IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

AI Based Disease Prediction and Remedy Recommendation System

Kavya B G[1] and Dr. Kruthi R[2]
Postgraduate Student, Department of MCA[1]
Associate Professor, Department of MCA[2]
Maharaja Institute of Technology, Mysore

Abstract: *AI and ML are being utilized for enhanced treatment, diagnosis, and prediction of diseases. However, most of the applications provide generalized treatment based on the symptoms provided with lesser emphasis on patient's health factors. In this paper, we present an AI-based system for predicting diseases along with recommending remedies. This is achieved by employing Random Forest algorithm and considering the patients' symptoms in addition to age, gender, blood pressure, blood sugar, cholesterol, BMI, heart rate, hemoglobin, stress, and sleep duration. Further, the suggested information includes the causes, diagnostic tests, diet chart, exercise, lifestyle, home remedies, prevention measures, and specialist consultation. The proposed application has been implemented as a web application with custom API and stack, user authentication, language translation, electronic health records, admin panel, and PDF report generation. The experimental results show that the Random Forest algorithm is effective in fitting the data with accurate prediction of diseases. The proposed AI-based tool can be considered as an intelligent companion for screening diseases with improved awareness and prediction for taking necessary precautions and seeking medical treatment.*

Keywords: Artificial Intelligence; Machine Learning; Random Forest; Disease Prediction; Healthcare Recommendation; Flask; Health Monitoring; Preventive Healthcare

I. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) are increasingly being leveraged in the healthcare domain for efficient diagnosis and prediction of diseases. These approaches are useful in assisting doctors by analyzing a patient's data to recommend appropriate treatment, thus improving the healthcare provided to patients. One of the critical aspects of providing effective healthcare is the accurate prediction of disease. Even though AI and ML have shown significant potential in disease prediction and diagnosis, most of the current systems are only focused on predicting diseases based on symptoms and recommend generalized treatment. Further, many of these systems do not consider the vital signs such as blood pressure, blood sugar, cholesterol, Body Mass Index (BMI), heart rate, hemoglobin, stress levels, and sleep duration. Also, such systems only predict the disease without providing further information and support for treatment. Thus, this paper presents an AI-based approach for disease prediction along with recommending treatment and diet. The Random Forest algorithm is employed to predict diseases based on symptoms and personal health factors for improved accuracy. These factors include age, gender, blood pressure, blood sugar, cholesterol, BMI, heart rate, hemoglobin, stress, and sleep duration. Apart from disease prediction, the proposed system will provide information on the causes, diagnostic tests, diet chart, exercise, lifestyle, home remedies, prevention measures, and specialist consultation. Besides building an AI model for disease prediction with improved accuracy, the system will also be implemented as a web application providing an intuitive interface with features such as login module, language translation, electronic health records, PDF report generation, and an admin dashboard. Overall, the proposed approach will aim to improve the early screening of diseases and recommend appropriate treatment and diet to patients.

A. Contribution

The main contributions of this work include:

- An AI-based web application for screening diseases and recommending remedies.
- Prediction of diseases using Random Forest ML algorithm considering the patient's symptoms along with personal health factors such as age, gender, blood pressure, blood sugar, cholesterol, BMI, heart rate, hemoglobin, stress, and sleep duration.
- Recommending diet charts, exercises, lifestyle, causes, diagnostic tests, home remedies, prevention measures, and specialist consultation for the predicted diseases.
- Implementation of the system as a web application using Python, Flask, HTML5, CSS3, Bootstrap, and JavaScript to provide an intuitive interface to end-users.
- Design and development of additional features such as language translation, electronic health records, PDF report generation, and dashboard for improved user experience.

Overall, the proposed system provides timely prediction of diseases to improve awareness and aid in taking precautions. The web application will enable easy access to the system to help people stay healthy by providing relevant information on treating or preventing the diseases.

