Enhancing The Recommendation Model For Disease Predictionbased On User Symptoms Using Machine Learning

Sasikala M[1], Dr.Josephine Mary L[2]
[1] Dept of MCA
[2]Assistant Professor, Dept of MCA
[1, 2] Sri Muthukumaran Institute of Technology Chennai, India,

Abstract- *Healthcare is one of the most important research fields with the rapid improvement of technology and increase in data. It is difficult to handle huge amounts of patient data. Big Data Analytics can be used to handle such data. There are a lot of procedures for the treatment of multiple diseases across the world. Machine Learning is a prominent approach that helps in prediction and diagnosis of a disease.In the existing system,diagnosis at an early stage can be difficult as a lot of diseases might have very common symptoms and require a professional to identify the illness. Besides, a lot of patients delay visiting a doctor because of the lack of information and availability. Therefore, to tackle this problem, a system that will be able to give some hints about the patient's health issues is developed. The proposed system, "Enhancing the Recommendation Model for Disease Prediction Based on User Symptoms Using Machine Learning," presents an effective approach for predicting diseases and providing user-centered health recommendations. By integrating multiple machine learning algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors through an Ensemble Voting Model, the system achieves improved prediction accuracy, robustness, and reliability compared to traditional single-model approaches. The implementation of pre-processing and feature engineering techniques enhances data quality and optimizes model performance, enabling the system to handle noisy and diverse symptom data effectively. In addition to disease prediction, the inclusion of a recommendation module provides meaningful precautions and lifestyle suggestions, increasing the practical usefulness of the system for end users. The developed interface ensures a user-friendly and accessible environment for symptom analysis and prediction. Performance evaluation demonstrates that the proposed approach produces accurate and consistent results while supporting continuous improvement through feedback mechanisms.*

Keywords: Disease Prediction, Machine Learning, Symptom Analysis, Healthcare AI, Clinical Decision Support, Explainable AI

I. INTRODUCTION

Diseases often begin with mild symptoms that gradually affect a person's health and are not easily recognized during the early stages. Therefore, early diagnosis is essential for effective treatment and prevention. In medical practice, diagnosis is generally based on patient symptoms, medical history, and clinical examination. However, increasing patient volume and limited consultation time make it difficult to analyze every symptom carefully. In addition, many diseases share similar symptoms, which increases the complexity of accurate diagnosis during the initial stages. Most existing disease prediction systems rely on limited datasets or focus on individual symptoms rather than analysing multiple symptoms collectively. Such approaches fail to capture the complexity of real-world medical conditions where symptoms often overlap and evolve together. As a result, prediction accuracy and reliability are reduced, especially when handling noisy or incomplete symptom data.

Recent advancements in machine learning and artificial intelligence provide significant opportunities to improve disease prediction systems. Machine learning algorithms can identify hidden patterns in symptom data and support accurate disease prediction. The proposed system adopts a comprehensive approach by analysing multiple symptoms collectively using machine learning techniques. By learning symptom–disease relationships from structured datasets, the system provides accurate and reliable predictions, supporting early disease identification and encouraging timely medical consultation.

1.1 PROBLEM STATEMENT

Accurate and early disease prediction based on user symptoms remains a major challenge in modern healthcare systems. Although machine learning techniques have improved prediction accuracy, many existing systems focus mainly on classification performance while neglecting recommendation mechanisms. Most systems rely on single algorithms, which often produce inconsistent results when handling noisy, incomplete, or diverse symptom data. Existing disease prediction systems also lack effective integration of multiple machine learning algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors. In addition, limited attention is given to pre-processing and feature engineering techniques, which reduces model efficiency, reliability, and overall performance. Another important limitation is the absence of a structured recommendation mechanism. Many applications provide only disease predictions without offering precautions, lifestyle suggestions, or healthcare guidance, reducing their practical usefulness for users.

