ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

MACHINE LEARNING BASED DISEASE PREDICTION SYSTEM USING PYTHON

Sunil.B[1], M. Nagaraju[2]
[1]*Master of Technology in Computer Science, Gonna Institute of Information Technology & Sciences*
[2]*Associate Professor, Gonna Institute of Information Technology & Sciences,*
(Approved by AICTE, New Delhi Affiliated to JNTU-Gurajada Vizianagaram)

ABSTRACT
Early disease detection plays an important role in improving healthcare outcomes and reducing health risks. This paper presents a Machine Learning Based Disease Prediction System with Health Risk Assessment that helps users obtain preliminary health insights based on symptoms and key health parameters.The system consists of two modules: disease prediction and health risk assessment. The disease prediction module identifies the most probable disease from user-provided symptoms, while the health risk assessment module evaluates factors such as age, blood pressure, blood sugar, BMI, cholesterol level, and heart rate to determine potential health risks.To develop the system, healthcare datasets were preprocessed and analyzed using machine learning algorithms including Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine (SVM). Among these, Random Forest achieved the highest accuracy of 94%. The application was implemented using Python, Flask, Pandas, and Scikit-Learn with a user-friendly web interface.The proposed system provides quick and accessible health insights and can serve as a useful support tool for early disease prediction and health monitoring. While it is not a substitute for professional medical advice, it can assist users in making informed healthcare decisions.

KEYWORDS: *Machine Learning, Disease Prediction, Health Risk Assessment, Random Forest, Healthcare Analytics, Predictive Healthcare.*

1. INTRODUCTION

Healthcare plays a crucial role in improving the quality of life by enabling timely diagnosis and treatment of diseases. However, increasing patient populations, limited healthcare accessibility, and delays in medical consultation often make early disease detection challenging. As a result, there is a growing need for intelligent systems that can assist individuals in obtaining preliminary health assessments quickly and efficiently.

Recent advancements in Machine Learning (ML) have transformed the healthcare sector by enabling data-driven decision-making and predictive analysis. ML algorithms can analyze large volumes of healthcare data, identify hidden patterns, and generate accurate predictions. These capabilities have made machine learning a valuable tool for disease prediction, health monitoring, and medical decision support.

Disease prediction systems help identify potential illnesses based on symptoms and medical information, allowing users to seek timely medical attention. Similarly, health risk assessment systems evaluate important health indicators such as age, blood pressure, blood sugar level, cholesterol, heart rate, and Body Mass Index (BMI) to estimate the likelihood of developing health-related complications. Such systems promote preventive healthcare and improve health awareness among individuals.

This project presents a Machine Learning Based Disease Prediction System with Health Risk Assessment that combines symptom-based disease prediction and health parameter analysis within a web-based platform. The system utilizes machine learning algorithms including Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine (SVM) to analyze healthcare data and generate predictions. Among the evaluated models, Random Forest achieved the highest prediction accuracy of 94%.

The main objective of the proposed system is to provide users with a simple, accessible, and reliable platform for obtaining preliminary health insights. The application is developed using Python and Flask, enabling users to predict diseases based on symptoms and assess health risks using key medical parameters. Although the system is not intended to replace professional medical diagnosis, it serves as an effective decision-support tool for early disease detection and health monitoring.

2. LITERATURE REVIEW

Machine Learning has become an important technology in healthcare due to its ability to analyze large volumes of medical data and support disease diagnosis. Several studies have demonstrated the effectiveness of machine learning algorithms in predicting diseases based on symptoms, medical history, and health parameters. Algorithms such as Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine (SVM) have been widely used for disease classification and prediction because of their ability to identify hidden patterns within healthcare datasets.

Existing disease prediction systems have shown promising results in detecting conditions such as diabetes, heart disease, kidney disorders, and other chronic illnesses. These systems assist healthcare professionals by providing preliminary predictions and improving diagnostic accuracy. In addition, health risk assessment models have been developed to evaluate important factors such as blood pressure, blood sugar levels, cholesterol, body mass index (BMI), and heart rate, enabling early identification of potential health risks.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Despite significant progress, many existing systems focus only on a single disease or rely solely on symptom-based analysis. Some applications lack comprehensive health risk assessment features, while others provide limited user interaction and accessibility. Furthermore, disease prediction and health risk assessment are often implemented as separate systems, reducing their overall effectiveness in providing complete healthcare support.

