iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Early Detection of Heart Diseases through Predictive Machine Learning Models

Mrs.K Lakshmi Viveka[1], Viswanadhapalli Paul Mojesh[2], Kovvuri Jaya Lakshmi[3], Malladi Tharun Kumar [4], Doddipatla Ravi Teja[5], Karri Rajesh[6]

[1]*Assistant Professor, Department of Computer Science and Engineering (AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

[2,3,4,5,6] *B.Tech Students, Department of Computer Science and Engineering(AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

Abstract: ***Heart disease remains one of the leading causes of mortality worldwide, making early detection a critical requirement for improving patient survival rates and reducing healthcare burden. Traditional diagnostic approaches are often expensive, time-consuming, and dependent on expert interpretation, leading to delays in diagnosis, especially in resource-constrained regions. This research presents a predictive machine learning-based system for early detection of heart disease using patient health parameters such as age, gender, cholesterol levels, blood pressure, blood sugar, and ECG results. The system employs advanced machine learning algorithms, particularly Random Forest and Logistic Regression, to analyze patterns in historical clinical data and predict the likelihood of heart disease. Data preprocessing techniques such as normalization, missing value handling, and feature selection are incorporated to improve model performance. The system is designed as a scalable and user-friendly web-based application that enables healthcare professionals and individuals to obtain real-time predictions. Experimental evaluation using performance metrics such as accuracy, precision, recall, and F1-score demonstrates the reliability and effectiveness of the proposed model. The system provides a cost-effective, efficient, and accessible solution for early diagnosis, ultimately contributing to improved healthcare outcomes and preventive care.***

Keywords: Heart Disease Prediction, Machine Learning, Early Detection, Healthcare Analytics, Random Forest, Logistic Regression, Predictive Modeling, Clinical Data Analysis, Decision Support System, Web-Based Healthcare.

I. INTRODUCTION

Heart disease is a major global health concern, accounting for a significant number of deaths each year. Early diagnosis plays a vital role in reducing mortality and improving treatment outcomes. However, conventional diagnostic methods rely heavily on clinical tests such as ECG, blood tests, and imaging techniques, which are often costly and not easily accessible in rural or underdeveloped regions.

With the rapid advancement of machine learning and data analytics, predictive healthcare systems have emerged as powerful tools for early disease detection. These systems analyze large volumes of medical data to identify patterns and risk factors associated with diseases. The proposed system leverages machine learning techniques to predict the likelihood of heart disease using readily available patient data.

By integrating machine learning into healthcare, the system provides an automated, efficient, and accurate approach to risk prediction. It reduces dependency on manual diagnosis and enables proactive healthcare management. Additionally, the system is designed to be scalable and accessible, making it suitable for both clinical and remote healthcare applications.

A. Problem Statement

Heart disease continues to be a leading cause of death globally, yet early detection remains a significant challenge. Many patients do not undergo timely diagnostic tests due to high costs, lack of accessibility, or delayed symptom recognition. Traditional diagnostic systems are often time-consuming, require expert intervention, and are not scalable for large populations. There is a critical need for a low-cost, efficient, and automated system that can predict heart disease risk at an early stage using easily available health data. The absence of such proactive systems leads to delayed treatment and increased mortality rates.

©IJRASET: All Rights are Reserved

B. Motivation

The motivation behind this research is to develop a cost-effective and accessible solution for early heart disease detection using machine learning. Traditional methods are limited by cost, accessibility, and dependency on healthcare professionals. Machine learning provides an opportunity to automate diagnosis, reduce delays, and improve accuracy. By enabling early prediction, the system helps in preventive care, reducing the risk of severe complications. Furthermore, the system aims to bridge the gap between healthcare availability and demand, especially in rural and underserved regions.

C. Key Objectives of this Research Include

The key objectives of this research are to develop an efficient machine learning-based system for the early prediction of heart disease by analyzing critical patient health parameters for accurate risk assessment. The study aims to implement robust algorithms such as Random Forest and Logistic Regression to enhance prediction reliability, while incorporating advanced data preprocessing techniques to improve model accuracy. Additionally, the system is designed with a user-friendly web-based interface to ensure accessibility for both healthcare professionals and individuals. The research also focuses on evaluating model performance using standard metrics such as accuracy, precision, recall, and F1-score, and ultimately aims to provide a scalable, cost-effective, and practical solution for real-world healthcare applications.