Therefore, there is a need to develop an enhanced disease prediction system that improves prediction accuracy through the integration of multiple machine learning models while also incorporating an effective recommendation mechanism. The system should efficiently handle noisy symptom data, apply proper feature selection, and provide reliable and user-friendly healthcare support.

1.2. OBJECTIVES

- To develop an enhanced disease prediction system using machine learning based on user symptoms
- To improve prediction accuracy by integrating multiple algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors
- To design an efficient data pre-processing mechanism for handling missing, noisy, and inconsistent symptom data
- To implement effective feature engineering techniques to select the most relevant symptoms and improve model performance
- To develop a recommendation module that provides precautions and lifestyle suggestions based on predicted diseases
- To ensure the system is user-friendly and provides easily understandable results
- To evaluate and compare the performance of different machine learning models
- To incorporate a feedback mechanism for continuous system improvement and learning
- To build a scalable system that can be extended with additional datasets and advanced techniques in the future

The existing systems for disease prediction based on user symptoms primarily rely on basic machine learning models or rule-based approaches to identify possible diseases. Most systems take user symptoms as input and compare them with predefined datasets to generate predictions. Commonly used algorithms include Decision Tree, Naïve Bayes, and K-Nearest Neighbors, which provide moderate accuracy but often struggle with complex and large-scale data.

Most existing approaches focus mainly on disease prediction rather than providing meaningful recommendations. These systems generally output only the predicted disease without offering precautions or lifestyle suggestions, which limits their practical usefulness for end users.

Another limitation is the use of single-model techniques, which reduces robustness and may lead to inconsistent results when handling noisy or incomplete symptom data. In addition, limited preprocessing and feature selection techniques negatively affect model performance and accuracy.

Furthermore, many systems lack interpretability, user-friendly interfaces, and feedback mechanisms for continuous improvement. Overall, existing disease prediction systems provide only a basic framework for disease identification and fail to achieve higher accuracy, reliability, recommendation capability, and user-centered healthcare support.

The proposed system develops an enhanced disease prediction and recommendation model that improves accuracy, reliability, and usability by integrating multiple machine learning techniques with an intelligent recommendation mechanism. Unlike existing systems that rely on single-model approaches, the proposed system combines algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors to achieve higher prediction accuracy and robustness through an ensemble approach. The system collects user symptoms through an interactive interface and performs preprocessing to handle noisy, missing, and inconsistent data. Feature engineering techniques are applied to select relevant symptoms and optimize model performance. The machine learning layer processes the refined data and generates disease predictions using multiple algorithms. The outputs are combined using a voting mechanism to produce a final and more accurate prediction. This multi-model strategy improves system stability and reliability.

A major feature of the proposed system is the recommendation engine, which provides precautions, lifestyle suggestions, and healthcare guidance based on the predicted disease. The system also includes a knowledge base, database support, and a feedback mechanism for continuous learning and performance improvement.

Overall, the proposed system provides a comprehensive, scalable, and user-friendly solution by combining accurate disease prediction with effective healthcare recommendations and continuous system enhancement.

1.3. METHODOLOGY

The working process of the system consists of several simple steps:

1. Collect symptom–disease dataset and gather user input symptoms
2. Perform data pre-processing by cleaning, handling missing values, and encoding data
3. Apply normalization to standardize the dataset
4. Conduct feature engineering to select important symptoms and remove irrelevant data
5. Split the dataset into training and testing sets
6. Train multiple machine learning models such as Random Forest, Support Vector Machine, Naïve
7. Bayes, Decision Tree, and K-Nearest Neighbors
8. Combine model outputs using an ensemble method (e.g., voting)
9. Predict disease based on user symptoms
10. Generate recommendations such as precautions and lifestyle suggestions
11. Evaluate model performance using metrics like accuracy, precision, recall, and F1-score
12. Incorporate feedback for continuous improvement of the system

The remainder of this paper is organized as follows. Section II describes with Literature survey, Section III deals with the implementation of enhancing the recommendation model for disease prediction based on user symptoms using machine learning. Section IV discusses with conclusion and future enhancement.

