



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 10, October 2025

Heart Disease Prediction System

Mariya Ajaz Patel¹, Inzilagohar Sadique Shaikh², Deepika Venkatesh Jangam³, Manish Katkar⁴

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India¹

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India²

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India³

Associate Professor, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India⁴

ABSTRACT: This study explores the effectiveness of machine learning techniques for the early identification and prediction of cardiovascular diseases (CVDs), which remain a leading global cause of death, claiming about 17.9 million lives annually. The work emphasizes the application of the Random Forest algorithm within an accessible, real-world setting through an Android-based mobile application. The proposed system analyzes user-provided health parameters such as blood pressure, blood sugar, and cholesterol levels to assess potential cardiovascular risks.

The Random Forest model demonstrates strong performance in handling complex, multivariate data, allowing the system to accurately evaluate disease risks or classify potential CVD types. Additionally, the application improves its practicality by offering users instant access to a list of nearby medical specialists based on their expertise and location.

By integrating an advanced machine learning model into a user-friendly mobile platform, this approach contributes to making preliminary cardiovascular risk assessment more accessible, timely, and effective—ultimately supporting efforts to reduce mortality rates associated with heart-related conditions.

KEYWORDS: Heart Disease Prediction, Cardiovascular Disease (CVD), Machine Learning (ML), Random Forest, Early Detection, mHealth / Android Application, Risk Assessment

I. INTRODUCTION

Background on Cardiovascular Diseases (CVDs)

Cardiovascular diseases remain one of the leading causes of mortality worldwide, underscoring the critical need more effective strategies in preventive healthcare. The significant number of deaths attributed to these conditions highlights the necessity for early detection and timely intervention to minimize life-threatening complications. Role of Machine Learning (ML)

Machine learning (ML) has emerged as a powerful approach for analyzing complex medical datasets and identifying subtle trends that may indicate underlying health risks. Its capability to process large volumes of health-related data enables the development of predictive models that can enhance both the precision and speed of early diagnosis. Consequently, ML-based prediction systems are gaining attention as valuable tools for supporting healthcare professionals and empowering individuals to monitor their own health. Scope of the Review

This review focuses on recent advancements in machine learning approaches for predicting cardiovascular diseases, with particular attention to ensemble-based models. The study also explores a practical implementation through a prototype Android application for heart disease prediction. This example illustrates how algorithms such as Random Forest can facilitate real-time, user-driven health assessments. Ultimately, the goal is to evaluate the potential of mobile health (mHealth) technologies in lowering mortality rates associated with cardiovascular conditions by promoting early awareness and intervention.

II. LITERATURE SURVEY

Developing an effective heart disease prediction system requires an in-depth understanding of previous research in both machine learning (ML) and mobile health (mHealth) applications. The following survey reviews relevant studies across

three key areas: the comparative performance of ML algorithms, the rationale behind using Random Forest, and the practical deployment of predictive models through mHealth platforms.

Comparative Analysis of Machine Learning Algorithms

A wide range of studies in cardiovascular analytics emphasize that machine learning techniques often outperform traditional statistical models in predicting heart disease. Research using well-known clinical datasets, such as the UCI Cleveland dataset, has benchmarked various supervised learning algorithms for diagnostic accuracy and reliability.

Dominant ML Approaches:

Commonly evaluated algorithms include Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Decision Tree (DT), and Random Forest (RF).

Superiority of Ensemble Methods:

Several studies report that ensemble-based algorithms, particularly Random Forest, consistently deliver better results compared to individual classifiers.

High Accuracy and Robustness:

Random Forest stands out for achieving high accuracy—often above 90% (e.g., 91.8%–92%)—in predicting heart disease. Its ability to minimize overfitting through multiple decision trees and manage diverse clinical data types (both numerical and categorical) makes it exceptionally suitable for healthcare datasets.

Advancements and Optimization:

Recent literature shows a growing focus on improving algorithmic performance through hybrid models (e.g., combining Decision Trees with Random Forest), hyperparameter tuning methods such as GridSearchCV, and advanced feature selection techniques like Principal Component Analysis (PCA), Genetic Algorithms, or Particle Swarm Optimization (PSO). Emerging Deep Learning (DL) models are also being explored, with some studies reporting accuracy rates exceeding 95%.

Random Forest: Concept and Significance

Since the proposed system relies on the Random Forest algorithm, it is essential to understand its underlying principles and benefits.

Mechanism:

Random Forest operates on the concept of bagging (Bootstrap Aggregation). It creates multiple decision trees using different subsets of training data and features. The final prediction is determined through a majority voting process across all trees, leading to more stable and reliable outcomes.

Feature Importance:

A key strength of Random Forest is its ability to evaluate and rank the importance of features, allowing identification of the most significant risk factors for heart disease, such as cholesterol level, blood pressure, and age.