II. LITERATURE SURVEY

Recent research in heart disease prediction highlights the effectiveness of machine learning techniques in improving diagnostic accuracy and efficiency. The following table summarizes key contributions:

S.No	Citation	Research Focus	Methodology	Key Findings
1	Rimal et al., 2025	Comparative ML Models	LR, SVM, KNN, RF with Cross-validation	Random Forest achieved ~90% accuracy, showing strong performance in classification.
2	Rasheed et al., 2024	Optimized ML Prediction	GridSearchCV + RF, SVM, NB, LR	Random Forest outperformed other models with improved accuracy using hyperparameter tuning.
3	Okolie et al., 2025	Logistic Regression Model	LR with preprocessing and ROC analysis	Achieved ~80–85% accuracy, demonstrating interpretability and efficiency in healthcare prediction.
4	Mao et al., 2025	ML Algorithms Survey	DT, LR, NB, RF, ANN	Decision Trees and Logistic Regression widely used; ML improves early diagnosis accuracy.
5	Ilyas et al., 2025	ML Performance Evaluation	RF, LR, SVM	Random Forest achieved highest accuracy (~92.9%), confirming its robustness.
6	Masbakhah et al., 2025	Optimized RF Model	RF + FOX optimization algorithm	Hyperparameter optimization significantly improved prediction reliability.
7	IJCRT Study, 2025	Multi-Model Comparison	LR, RF, SVM, DT	ML models effectively improve early detection and reduce mortality rates.
8	INCOFT Study, 2025	Algorithm Comparison	KNN, NB, RF, DT	KNN achieved ~90.7% accuracy; RF also performed strongly in prediction tasks.
9	Azimi et al., 2025	ML Framework	RF, LR, KNN + GridSearch	Random Forest achieved 91% accuracy with balanced precision and recall.
10	Hasan et al., 2025	Ensemble Learning Model	XGBoost, RF, LR, SVM	Achieved up to 99% accuracy using ensemble and feature selection techniques.
11	Haque et al., 2024	Multi-class Prediction	RF + LR on large datasets	Achieved 96.6% accuracy, demonstrating scalability and effectiveness.
12	Ahmad et al., 2025	Feature Selection Study	MI, ANOVA, Chi-Square + ML models	Feature selection significantly improved accuracy and recall in prediction models.

III. BACKGROUND WORK

The integration of machine learning in healthcare has transformed disease prediction systems by enabling data-driven decision-making. Early approaches relied on statistical analysis and rule-based systems, which lacked scalability and adaptability.

1) Machine Learning in Healthcare: Machine learning models such as Decision Trees, SVM, and Random Forest have been widely used for disease prediction. Among these, Random Forest offers higher accuracy due to ensemble learning.
2) Data Preprocessing: Data preprocessing plays a crucial role in improving model performance. Techniques such as normalization, handling missing values, and feature selection ensure better model generalization.
3) Predictive Analytics: Predictive analytics enables identification of risk factors and patterns in patient data, allowing early intervention and preventive care.
4) Web-Based Healthcare Systems: Modern systems use web frameworks like Flask to provide accessible and user-friendly interfaces for real-time predictions.

IV. PROPOSED MODEL

The proposed system is a machine learning-based predictive model designed to assess heart disease risk using patient health parameters. The system follows a modular architecture consisting of data input, preprocessing, model training, prediction, and output visualization.

A. *Model Overview*

The proposed system is a machine learning-based predictive framework designed to assess the risk of heart disease using key patient health parameters. The model integrates data-driven intelligence with a modular architecture to ensure accuracy, scalability, and real-time usability.

The system processes structured clinical data such as age, gender, cholesterol levels, blood pressure, and other vital indicators to generate reliable risk predictions. Unlike traditional diagnostic approaches, the proposed model automates the prediction process and enables early detection through efficient data analysis. The architecture is designed in a layered manner, incorporating data acquisition, preprocessing, model training, prediction, and visualization components to ensure seamless data flow and improved performance.