II. LITERATURE SURVEY

Ramachandiran*et al.* proposed a symptom-based disease prediction system using machine learning techniques for early diagnosis. Their study demonstrated that machine learning algorithms can effectively analyze symptom patterns and improve disease prediction accuracy. However, the system mainly focused on disease classification and lacked a recommendation mechanism for user guidance [1].Mahajan *et al.* presented a detailed study on ensemble learning methods in healthcare applications. The research highlighted that ensemble techniques improve prediction accuracy and robustness by combining multiple classifiers. Methods such as voting, bagging, and boosting were found to outperform individual machine learning models in disease prediction tasks [2].Saha*et al.*analyzed hybrid ensemble frameworks for healthcare prediction systems and reported that ensemble models integrated with feature selection techniques achieved higher classification performance. Their study emphasized the importance of combining multiple machine learning algorithms to improve prediction stability and reliability [3].
Several studies have focused specifically on symptom-based disease prediction using machine learning algorithms such as Random Forest, Support Vector Machine, and Naïve Bayes. Research findings showed that ensemble voting approaches significantly reduced prediction errors and improved the consistency of healthcare prediction systems [4].Islam *et al.* reviewed machine learning techniques for chronic disease prediction and concluded that pre-processing and feature engineering are essential for improving model performance. The study also highlighted the role of machine learning in early detection of diseases such as diabetes, heart disease, and kidney disorders [5].Recent studies on electronic healthcare systems demonstrated that ensemble learning models provide better predictive capability compared to standalone classifiers when handling large and complex medical datasets. The research further emphasized the importance of healthcare recommendation systems in supporting clinical decision-making [6].

Rajora*et al.* developed a disease prediction and recommendation system using machine learning algorithms including Decision Tree, K-Nearest Neighbors, and Random Forest integrated through ensemble voting. Their system provided disease prediction along with basic recommendations and demonstrated improved prediction accuracy [7].Recent advancements in healthcare AI have also focused on multilingual symptom datasets and personalized healthcare systems. Ensemble learning approaches using hard and soft voting mechanisms achieved high prediction accuracy while improving healthcare accessibility and user interaction [8].Furthermore, deep learning and ensemble-based disease diagnosis systems have shown improved predictive capability for chronic disease analysis. However, many existing systems still lack interpretability, efficient recommendation support, and user-centered healthcare guidance [9].

III. IMPLEMENTATIONOF ENHANCING THE RECOMMENDATION MODEL FOR DISEASE PREDICTION BASED ON USER SYMPTOMS USING MACHINE LEARNING.

3.1. SYSTEM ARCHITECTURE

The System architecture of proposed system is shown in Fig 1.

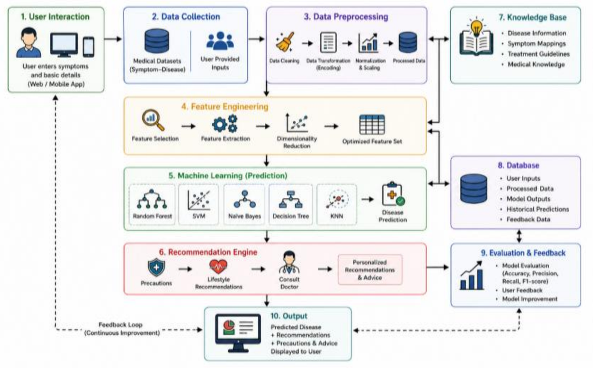

Fig.1.System Architecture

3.1.1. User Interaction

The process starts with the **User Interaction Layer**, where the user enters symptoms and basic personal details such as age or gender through a web or mobile interface. This layer acts as the entry point of the system, ensuring that the input is collected in a structured and user-friendly way.

3.1.2. Data Collection

The data then moves into the **Data Collection Layer**, where the system combines user-provided symptoms with existing medical datasets that contain relationships between symptoms and diseases. All this information is stored as raw data, forming the base for further processing.