To address these limitations, the proposed Machine Learning Based Disease Prediction System with Health Risk Assessment integrates both disease prediction and health parameter evaluation within a single web-based platform. The system utilizes machine learning algorithms to analyze symptoms and health metrics, providing users with preliminary health insights and supporting preventive healthcare practices.

The literature review highlights the growing role of machine learning in healthcare and establishes the need for an integrated system that combines disease prediction and health risk assessment to improve healthcare accessibility and early disease detection.

3. PROPOSED SYSTEM

The proposed Machine Learning Based Disease Prediction System with Health Risk Assessment is designed to provide preliminary healthcare insights through disease prediction and health risk evaluation. The system integrates machine learning techniques with a web-based application to assist users in identifying potential diseases and assessing health risks based on symptoms and medical parameters.

The system consists of two main modules: the Disease Prediction Module and the Health Risk Assessment Module. In the Disease Prediction Module, users select symptoms through a web interface, and the trained machine learning model predicts the most probable disease based on the provided inputs. This helps users obtain early health insights and encourages timely medical consultation.

The Health Risk Assessment Module evaluates key health parameters such as age, blood pressure, blood sugar level, cholesterol level, heart rate, and Body Mass Index (BMI). Based on these parameters, the system determines the user's health risk level and provides a preliminary assessment of potential health concerns.

The application is developed using Python and Flask, while HTML and CSS are used to create an interactive user interface. Machine learning algorithms including Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine (SVM) are trained using healthcare datasets and integrated into the web application for real-time prediction.

The workflow of the proposed system begins with user input, followed by data processing and prediction through trained machine learning models. The generated results are then displayed through the web interface, providing users with quick and accessible healthcare insights.

The proposed system offers several benefits, including early disease detection, health risk monitoring, improved healthcare awareness, and easy accessibility through a web platform. By combining disease prediction and health risk assessment within a single application, the system provides a more comprehensive healthcare support solution compared to existing approaches.

4. METHODOLOGY

4.1 Dataset Description

The proposed system utilizes two healthcare datasets. The first dataset is a symptom-based dataset containing common symptoms such as fever, cough, headache, fatigue, chest pain, nausea, dizziness, and weakness associated with various diseases including Flu, COVID-19, Dengue, Malaria, Diabetes, Hypertension, and Heart Disease. Symptoms are represented in binary format, where 1 indicates the presence of a symptom and 0 indicates its absence.

The second dataset is a Health Risk Assessment Dataset containing health parameters such as age, blood pressure, blood sugar level, cholesterol level, heart rate, and Body Mass Index (BMI). These parameters are used to assess health risks related to conditions such as diabetes, hypertension, obesity, and cardiovascular diseases.

4.2 Data Preprocessing

Data preprocessing was performed to improve data quality and model performance. The datasets were cleaned by removing duplicate records and handling missing values. Symptom data was transformed into binary numerical values, enabling machine learning algorithms to process the information effectively.

Feature selection was applied to identify the most relevant attributes for prediction. Important features included symptoms such as fever, cough, fatigue, nausea, and dizziness, along with health parameters including blood pressure, blood sugar level, cholesterol level, heart rate, age, and BMI. The datasets were further normalized to ensure consistency among numerical values.

For model development, the datasets were divided into training and testing sets using an 80:20 ratio.

4.3 Machine Learning Algorithms

Several machine learning algorithms were evaluated for disease prediction and health risk assessment.

Decision Tree: A classification algorithm that predicts outcomes using a tree-like structure based on feature values.

Random Forest: An ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Naïve Bayes: A probabilistic classifier based on Bayes' theorem, known for its simplicity and efficiency in classification tasks.

Support Vector Machine (SVM): A supervised learning algorithm that identifies an optimal decision boundary for classification problems.

4.4 Model Evaluation

The performance of the machine learning models was evaluated using standard classification metrics including Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. The results obtained are shown in Table 1.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Algorithm	Accuracy
Decision Tree	88%
Random Forest	94%
Naïve Bayes	86%
SVM	92%

Among all evaluated algorithms, Random Forest achieved the highest accuracy of 94% and was selected as the final model for deployment. The evaluation results demonstrate the effectiveness of machine learning techniques in disease prediction and health risk assessment.

5. IMPLEMENTATION

5.1 Development Environment

The ML-Based Disease Prediction System was developed using Python and the Flask framework. The application was designed using HTML and CSS for the user interface. Machine learning models were developed using Scikit-Learn, while Pandas and NumPy were used for data preprocessing and analysis. The development was carried out using Visual Studio Code on a Windows operating system.