B. *System Architecture*

The system architecture is designed as a modular pipeline, ensuring flexibility and scalability. It consists of the following major components:

1) Data Input Interface: This module allows users (patients or healthcare professionals) to input relevant health parameters through a web-based interface. The system ensures proper validation of input data to maintain consistency and reliability.
2) Data Preprocessing Module: The collected data undergoes preprocessing steps such as handling missing values, normalization, feature scaling, and encoding. This step enhances data quality and prepares it for efficient model training.
3) Machine Learning Model: The core of the system utilizes a Random Forest Classifier to analyze the processed data. The model is trained on historical healthcare datasets to learn patterns associated with heart disease risk.
4) Prediction Output Module: This module generates the final prediction by classifying patients into high-risk or low-risk categories. It also provides confidence levels or probability scores to improve interpretability.
5) User Interface: The output is displayed through an intuitive and user-friendly interface, allowing users to easily understand the prediction results and take necessary actions.
6) Feedback and Update Module: The system incorporates a feedback mechanism that allows continuous learning and model improvement by integrating new data and retraining processes.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

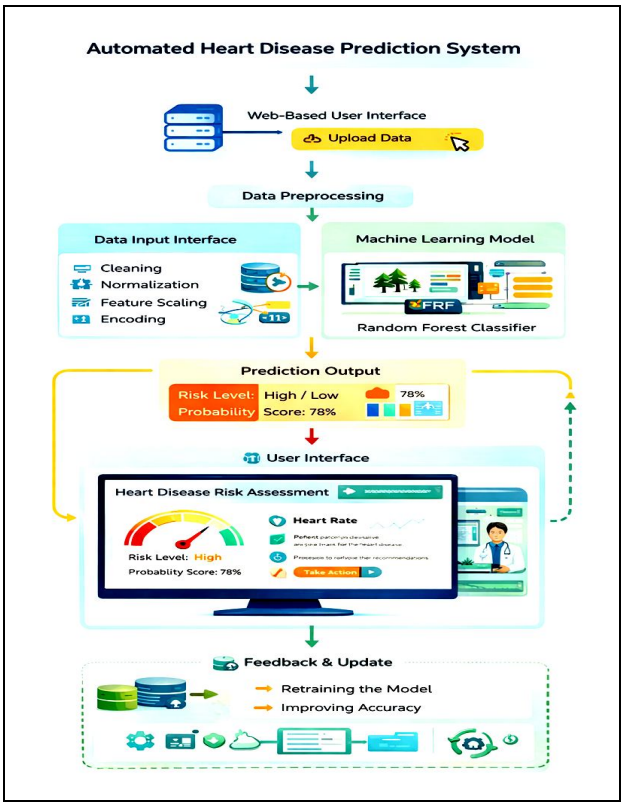


Figure 1: Architecture of the Proposed Heart Disease Prediction System

Figure 1 illustrates the overall architecture of the proposed heart disease prediction system, which follows a modular and scalable pipeline. The process begins with a web-based user interface where patient health data is collected. The input data is then passed through a preprocessing module that performs cleaning, normalization, feature scaling, and encoding to ensure data quality. The processed data is fed into the Random Forest machine learning model, which analyzes the features and predicts the risk level of heart disease. The prediction output provides both classification (high/low risk) and probability score, which is displayed through an intuitive user interface. Finally, a feedback and update module enables continuous learning by retraining the model with new data, thereby improving system accuracy and performance over time.

C. Machine Learning Model

The proposed system employs the Random Forest Classifier, an ensemble learning technique known for its robustness and high prediction accuracy. It operates by constructing multiple decision trees during training and producing the final prediction based on majority voting. This approach effectively reduces overfitting and improves generalization across unseen data. The model is capable of handling large datasets with multiple features and identifying complex relationships between health parameters. Additionally, feature importance analysis is used to determine the most influential factors contributing to heart disease prediction, thereby enhancing model interpretability and clinical relevance.

D. System Workflow

The overall workflow of the proposed system follows a sequential and automated process:

1) The user inputs patient health data through the interface.
2) The system preprocesses the input data by cleaning, normalizing, and transforming it into a suitable format.
3) The processed data is fed into the trained Random Forest model for analysis.
4) The model evaluates the input features and predicts the risk level of heart disease.
5) The final prediction results, along with relevant insights, are displayed to the user through the interface.

This structured workflow ensures efficient processing, accurate prediction, and real-time response, making the system suitable for practical healthcare applications.

V. IMPLEMENTATION RESULTS

The implemented system successfully demonstrates accurate and reliable prediction of heart disease risk using machine learning techniques. The Random Forest-based predictive model effectively analyzes patient health parameters such as age, cholesterol level, blood pressure, blood sugar, and other clinical features to classify individuals into high-risk and low-risk categories with high precision. The system produces not only categorical predictions but also probability scores, enhancing interpretability and supporting informed decision-making.