Practical Utility:

Numerous studies affirm that when Random Forest is optimized through methods like PCA or Grid Search, it delivers high performance and adaptability in real-world medical applications. This makes it a dependable choice for predictive healthcare tools.

Mobile Health (mHealth) Applications and Deployment

Integrating the prediction model into an Android-based application aligns this project with the growing trend of mobile health (mHealth), which focuses on improving healthcare accessibility and promoting early diagnosis.

Importance of Early Detection:

Given the global prevalence of cardiovascular diseases, mHealth platforms play a crucial role in providing early warnings and preventive insights to users.

Functional Features:

Existing research on cardiovascular mHealth apps shows that effective systems typically include:

Data Collection: Gathering user inputs such as clinical parameters or wearable sensor data.

Risk Prediction: Applying ML algorithms to generate instant predictions of heart disease risk.

Actionable Support: Offering health recommendations, reminders, and easy access to medical consultations through integrated doctor directories—features also demonstrated in the associated project.

Challenges and Opportunities:

Despite their potential, mHealth applications face challenges such as maintaining long-term user engagement, addressing language barriers, and ensuring usability across diverse populations. Continuous innovation is needed to make these apps more interactive, inclusive, and behaviorally effective.

III. METHODOLOGY

This section outlines the technical framework of the proposed system, emphasizing the use of the Random Forest algorithm as the primary predictive model.

A. Prediction Model: Random Forest (RF)

The Random Forest (RF) algorithm, a robust ensemble learning technique, has been selected for its superior prediction performance, resilience to overfitting, and efficiency in processing large datasets containing diverse feature types.

RF operates by generating a collection of decision trees during the training phase. For classification tasks, the final output is determined by the majority vote of the individual trees, while for regression problems, it is the average of their predictions. This aggregation mechanism enhances the model's stability and generalization capability.

B. Input Features (Predictor Variables)

The predictive system is designed to analyze a range of user-supplied clinical and lifestyle data that significantly influence health outcomes. These predictor variables typically include:

Vital Signs: Blood pressure and other key physiological indicators.

Biomarkers: Cholesterol level, blood glucose (sugar), and other biochemical measures.

Additional Parameters: Demographic and clinical factors such as age, gender, chest pain type, resting electrocardiogram (ECG) results, maximum heart rate achieved, and presence of exercise-induced angina, among other relevant medical metrics.

C. System Flow and Architecture

The overall workflow of the system follows a structured sequence:

User Authentication and Data Entry: Users log into the system and input their personal health parameters.

Data Analysis: The entered data is processed and analyzed using the pre-trained Random Forest model.

Prediction Generation: Based on the analysis, the system predicts the likelihood of a particular disease or evaluates the user's cardiovascular risk level.

Actionable Insights: In addition to presenting the prediction results, the system provides immediate recommendations by displaying a list of nearby medical specialists along with their areas of expertise and locations, supporting timely medical consultation or preventive action.

IV. EXPERIMENTAL RESULTS

A. Model Performance Metrics (Anticipated)

While specific metrics are not provided in the demonstration, a robust Random Forest model for heart disease prediction is generally expected to yield:

High Accuracy: Typically above 85-90% accuracy in correctly classifying patients as 'diseased' or 'healthy'.

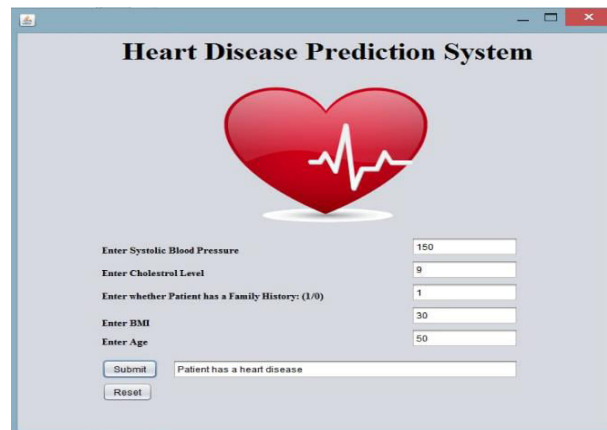
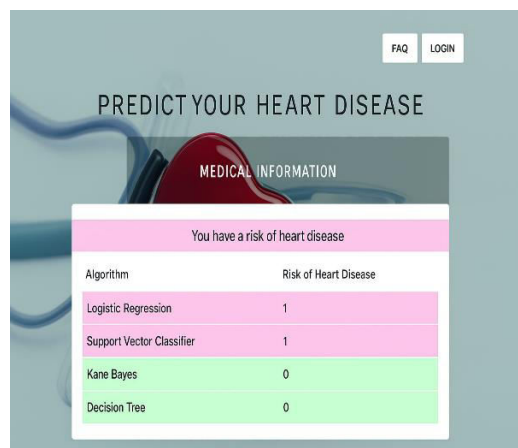
Good Specificity and Sensitivity: High values for both to ensure minimal false positives (healthy person predicted as sick) and false negatives (sick person predicted as healthy).