5.2 Model Development

The prepared datasets were preprocessed and divided into training and testing sets using an 80:20 ratio. Multiple machine learning algorithms including Decision Tree, Naïve Bayes, Support Vector Machine, and Random Forest were evaluated. Among these, the Random Forest classifier achieved the highest accuracy of 94% and was selected for deployment.

5.3 Web Application Implementation

The trained machine learning models were integrated with a Flask-based web application. The application consists of a Login Page, Dashboard, Symptom-Based Disease Prediction Module, Health Risk Assessment Module, and Result Pages. User inputs are collected through web forms and processed by the trained models to generate prediction results in real time.

5.4 System Workflow

The system follows a simple workflow:

1. User logs into the application.
2. User selects symptoms or enters health parameters.
3. Flask processes the input data.
4. The trained machine learning model generates predictions.
5. Results are displayed to the user through the web interface.

5.5 Implementation Outcome

The developed system successfully predicts diseases based on symptoms and assesses health risks using medical parameters. The integration of machine learning models with Flask enables quick and user-friendly access to prediction results, demonstrating the practical application of machine learning in healthcare.

6. RESULTS AND DISCUSSION

Figure 1. Login Page

The login page provides secure access to the system.

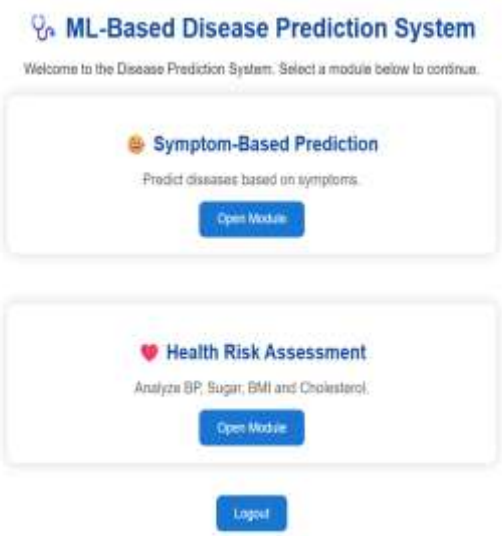

Figure 2. Dashboard Page

The dashboard allows navigation between disease prediction and health risk assessment modules.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