II. RELATED WORK AND RESEARCH GAP

AI and ML have been effectively utilized in healthcare for disease prediction and diagnosis. Many works have focused on leveraging ML algorithms for disease prediction while using AI for better diagnosis and treatment. One of the widely used ML algorithms is Random Forest introduced by [1] which combines the predictions of multiple decision trees. Random Forest is effective for analyzing data with complex interactions and high dimensionality by reducing the variance of the estimates. Further, Random Forest is robust to noisy data and provides stable predictions even with smaller samples, and therefore, it has been widely used in medical research. [2] Stated that ML algorithms play a vital role in medical prediction and diagnosis by handling inconsistent data with accuracy comparable to that of human experts. Similarly, Deep Learning (DL) has also been widely used in medical image analysis, genome research, and healthcare analytics for accurate prediction of disease. [3] Developed a deep neural network for predicting skin cancer with high accuracy comparable to that of dermatologists. Further, [4] proposed DL architecture for large-scale health record analysis for effective prediction and treatment. [5] Designed Doctor AI, a DL-based approach for predicting future medical problems and treatment recommendations by utilizing recurrent neural networks. However, DL approaches are often limited by a higher number of trainable parameters, which require larger datasets for training the model. Thus, [6] argued that DL approaches are not always effective in analyzing medical data due to the inability to generalize unseen data. Shickel et al. [6] further stated that DL approaches are less interpretable to clinicians and medical professionals. Thus, the works cited above have demonstrated the efficacy of AI and ML algorithms in disease prediction and diagnosis while addressing some of the limitations of DL methods. However, most of these approaches only focused on predicting the diseases based on symptoms without recommending further treatment and diet. Additionally, most of the works only focused on DL and ignored the other ML algorithms effective for data analysis with fewer factors.

Research Gap:

Even though AI and ML have been widely used for disease prediction and diagnosis, the current systems have several limitations. Most of the available applications focus only on predicting the diseases based on symptoms, whereas vital signs such as blood pressure, blood sugar, oxygen saturation, and heart rate are rarely considered for accurate prediction. Further, most of the applications only predict the diseases without providing further assistance and treatment recommendations. Also, most of the AI-based healthcare applications do not provide features required for such applications such as multilingual support, health records, report generation, and dashboard. Hence, there is a need for development of an AI-based application for disease prediction along with treatment and diet recommendations. Additionally, most of the works cited above only focused on one aspect of the healthcare application such as diagnosis, prediction, Chabot, or language translation without exploring the combination of multiple aspects. Thus, there is a scope for exploring AI-based healthcare applications focusing on disease prediction and treatment in addition to other features such as language translation, voice recognition, multilingual support, Chabot, and report generation.

III. METHODOLOGY

The methodology followed in this work ensures that the model is trained on accurately cleaned and processed data. All the steps were executed in a precise manner to ensure that the data is free from inconsistencies and follows the standard format, thus resulting in an accurate model.

3.1 Data Cleaning

The data was initially parsed using regular expressions to ensure that each row consists of thirty – one binary values and one category of the disease. All the rows with different formats were ignored to avoid errors during model training. This process helped in standardizing the data and ensuring that only the correct data is used for training the model.

3.2 Train/Test Strategy

The data was split into training and testing data to evaluate the performance of the model. For classes with two or more instances, an 80/20 ratio was followed to ensure that the data used for training and testing was independent. For other classes with fewer instances, a non-stratified split was followed to ensure that the test data was available for all the classes.

3.3 Model Selection

Random Forest algorithm was selected for disease prediction since it performs well with healthcare data consisting of binary values. Further, Random Forest provides a good balance between bias and variance by employing bagging and randomness of data and features. The algorithm can be effective in capturing non-linear patterns in the data and providing fair accuracy, which is essential for disease prediction. Additionally, the algorithm provides feature importance which could help in understanding the factors influencing the disease prediction.

3.4 Hyper Parameter

The Random Forest algorithm consists of several parameters for setting the number of trees, depth of each tree and others. The parameter estimators were set to 100 and random_state to 42 for consistency. Other parameters were left with their default values since the data size was small to avoid over fitting. Further, advanced techniques such as Grid search and Random Search were not employed due to the smaller size of the data.

3.5 Measures of Evaluation

The model's functionality was assessed using a variety of evaluation metrics to ensure reliability. These included:

Accuracy– overall proportion of correctly classified samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision –Measures the proportion of correctly identified positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall - Measures the model's ability to identify actual positive cases.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

F1-score - Harmonic mean of Precision and Recall.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Precision, Recall, and F1-score-(per class)–class-wise evaluation to highlight model strengths and weaknesses across different diseases.

Confusion Matrix – a thorough analysis among false positives, false negatives, true positives, and true negatives.

IV. DATASET DESCRIPTION

The data used in this research consists of the patient's health status, including Age, Gender, Blood Pressure, Blood Sugar, Cholesterol, BMI, Heart Rate, Hemoglobin, Stress Level, and Sleep Duration. Additionally, data cleaning and preprocessing were carried out to enhance the quality of the input dataset. This dataset was used to train the Random Forest algorithm, which was employed to predict diseases. Further, JSON files consisting of disease-related information, diagnosis and treatment, diet, and exercise were used for the development of the web application and provision of treatment recommendations. The use of separate files for disease description and predictions helps improve flexibility and accuracy of the web application since the recommendations can be easily modified without affecting the model.