3.1.3. Data Pre-processing

Next, the system enters the **Data Preprocessing Layer**, which prepares the raw data for machine learning. In this stage, missing values and duplicate entries are removed to ensure data quality. The system converts categorical symptom data into numerical form using encoding techniques, and normalization is applied so that all features are on a similar scale. By the end of this layer, the data becomes clean and consistent.

3.1.4.Feature Engineering

After preprocessing, the data flows into the **Feature Engineering Layer**, where the system improves the quality of input features. Important symptoms are selected while irrelevant ones are removed. New features may also be created to capture hidden patterns in the data. If the dataset is large, dimensionality reduction techniques are used to simplify it while keeping essential information. This step ensures that the model receives the most meaningful inputs.

3.1.5. Machine Learning

The refined data is then passed to the **Machine Learning Layer**, which is the core of your system. Here, multiple algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors are trained and used for prediction. Each model analyzes the symptoms differently, and their outputs are combined using a voting or aggregation method. This multi-model approach increases prediction accuracy and reliability.

3.1.6.Recommendation Engine

Once the disease is predicted, the system moves to the Recommendation Engine, which transforms the prediction into useful guidance. Instead of providing medications, the system focuses on precautions, lifestyle suggestions, and advice on whether medical consultation is needed.

3.1.7. Knowledge Base

In parallel, the system uses a Knowledge Base, which stores structured medical information such as symptom-disease relationshipsand general healthcare knowledge. This supports both prediction and recommendation processes.

3.1.8.Database Layer

The Database Layer works in the background, storing user inputs, processed data, model outputs, and historical records. This storage is essential for tracking system performance and enabling future improvements.

3.1.9. Evaluation and Feedback Layer

Finally, the Evaluation and Feedback Layer continuously monitors how well the system performs. It evaluates accuracy and collects user feedback, which is then used to retrain and improve the models. This creates a feedback loop, allowing the system to become more accurate over time. Overall, the architecture follows a logical flow where user input is gradually refined, analyzed using multiple machine learning techniques, and converted into meaningful predictions and recommendations, with continuous learning built into the system.

3.1.10.Output Layer

The results are then shown in the **Output Layer**, where the predicted disease and recommendations are displayed clearly to the user. This layer ensures that complex model results are presented in a simple and understandable format.

3.2. RESULT AND DISCUSSION

The interface represents the front-end working architecture of the disease prediction and recommendation system. It acts as the entry point where users interact with the application through a secure and user-friendly environment. The system provides separate access for patients and administrators, ensuring proper management and monitoring of healthcare data is shown in **Fig. 2.**

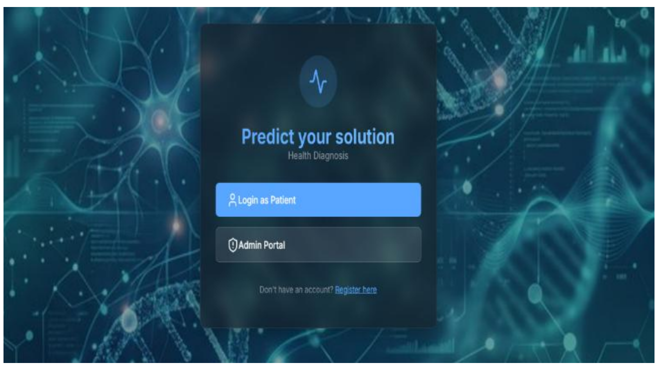

Fig.2Front-end Working Architecture of disease prediction and recommendation system

When the patient logs into the system, symptoms and health-related details are entered into the application. These inputs are then processed through the backend machine learning architecture, where pre-processing, feature engineering, and prediction models analyze the symptom data to identify possible diseases. Based on the prediction results, the system generates recommendations such as precautions and lifestyle suggestions.

Frontend:

BuiltusingHTMl,CSS and JavaScript. This is the website interface where users can input the symptoms to predict disease.