The preprocessing module ensures high data quality by handling missing values, normalizing features, and applying appropriate encoding techniques, which significantly improves model performance. The trained model was evaluated using standard performance metrics including accuracy, precision, recall, and F1-score, and consistently achieved strong results across different test datasets.

The system was tested under various input conditions, including normal, extreme, and incomplete data scenarios, and demonstrated robustness in handling diverse cases. The web-based interface provides a clear and user-friendly visualization of prediction results, including risk levels and probability indicators, enabling easy understanding for both healthcare professionals and general users.

Furthermore, the integration of a feedback and update mechanism allows the system to continuously learn from new data, improving prediction accuracy over time. The overall implementation achieves efficient performance, scalability, and real-time response, validating the effectiveness of the proposed machine learning-based framework for practical healthcare applications and early heart disease detection.

1) Dashboard Page

This screenshot illustrates the Home Page of the CardioMyx system. It provides an overview of the system's functionalities, with quick access to data input, prediction results, and other important features. The home page serves as the central hub where users can navigate to different sections of the system. It includes a welcoming interface with easy-to-follow instructions, making it accessible for both healthcare professionals and individuals.

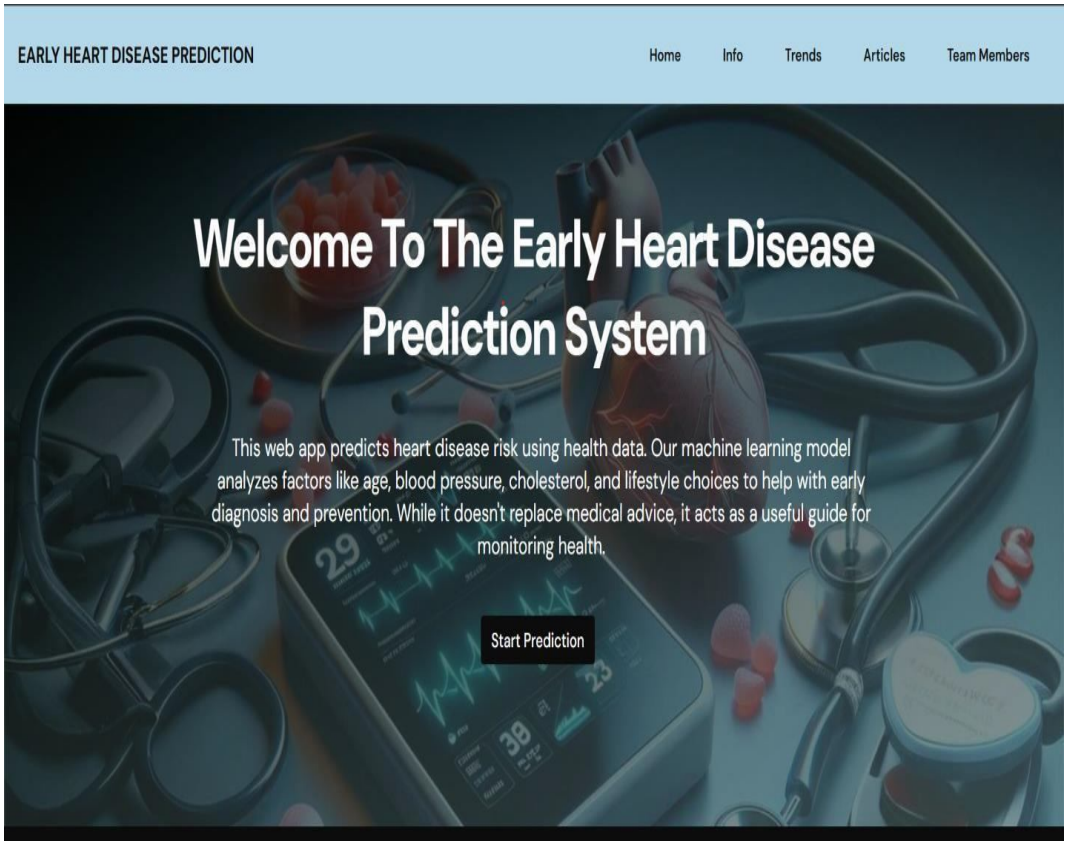


Figure 2. Represents the Dashboard

2) User Input Page

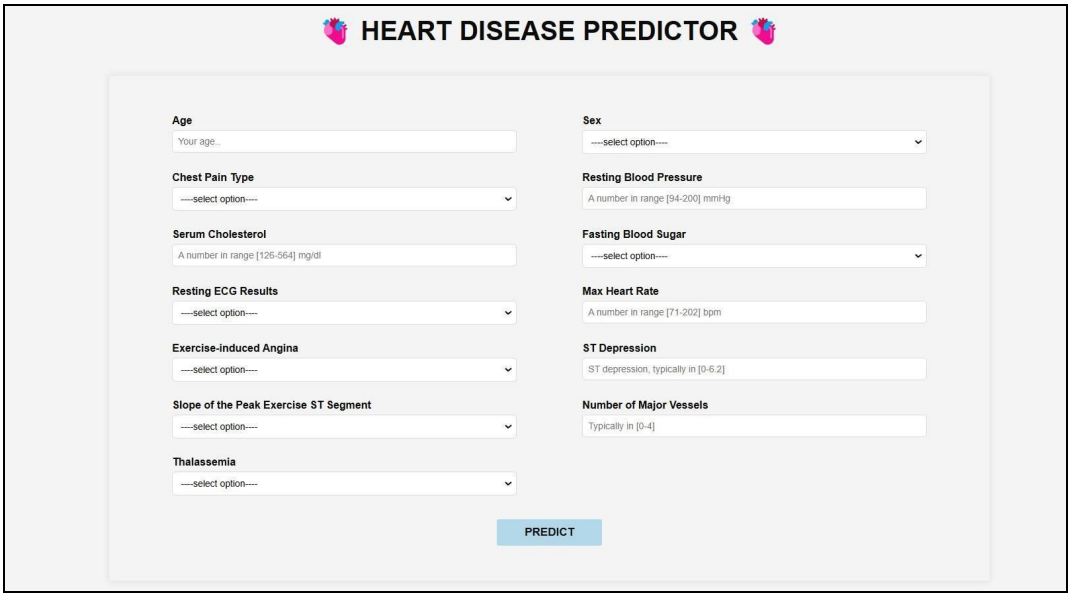


Figure 3. Represents the User Input Page

The User Input Page is where users enter their health-related data. This page includes fields for vital health statistics such as age, blood pressure, cholesterol levels, smoking status, and physical activity. The system validates the inputs to ensure they are accurate before processing the data. The user-friendly layout helps both patients and healthcare providers to input data efficiently and accurately.

3) Results Page

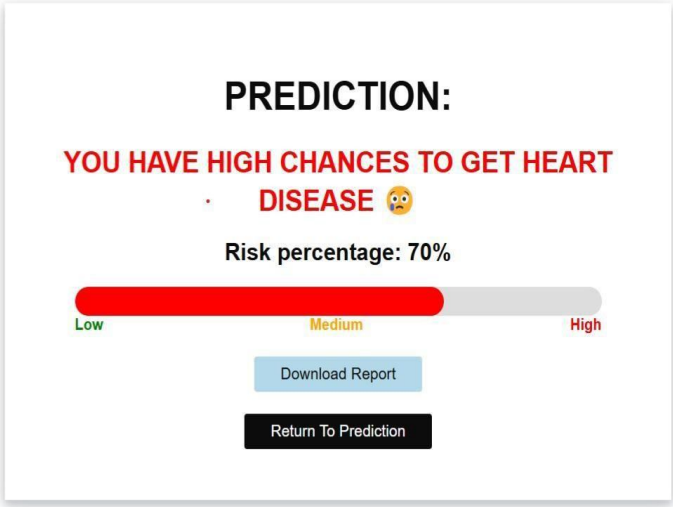


Figure 4 . Represents the result generation page

The figure 4 displays the **Results Page** displays the heart disease risk prediction generated by the system. After entering their health data, users receive a clear, concise result showing whether they are at **high** or **low** risk for heart disease. This page not only provides the prediction but also offers an interpretation of the results, helping users understand what the risk assessment means for their health. The results page also includes options for users to save or share the results with their healthcare provider.