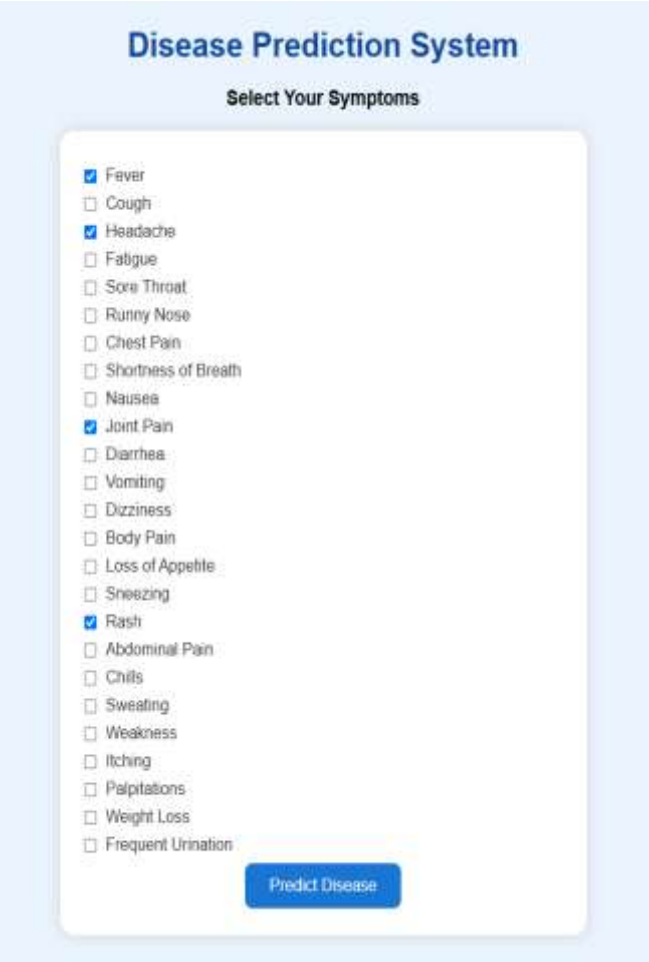

Figure 3. Symptom Based Disease Prediction Module

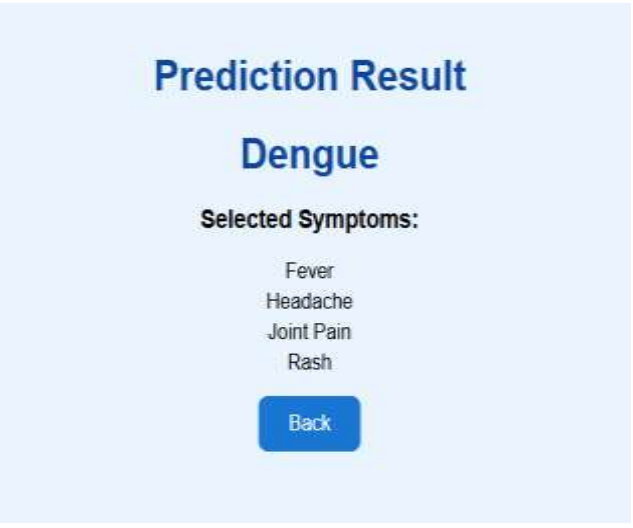

Figure 4. Disease Prediction Result

The system successfully predicts the most probable disease based on the selected symptoms.

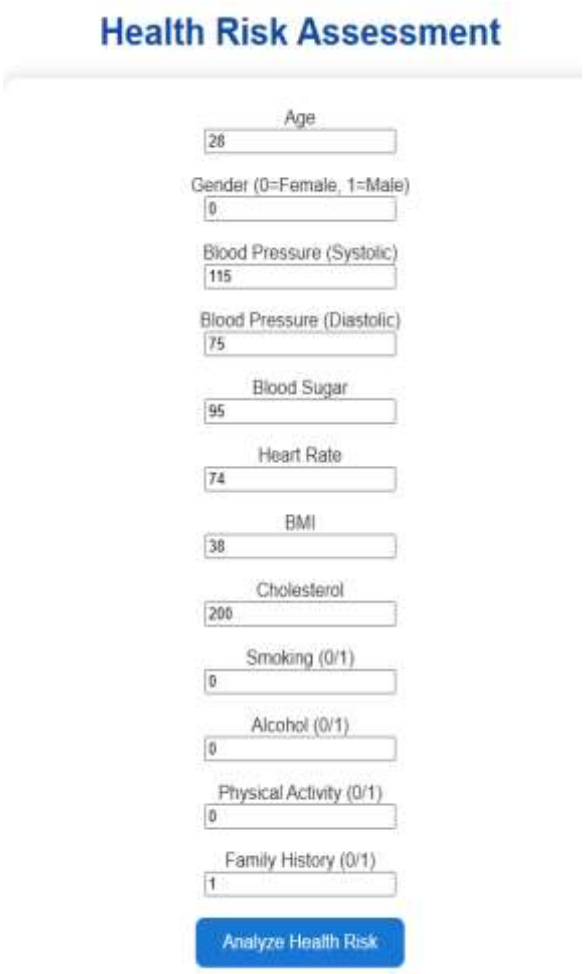

Figure 5. Health Risk Assessment Module

Figure 6. Health Risk Assessment Result

The system analyzes health parameters and predicts the corresponding health risk category.

The experimental results demonstrate that the developed system successfully performs disease prediction and health risk assessment through a user-friendly web interface. The integration of machine learning models with Flask enables real-time prediction and efficient healthcare support.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

7. TESTING

7.1 Testing Overview

Testing was conducted to verify the correctness, reliability, and performance of the ML-Based Disease Prediction System with Health Risk Assessment. Different testing techniques were applied to ensure that all modules functioned according to the specified requirements. The testing process helped identify and eliminate errors while improving system stability and prediction accuracy.

7.2 Unit Testing

Unit testing was performed on individual modules such as Login Module, Disease Prediction Module, Health Risk Assessment Module, and Logout Module. Each module was tested independently to verify its functionality.

Test Case	Expected Result	Status
User Login	User successfully logged in	Passed
Disease Prediction	Disease predicted correctly	Passed
Health Risk Assessment	Risk category displayed correctly	Passed
Logout	User logged out successfully	Passed

7.3 Integration Testing

Integration testing was conducted to ensure smooth communication between different modules of the application.

Modules Integrated	Expected Result	Status
Login + Dashboard	Dashboard opens after login	Passed
Dashboard + Disease Prediction	Prediction generated successfully	Passed
Dashboard + Health Risk Assessment	Risk assessment displayed correctly	Passed

7.4 System Testing

System testing was performed on the complete application to verify overall functionality.