Backend:

Built using python and Flask. This connects the website to the scrapers and the AI model.

3.2.1.Ensemble Voting Model:

An Ensemble Voting Model is a machine learning technique that combines the predictions of multiple algorithms to produce a final prediction. Instead of depending on a single model, the system uses several classifiers together and selects the output based on majority voting. This improves prediction accuracy, reliability, and robustness. The ensemble voting model combines the outputs of Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors. Each model independently predicts the disease based on user symptoms. The final disease prediction is then selected using a voting mechanism where the prediction receiving the highest number of votes becomes the final output. The ensemble approach reduces the weaknesses of individual models and utilizes their strengths collectively.

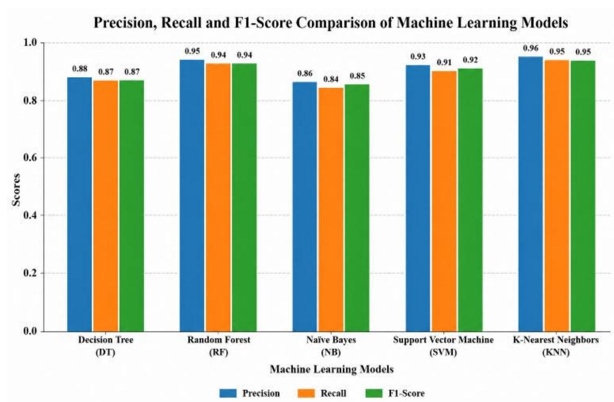

Fig.3 Performance Comparison of Machine Learning Algorithms

Fig. 3 shows the comparison of Precision, Recall, and F1-Score for different machine learning algorithms used in the proposed system. K-Nearest Neighbors and Random Forest achieved better performance compared to other models, demonstrating improved prediction accuracy and reliability.

3.2.2. Machine Learning Module:

The Machine Learning Module is the main part of the disease prediction system. It analyzes the user symptoms and predicts the possible disease using different machine learning algorithms. After pre-processing and feature engineering, the cleaned data is given to the models for prediction. In this paper, algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors are used. Each model predicts the disease separately, and the final result is selected using an Ensemble

Voting Model. This module improves prediction accuracy, reliability, and overall system performance.

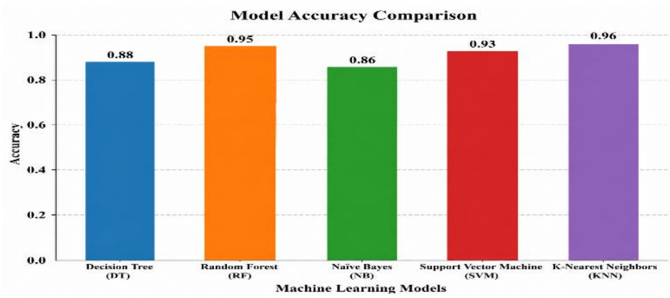

Fig.4Model Accuracy Comparison of Machine Learning Algorithms

Fig. 4 shows the accuracy comparison of different machine learning models used in the proposed system. K-Nearest Neighbors and Random Forest achieved higher accuracy compared to other algorithms, demonstrating improved prediction performance and reliability.

Deployment:

Cloud services like AWS, Google Cloud, or Microsoft Azure can be used to host the model, manage storage.

The following **Fig.5**showsthe Create Account page of the Disease Prediction system. It allows new users to register and access the platform for disease prediction services. This module enables secure user registration and stores patient information required for personalized disease prediction and health analysis.