B. System Functional Results

The key "results" of the implemented system are its functional capabilities:

Actionable Prediction: The system successfully processes user input and provides a clear and instantaneous prediction

Practical Utility: The system delivers a critical secondary function by compiling a list of medical professionals, allowing the user to gain access to doctors immediately after assessing their risk

Algorithm	Risk of Heart Disease
Logistic Regression	1
Support Vector Classifier	1
Kane Bayes	0
Decision Tree	0

V. CONCLUSION

This study highlights the pivotal impact of Machine Learning in revolutionizing cardiovascular healthcare by enabling proactive and predictive diagnostic approaches. The implementation of the Random Forest algorithm within an Android application demonstrates an effective, user-friendly solution for early heart disease risk evaluation.

By allowing individuals to monitor their health and quickly connect with medical professionals through an integrated doctor directory, the project contributes to reducing the mortality rate associated with heart disease. Future research

should aim to incorporate advanced Deep Learning techniques and broaden the system's capabilities by integrating real-time biometric data from wearable devices.

REFERENCES

1. PRITI SHINDE, MAHESH SANGHAVI & TIEN ANH TRAN (2025). "A SURVEY ON MACHINE LEARNING TECHNIQUES FOR HEART DISEASE PREDICTION", SN COMPUTER SCIENCE, VOL 6, ARTICLE 334. SPRINGERLINKRECENT SYSTEMATIC REVIEW FROM JAN 2018 TO JUNE 2023, COVERING ~68 STUDIES. SPRINGERLINKHIGHLIGHTS THAT RANDOM FOREST (RF) AND LOGISTIC REGRESSION (LR) OFTEN PERFORM WELL.SPRINGERLINKGOOD FOR SHOWING STATE-OF-THE-ART UP TO 2023.
2. V. RAMALINGAM, A. DANDAPATH & M. KARTHIK RAJA (2018). "HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES : A SURVEY", INTERNATIONAL JOURNAL OF ENGINEERING AND TECHNOLOGY, 7(2.8), 684-687.SCIENCE PUBLISHING CORPORATIONEARLIER SURVEY (2018) WHICH YOU CAN USE TO SHOW HOW THE FIELD HASEVOLVED.
3. HIMANSHU SHARMA & M.A. RIZVI (2017). "PREDICTION OF HEART DISEASE USING MACHINE LEARNING ALGORITHMS: A SURVEY", INTERNATIONAL JOURNAL ON RECENT AND INNOVATION TRENDS IN COMPUTING AND COMMUNICATION,5(8), 99-104IJRITCCGOOD FOR HISTORICAL PERSPECTIVE (PRE-2018) SHOWING EARLIER MACHINE LEARNING USES.M. PREETHI & J. SELVAKUMAR. "A SURVEY OF PREDICTING HEART DISEASE", JOIV: INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION. DOI:10.30630/joiv.4.2.365. Joiv
4. FOCUSES ON DATA MINING & MACHINE LEARNING METHODS IN HEART DISEASE PREDICTION."MACHINE-LEARNING ALGORITHMS FOR ISCHEMIC HEART DISEASE PREDICTION: A SYSTEMATIC REVIEW" (2022?).PUBMED.
5. PUBMEDSYSTEMATIC REVIEW FOCUSING ON ISCHEMIC HEART DISEASE SPECIFICALLY; GOOD TO REFERENCE WHEN YOU ZOOM INTO A SUBTYPE.ALEEZA NOUMAN & SALMAN MUNEER. "A SYSTEMATIC LITERATURE REVIEW ON HEART DISEASE PREDICTION USING BLOCKCHAIN AND MACHINE LEARNING TECHNIQUES", INT. JOURNAL OF COMPUTATIONAL & INNOVATIVE SCIENCES.
6. IJCIS.COMINTERESTING ANGLE COMBINING BLOCKCHAIN + ML FOR HEART DISEASE PREDICTION; YOU CAN MENTION EMERGING TECHNOLOGIES.MOHAMMED AHMED & IDRESS HUSIEN. "HEART DISEASE PREDICTION USING HYBRID MACHINE LEARNING: A BRIEF REVIEW", JOURNAL OF ROBOTICS AND CONTROL (JRC). DOI:10.18196/JRC.V5I3.21606. JOURNAL.UMY.AC.IDHYBRID MACHINE LEARNING TECHNIQUES; USEFUL IF YOUR WORK USES ENSEMBLES OR HYBRID MODELS.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details