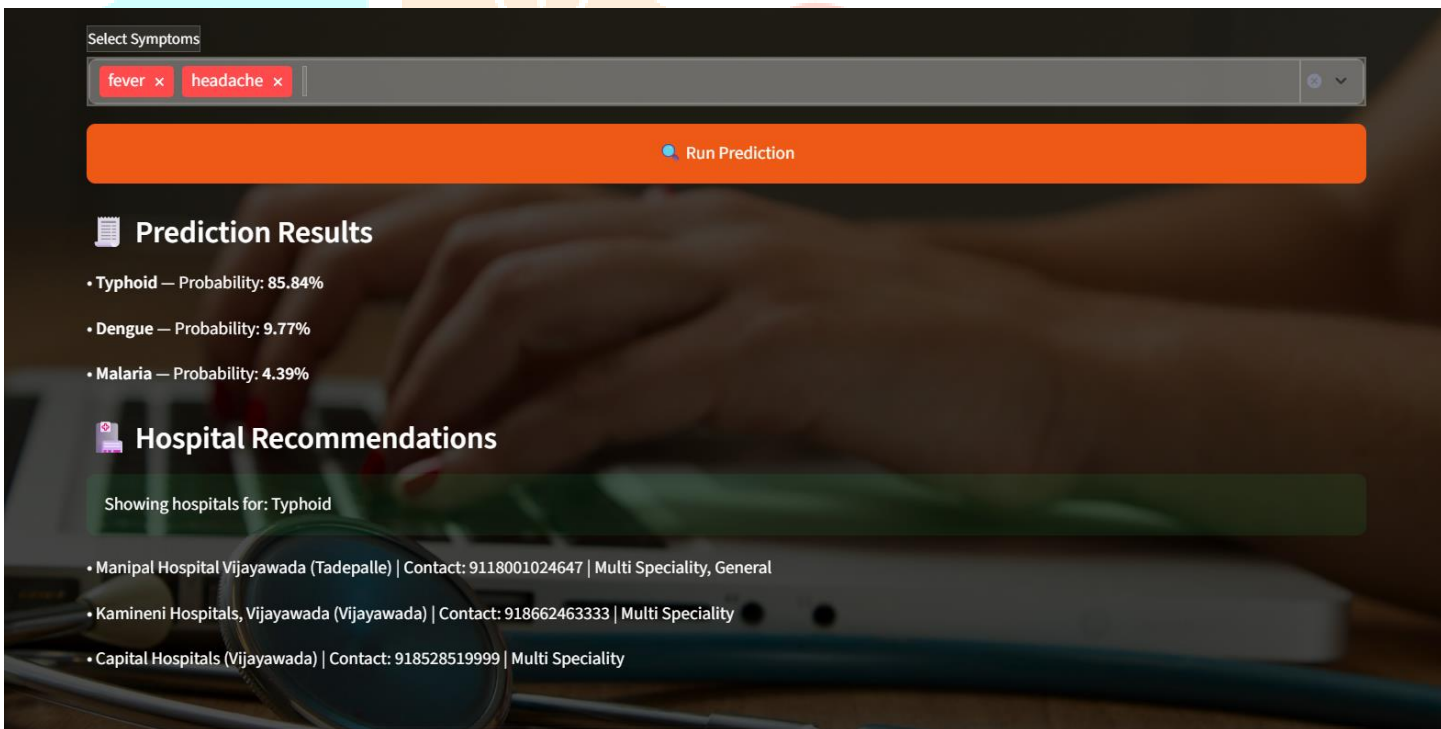


Fig. 4: Prediction Results Display with Top Three Diseases and Hospital Recommendations

VI. WEB APPLICATION

A web-based application is developed to make the system accessible and easy to use. It allows users to enter their symptoms and receive real-time disease predictions.

Features of the Web Application:

- User-friendly symptom input form
- Real-time disease prediction
- Top three predicted diseases with probability scores
- Hospital recommendation based on predictions
- Simple and interactive interface

This system helps users quickly understand possible health conditions and take appropriate steps for medical care.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to DR. S. SAI KUMAR, Ph.D., Sr. Assistant Professor, for his valuable guidance, constant support, and encouragement throughout the development of this project. His insightful suggestions and continuous motivation played a significant role in the successful completion of this work.

The authors also extend their heartfelt thanks to the Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology, for providing the necessary infrastructure, resources, and technical support required for carrying out this project. The supportive academic environment and encouragement from the faculty members greatly contributed to enhancing the quality of this work.

Finally, the authors express their sincere appreciation to the management of Prasad V Potluri Siddhartha Institute of Technology for their continuous support and encouragement in successfully completing this project and its presentation.

CONCLUSION

The proposed Multiple Disease Prediction System using machine learning provides an efficient and user-friendly solution for early disease detection based on symptoms. By utilizing the XGBoost algorithm, the system achieves reliable prediction performance and generates the top three probable diseases along with their probability scores, offering more informative insights than traditional single-output models. The integration of a hospital recommendation feature further enhances its practical usefulness by guiding users toward appropriate healthcare services. With its web-based interface, the system ensures real-time accessibility and ease of use, making it a valuable tool for preliminary health assessment and decision support. Although it is not a replacement for professional medical advice, it serves as an effective aid for early diagnosis and improved healthcare awareness.

REFERENCES

- [1] JIEEE – Journal of Informatics, Electrical, and Electronics Engineering. (n.d.). Machine Learning–Based Disease Prediction System.
— Referenced to justify the feasibility of using machine learning techniques for multi-disease prediction.
- [2] nevonProjects.com. (n.d.). Multiple Disease Prediction System using ML.
— Example project description and reference for similar system architecture and implementation approach.

- [3] Tutorialspoint, Data Flow Diagram (DFD) – System Analysis and Design.
— Referred to know about the concept behind the data flow.
- [4] Kaggle. (n.d.). Heart Disease Dataset. Kaggle Inc.
— Dataset used for heart disease prediction (johnsmith88/heart-disease-dataset).
- [5] Kaggle. (n.d.). Diabetes Prediction Dataset. Kaggle Inc.
— Dataset used for diabetes prediction (iammustafatz/diabetes-prediction-dataset).
- [6] Kaggle. (n.d.). Predict Liver Disease – 1700 Records Dataset. Kaggle Inc.
— Dataset used for liver disease prediction (rabieelkharoua/predict-liver-disease-1700-records-dataset).