Test Scenario	Expected Result	Status
Complete Application Execution	All modules work correctly	Passed
Navigation Between Pages	Smooth navigation	Passed
Prediction Generation	Accurate output displayed	Passed

7.5 Functional Testing

Functional testing ensured that every feature operated according to the requirements.

Function	Expected Result	Status
User Authentication	Valid login accepted	Passed
Disease Prediction	Correct disease predicted	Passed
Health Risk Analysis	Correct risk assessment generated	Passed

7.6 Performance Testing

Performance testing evaluated the response time and efficiency of the application.

Parameter	Result
Login Response Time	Less than 2 seconds
Disease Prediction Time	Less than 3 seconds
Health Risk Assessment Time	Less than 3 seconds
Overall Performance	Satisfactory

7.7 User Acceptance Testing

User Acceptance Testing (UAT) was conducted to ensure that the system met user expectations. Users were able to navigate the application easily and obtain prediction results without difficulty. Feedback indicated that the system was user-friendly and easy to understand.

7.8 Testing Results

The testing process confirmed that all modules functioned correctly and produced the expected outputs. The application successfully generated disease predictions and health risk assessments with satisfactory performance and reliability. No major functional issues were observed during testing.

Table 7.1 Overall Testing Summary

Testing Type	Result
Unit Testing	Passed
Integration Testing	Passed
System Testing	Passed
Functional Testing	Passed
Performance Testing	Passed
User Acceptance Testing	Passed

8. CONCLUSION AND FUTURE SCOPE

8.1 Conclusion

The ML-Based Disease Prediction System with Health Risk Assessment was successfully developed and implemented using Machine Learning and Flask technologies. The system provides disease prediction based on user-selected symptoms and evaluates health risks using important health parameters such as age, blood pressure, blood sugar level, BMI, cholesterol level, and heart rate.

Various machine learning algorithms including Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine were analyzed and evaluated. Among them, the Random Forest algorithm achieved the highest prediction accuracy of 94% and was selected as the final prediction model. The developed web application provides a user-friendly interface that allows users to obtain prediction results quickly and efficiently.

The testing results confirmed that all modules function correctly and generate reliable outputs. The system successfully

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

demonstrates the practical application of machine learning techniques in healthcare and can serve as a supportive tool for preliminary disease prediction and health risk assessment. Overall, the project achieved its objectives and provides an effective solution for improving healthcare awareness and early risk identification.

8.2 Future Scope

Although the developed system performs effectively, several enhancements can be implemented in the future to improve its functionality and accuracy.

Future improvements may include:

- Integration of larger and real-world hospital datasets for improved prediction accuracy.
- Inclusion of additional diseases, symptoms, and health parameters.
- Development of a doctor recommendation system based on prediction results.
- Integration with Electronic Health Records (EHR) and wearable health monitoring devices.
- Deployment of the application on cloud platforms for wider accessibility.
- Development of Android and iOS mobile applications.
- Implementation of advanced machine learning and deep learning techniques.
- Addition of multilingual support for users from different regions.
- Enhancement of security features for protecting healthcare data.
- Implementation of database systems such as MySQL or MongoDB for efficient storage and management of healthcare data.

These enhancements can improve the accuracy, scalability, and practical usability of the system, making it more suitable for real-world healthcare applications.

REFERENCES

1. *Python Software Foundation, "Python Documentation." Available: https://docs.python.org/*
2. *Flask Documentation, "Flask Web Framework Documentation." Available: https://flask.palletsprojects.com/*
3. *Pandas Documentation, "Pandas User Guide and API Reference." Available: https://pandas.pydata.org/*
4. *Scikit-Learn Documentation, "Machine Learning in Python." Available: https://scikit-learn.org/*
5. *F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.*
6. *A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed., O'Reilly Media, 2022.*
7. *I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.*
8. *World Health Organization (WHO), "Official Health Information Resources." Available: https://www.who.int/*
9. *Kaggle, "Healthcare and Disease Prediction Datasets." Available: https://www.kaggle.com/*
10. *Visual Studio Code Documentation. Available: https://code.visualstudio.com/docs*
11. *MDN Web Docs, "HTML Documentation." Available: https://developer.mozilla.org/*
12. *MDN Web Docs, "CSS Documentation." Available: https://developer.mozilla.org/*
13. *L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.*
14. *NumPy Documentation. Available: https://numpy.org/doc/*
15. *Machine Learning Mastery, "Machine Learning Tutorials and Resources." Available: https://machinelearningmastery.com/*