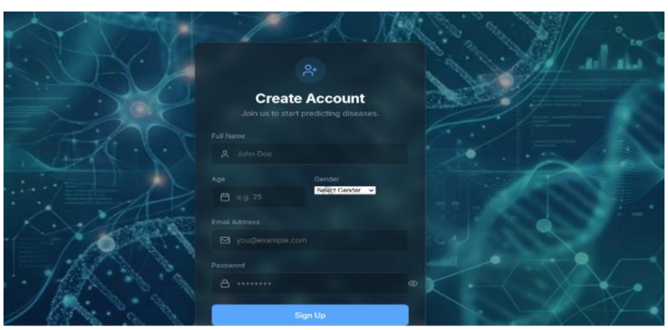

Fig.5.User Registration Page

Users can start diagnosis by entering symptoms to predict disease is shown in **Fig.6.**

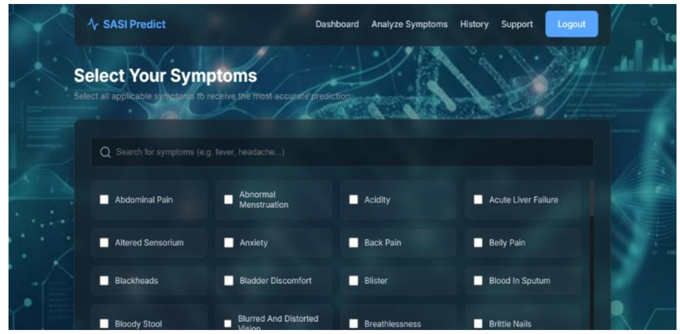

Fig.6 Analyze the Symptoms Page

It allows users to select multiple symptoms through an interactive and searchable interface. The selected symptoms are processed by the backend machine learning model to predict possible diseases with improved accuracy.

The **Fig. 7a) and b)**shows the Disease Prediction Result.It displays the predicted disease based on the user's entered symptoms.This module helps users understand their predicted health condition and provides basic guidance and precautionary measures for better healthcare awareness.

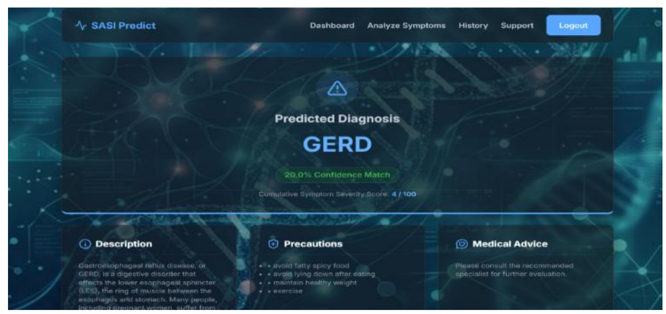

Fig. 7.a) Predicted Result Page

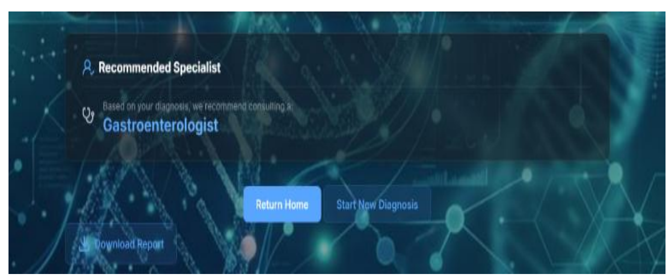

Fig.7.b)Prediction Result Page

Fig. 8.shows the admin portal login page of a Disease Prediction System. It is designed for authorized administrators to securely access the system.

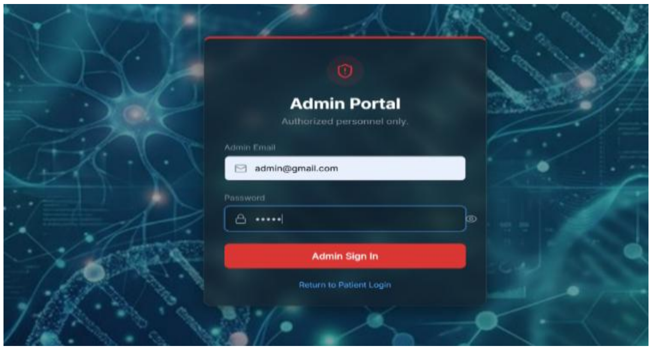

Fig.8Admin Portal Login Page

Fig.9 shows the Admin Dashboard Page of the Disease Prediction system. It provides administrators with an overview of platformactivities and system management features.