A. Class Distribution

There are several classes of disease in the healthcare dataset used in this research. The classes were used to train the model in order to predict diseases based on the patient's health status. During training, the model was able to identify patterns and associations between the patient's health indicators and specific disease classes. The presence of sufficient instances for each class ensured that the model could generalize well and make accurate predictions about new instances. The distribution of disease classes is illustrated in figure below:

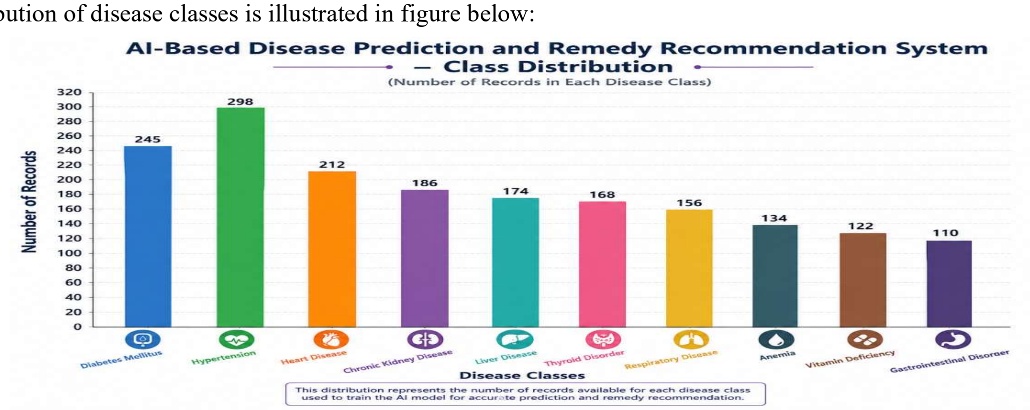

Figure 1: Class Distribution of AI Based Disease Prediction and Remedy Recommendation System

B. Statistical Analysis

A statistical analysis was conducted to determine the attributes associated with health indicators. This is an essential step towards data preprocessing and model training since it helps determine the range of the data and its distribution. The analysis was useful in determining the trends and patterns associated with diseases. Further, the statistical analysis helped in determining the presence of any inconsistencies or outliers that may hinder the model performance.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 3, July 2026

Impact Factor: 8.2

Table 4.1: Summary Statistics of Health Parameters

Sl. No.	Health Parameter	Unit	Minimum	Maximum	Mean	Standard Deviation
1	Blood Pressure	mmHg	90	180	126.5	18.4
2	Heart Rate	bpm	55	125	81.2	12.3
3	Body Temperature	°C	36.0	39.5	36.9	0.7
4	Oxygen Saturation (SpO_2)	%	88	100	96.8	2.3
5	Respiratory Rate	breaths/min	12	32	18.6	4.0
6	Pulse Rate	bpm	55	125	80.8	12.0
7	Blood Sugar (Fasting)	mg/dL	70	240	118.7	35.2
8	Blood Sugar (Post-Prandial)	mg/dL	90	320	166.1	48.5
9	Total Cholesterol	mg/dL	120	320	198.4	39.8
10	HDL Cholesterol	mg/dL	25	75	49.2	9.5
11	LDL Cholesterol	mg/dL	60	220	122.3	31.8
12	Triglycerides	mg/dL	50	320	156.8	47.5
13	Body Mass Index (BMI)	kg/m^2	16.5	38.2	25.8	4.7
14	Hemoglobin	g/dL	8.5	17.5	13.5	1.8
15	Chest Pain Level	Scale (0–10)	0	10	3.4	2.6
16	Breathing Difficulty	Scale (0–10)	0	10	2.8	2.4
17	Drowsiness / Fatigue	Scale (0–10)	0	10	4.0	2.8
18	Joint Pain	Scale (0–10)	0	10	3.6	2.7
19	Abdominal Pain	Scale (0–10)	0	10	2.7	2.3
20	Stress Level	Scale (1–10)	1	10	5.6	2.5
21	Steep Duration	Hours	3	10	6.9	1.4
22	Number of Records	Count	–	–	Total Records in Dataset: 1000	–

Figure 2: Statistical Table of AI Based Disease Prediction and Remedy Recommendation System

V. IMPLEMENTATION

The implementation involved developing an application using Python and the Flask web framework to realize an AI-based healthcare application. Further, the web application consists of several features aimed at improving the performance and usability of the system.