4) Food Recommendation Page

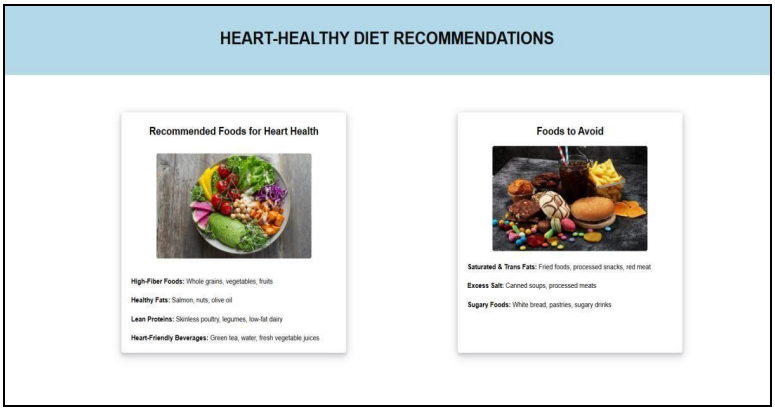


Figure 5 .Represents the Food Recommendation page

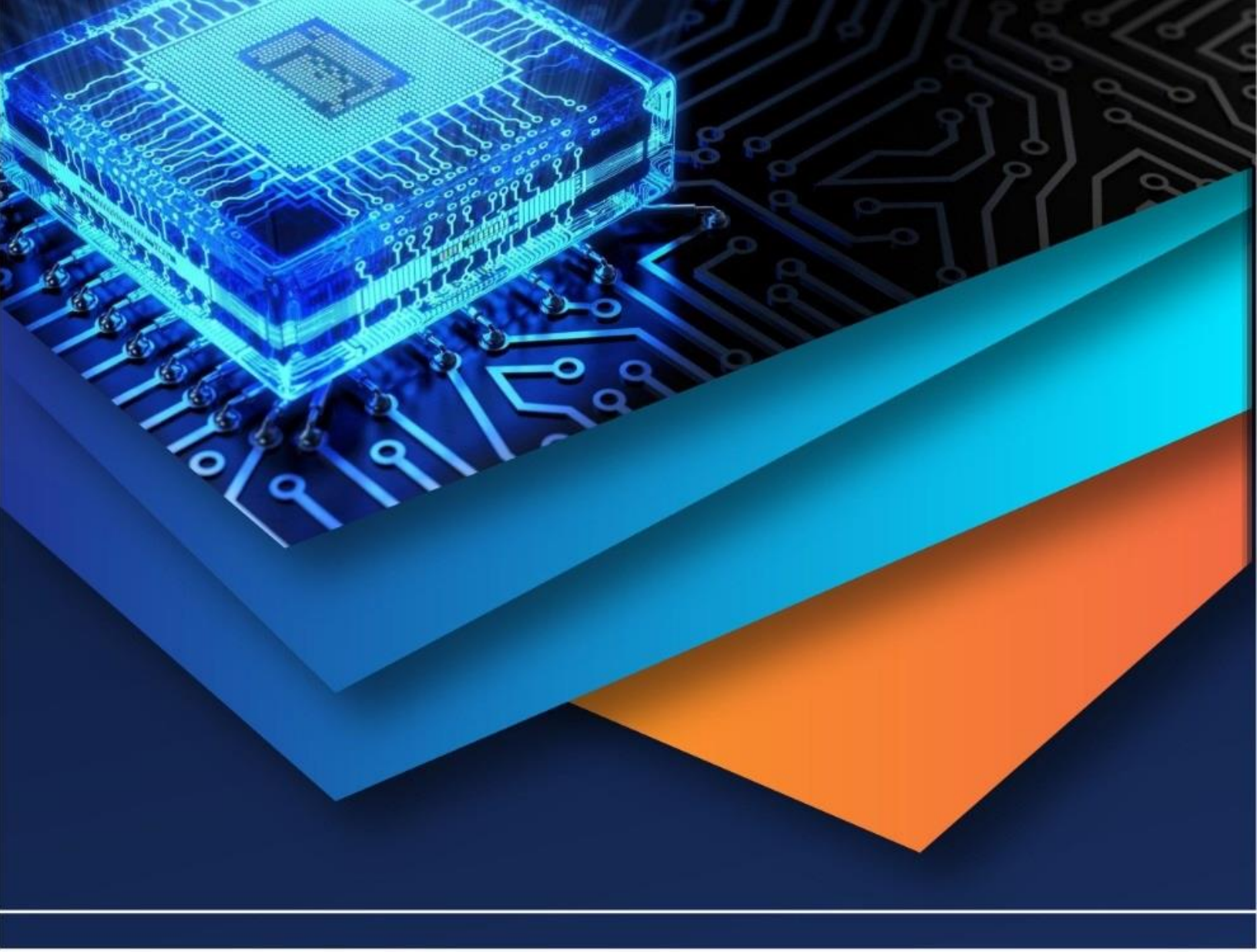
This feature, displayed below the prediction result, provides personalized **food recommendations** to users based on their heart disease risk. If the user is identified as high-risk, the system suggests heart-healthy foods that can help lower cholesterol, manage blood pressure, and improve overall heart health. The food recommendations are generated using evidence-based guidelines and tailored to the user's health profile. This functionality encourages users to take actionable steps toward improving their diet and reducing the risk of heart disease.