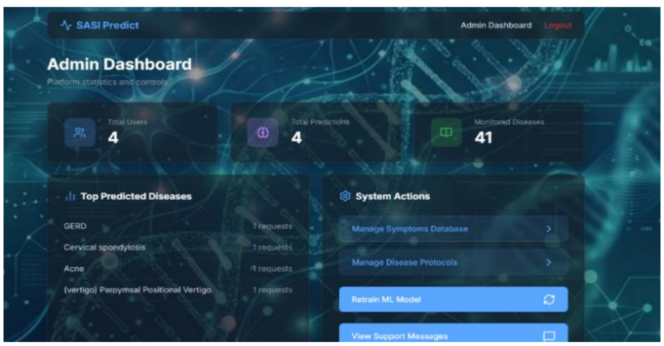

Fig.9Admin Dashboard

The proposed disease prediction system achieved improved accuracy and reliability by using an **Ensemble Voting Model** that combines Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree, and K-Nearest Neighbors. Preprocessing and feature engineering improved data quality and enhanced model performance. In addition to prediction, the recommendation module provided useful precautions and lifestyle suggestions, making the system more practical and user-friendly.

IV. CONCLUSION

A symptom-based Disease Prediction System is to help identify potential health conditions using machine learning algorithms**.** Firstly, users will need to provide the input data, and then the entered data will undergo analysis and processing. The symptoms provided by the user will be transformed into the correct format to make them suitable for analysis using machine learning. The system will analyze the user-entered symptoms and compare them with the data in the database where different diseases are stored and linked to particular symptoms. Thus, the system will predict the possible disease the person suffers from. Moreover, the use of explainable artificial intelligence methods in the system allows explaining how predictions were obtained, which adds a lot to the quality of the output. The experiment showed good predictionresults. Overall, the system provides a scalable, efficient, and intelligent healthcare support solution that can assist users in early disease identification and informed decision-making. Future enhancements may include real-time medical data integration, advanced deep learning techniques, and personalized healthcare analytics for further improving system capability and performance.

l

REFERENCES

[1] R. Ramachandiran*et al.*, “Symptoms Based Disease Prediction Using Machine Learning,” *International Journal of Research in Engineering and Science*, vol. 11, no. 5, pp. 45–50, 2023.

[2] A. Mahajan *et al.*, “Ensemble Learning Methods in Healthcare Prediction Systems,” *Healthcare*, vol. 11, no. 12, pp. 1–20, 2023.

[3] S. Saha*et al.*, “Hybrid Ensemble Frameworks for Healthcare Prediction,” *Journal of Intelligent Information Systems*, vol. 62, no. 3, pp. 455–472, 2024.

[4] P. Kumar and R. Sharma, “Symptom-Based Disease Prediction Using Ensemble Machine Learning,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 210–218, 2023.

[5] M. Islam *et al.*, “Machine Learning Techniques for Chronic Disease Prediction,” *Journal of Engineering Science and Technology Innovation*, vol. 9, no. 1, pp. 35–48, 2024.

[6] J. Lee *et al.*, “Ensemble Learning for Electronic Healthcare Systems,” *Digital Health*, vol. 9, pp. 1–15, 2023.

[7] M. Rajora*et al.*, “Disease Prediction and Recommendation System Using Machine learning,” *arXiv preprint arXiv: 2106.02813*, 2021.

[8] A. Verma and S. Patel, “Multilingual Symptom-Based Disease Prediction Using Ensemble Voting,” *International Journal of Artificial Intelligence in Healthcare*, vol. 5, no. 2, pp. 120–130, 2025.

[9] K. Zhang *et al.*, “Deep Ensemble Learning for Chronic Disease Diagnosis,” *IEEE Access*, vol. 10, pp. 65432–65445, 2022.