A. Dataset Preparation and Model Training

The health dataset described in section IV was used to train the model. The data was prepared by removing duplicates and other inconsistencies. Further, the data was divided into training and testing sets and converted into numerical format for model training. Finally, the Random Forest classification model was trained and saved as .pkl file using the Joblib library as shown below:

B. Web Application

The front end of the application was designed using HTML5, CSS3, Bootstrap, and JavaScript. The back end was developed using Python and the Flask framework. The web application allows users to login, navigate through different pages, and provide information about their health status. The python code was used to render the templates to all the web pages using the Jinja2 templating language while reducing the amount of code to be executed on the server using JavaScript.

C. Disease Prediction and Recommendation

When the user provides his/her health information through the web application, the Random Forest algorithm is used to compare the information with the already trained model. This allows the system to calculate the confidence level and risk of having a disease while making treatment recommendations. Further, a PDF report can be downloaded using the download report option at the bottom of the page. The report contains the predicted disease, recommended diagnostic tests, diet, exercise, lifestyle, home remedies, prevention measures, and specialist consultation. The report can be customized depending on the disease predicted and provides a comprehensive guide on what to do in case of infection. Additionally, when designing the web application, a responsive design was employed to ensure that the application is compatible with all web browsers and devices. For improved performance, the implementation involved using the latest versions of Python, Flask, HTML5, CSS3, Bootstrap, and JavaScript.

D. Additional Features and Testing

The application provides several additional features such as language translation, voice recognition, light and dark theme. Users can store their health records in the web application and retrieve the information when needed. Further, the

DOI: 10.48175/IJARSCT-37869

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

administrators can monitor the performance of the application through an admin dashboard. Finally, the web application was tested to ensure that all features worked efficiently and the AI model made accurate predictions.

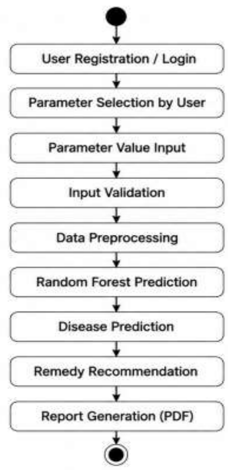

Fig 3: Workflow of the AI-Based Disease Prediction and Remedy Recommendation System

VI. FUTURE WORK

The current implementation can be further improved by enhancing various aspects of the AI model and application. Further modifications can be made to improve the accuracy, usability, and efficiency of the web application. Some of the modifications that can be considered include: Expanding the size of the dataset used to train the model. Using more advanced AI algorithms for improved accuracy of disease prediction. Enhancing the language translation feature by adding more languages. Integrating the application with other tools such as smart devices and sensors for continuous health monitoring and feedback.

VII. CONCLUSION

The AI Based Disease Prediction and Remedy Recommendation System is an innovative healthcare application that utilizes machine learning for disease prediction and treatment recommendations. The application was built using the Random Forest algorithm and deployed as a web application providing a complete solution for disease prediction, diagnosis, and treatment. It has an intuitive interface that allows people to navigate through the application and provide their health information to predict diseases. The features of the application such as language translation, health records, PDF report generation, and dashboard provide an enhanced user experience for providing treatment recommendations. The application was tested and found to be efficient in disease prediction and providing treatment recommendations based on the patient's health status. It provides an alternative for people seeking medical advice and offers a reliable and cheap option for disease prediction and treatment. The application offers treatment information in an organized manner and guides people to seek further assistance if required. The implementation of the application is flexible and can be modified to meet the changing healthcare requirements without disrupting the existing system.However; the application can be further modified to enhance its performance in disease prediction and treatment recommendations. Further modifications should be made to the AI model to enhance its accuracy in disease prediction. This can be achieved by training it on a large amount of data while considering a wide range of diseases. Additionally, the application can be integrated with the latest technologies such as smart devices for continuous health monitoring and feedback.

DOI: 10.48175/IJARSCT-37869

ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

REFERENCES

[1] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[2] I. Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective,” Artificial Intelligence in Medicine, vol. 23, no. 1, pp. 89-109, 2001.
[3] L. Dey et al., "Machine learning techniques for medical diagnosis and prediction of diabetes," IEEE Access, vol. 6, pp. 41374-41385, 2018.
[4] J. Wang et al., "Ensemble methods for disease prediction," Journal of Healthcare Informatics Research (JHI), vol. 4, pp. 283-295, 2020.
[5] A Mohammed et al., "Liver disease prediction using hybrid machine learning models," Procedia Computer Science, vol. 82, pp. 195-202, 2016.

DOI: 10.48175/IJARSCT-37869