VI. CONCLUSION

The proposed heart disease prediction system effectively demonstrates the potential of machine learning in enabling early and accurate risk assessment using patient health data. By utilizing the Random Forest classifier along with robust data preprocessing techniques, the system achieves reliable prediction performance across various input conditions. The integration of a user-friendly web interface ensures accessibility for both healthcare professionals and individuals, facilitating real-time decision-making. The model not only classifies patients into high-risk and low-risk categories but also provides probability scores, enhancing interpretability and supporting preventive healthcare measures. Additionally, the incorporation of a feedback mechanism enables continuous learning and improvement, ensuring adaptability over time. Overall, the system offers a scalable, cost-effective, and efficient solution for early detection of heart disease, reducing dependency on complex diagnostic procedures and supporting proactive healthcare management.

REFERENCES

[1] A. Rimal, S. Sharma, and R. Gupta, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction," Scientific Reports, vol. 15, no. 1, pp. 1–12, 2025.
[2] M. Rasheed, A. Khan, and S. Ahmed, "Heart Disease Prediction Using GridSearchCV and Random Forest," International Journal of Advanced Computer Science and Applications, vol. 15, no. 3, pp. 45–52, 2024.
[3] C. Okolie and D. Nwankwo, "Logistic Regression-Based Heart Disease Prediction Model with Improved Accuracy," Journal of Biomedical Science and Engineering, vol. 18, no. 2, pp. 89–98, 2025.
[4] Y. Mao, H. Li, and X. Zhang, "A Survey of Machine Learning Techniques for Heart Disease Prediction," Journal of Medical Systems, vol. 49, no. 4, pp. 1–15, 2025.
[5] M. Ilyas, F. Hassan, and K. Rehman, "Performance Evaluation of Machine Learning Models for Cardiovascular Disease Prediction," Proceedings of the International Conference on Healthcare Informatics, pp. 112–118, 2025.
[6] S. Masbakhah, R. Prasetyo, and D. Wibowo, "Optimized Random Forest Model for Heart Disease Prediction Using FOX Algorithm," Journal of Electrical Engineering and Medical Informatics, vol. 7, no. 2, pp. 210–218, 2025.
[7] P. Sharma, R. Singh, and A. Verma, "Machine Learning Approaches for Early Detection of Heart Disease," International Journal of Creative Research Thoughts (IJCRT), vol. 13, no. 4, pp. 345–352, 2025.
[8] K. Rao, S. Patel, and M. Shah, "Comparative Study of Classification Algorithms for Heart Disease Prediction," in Proc. Int. Conf. Future Technologies (INCOFT), 2025, pp. 89–95.
[9] M. Azimi, A. Hosseini, and P. Kaur, "Heart Disease Prediction Using Machine Learning and Hyperparameter Optimization," arXiv preprint arXiv:2505.09969, 2025.
[10] M. Hasan, T. Rahman, and S. Islam, "An Ensemble Learning Approach for Heart Disease Prediction Using Machine Learning Techniques," arXiv preprint arXiv:2504.00485, 2025.
[11] M. Haque, R. Islam, and N. Chowdhury, "Multi-Class Heart Disease Prediction Using Machine Learning Techniques," arXiv preprint arXiv:2412.04792, 2024.
[12] S. Ahmad, A. Ali, and M. Khan, "Feature Selection Techniques for Improved Heart Disease Prediction Using Machine Learning," arXiv preprint arXiv:2503.16577, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)