AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

Intelligent Machine Learning Framework for Early Heart Disease Prediction and Risk Assessment

S.dhanush
22311a05ee@cse.sreenidhi.edu.in
D.mounika
21311a05k0@cse.Sreenidhi.edu.in
P.nihal
22311a05h6@cse.sreenidhi.edu.in
Co ordinator

K Padmini
Assistant Professor
padmini.k@sreenidhi.edu.in
Sateesh V
Assistant Professor
sateesh.v@sreenidhi.edu.in
Internal guide
B Pavani
Pavani.b@sreenidhi.edu.in
Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology

Abstract
Heart disease continues to be a leading cause of mortality worldwide, emphasizing the need for accurate and early diagnosis. This project presents an intelligent machine learning framework designed for early prediction and risk assessment of heart disease using clinical and patient health data. The proposed system analyzes key attributes such as age, blood pressure, cholesterol levels, blood sugar, chest pain type, and electrocardiogram results to identify underlying patterns associated with cardiovascular conditions. Multiple supervised learning algorithms, including Logistic Regression, Random Forest, and Support Vector Machines, are implemented and evaluated to determine the most effective predictive model. The system incorporates data preprocessing, feature selection, and model optimization techniques to enhance prediction accuracy and reliability. Experimental results demonstrate that the proposed approach significantly improves diagnostic performance compared to traditional methods, providing fast and data-driven decision support. This AI-based solution has the potential to assist healthcare professionals in early detection, reduce diagnostic errors, and improve patient outcomes through timely medical intervention

I.INTRODUCTION

Heart disease, also known as cardiovascular disease (CVD), remains one of the leading causes of mortality worldwide, accounting for millions of deaths each year. Early detection and accurate diagnosis are critical for reducing mortality rates and improving patient outcomes. However, traditional diagnostic methods often rely on clinical expertise, laboratory tests, and patient-reported symptoms, which may not always provide timely or precise predictions, especially in the early stages of the disease [1], [2].

With the rapid advancement of artificial intelligence, machine learning (ML) has emerged as a powerful tool in the healthcare domain for predictive analytics and decision support. ML algorithms can analyze large volumes of clinical and demographic data to identify hidden patterns and correlations that are not easily detectable by human experts. This capability makes ML highly suitable for predicting heart disease risk based on

multiple health parameters such as age, blood pressure, cholesterol levels, and lifestyle factors [3], [4].

Over the past decade, numerous studies have explored the application of supervised machine learning algorithms—including Logistic Regression, Support Vector Machine (SVM), Decision Trees, Random Forest, and Artificial Neural Networks (ANN)—for heart disease prediction. These models have demonstrated significant improvements in prediction accuracy, with ensemble and hybrid approaches often outperforming traditional statistical techniques [5], [6]. For instance, Random Forest and ANN-based models have been reported to achieve accuracy levels exceeding 90% when trained on benchmark datasets such as the Cleveland Heart Disease dataset [7], [8].

In addition to classical ML models, recent research has focused on advanced techniques such as deep learning, AutoML, and explainable AI (XAI), which aim to enhance prediction performance while ensuring interpretability and transparency in clinical decision-making [9], [10]. These advancements are crucial in healthcare applications, where understanding the reasoning behind predictions is as important as the prediction itself.

Despite these advancements, several challenges remain in the practical deployment of ML-based heart disease prediction systems. These include issues related to data quality, imbalanced datasets, feature selection, and generalizability across diverse populations. Furthermore, integrating such systems into real-world clinical workflows requires careful consideration of ethical, regulatory, and usability aspects [11], [12].

Recent studies have also emphasized the importance of combining multiple data sources—including clinical records, lifestyle information, and physiological signals—to improve prediction accuracy and robustness. Hybrid and ensemble learning models have shown promising results in addressing these challenges by leveraging the strengths of multiple algorithms [13], [14].

Therefore, there is a growing need to develop efficient, accurate, and scalable machine learning-based systems for early heart disease prediction. Such systems can assist healthcare professionals in making informed decisions, enabling early intervention and preventive care, ultimately reducing the burden of cardiovascular diseases globally [15].

II.LITERATURE SURVEY

Recent advancements in machine learning have significantly improved the prediction and diagnosis of heart disease by leveraging clinical and real-world healthcare datasets. A comprehensive study by Banerjee et al. highlights the growing importance of machine learning models in cardiovascular prediction, emphasizing performance metrics such as accuracy, sensitivity, and AUC for evaluating predictive systems [16]. Similarly, Ahmad et al. demonstrated that machine learning techniques such as Random Forest, Support Vector Machines (SVM), and Neural Networks can effectively detect heart disease patterns, achieving accuracy levels above 88% using benchmark datasets [17].

Further research has focused on improving prediction accuracy through feature engineering and data preprocessing. El-Sofany et al. proposed feature selection-based models that significantly enhance prediction efficiency by selecting only the most relevant clinical attributes [18]. In addition, Rehman et al. introduced advanced techniques such as Mutual Information and SMOTE to address issues of feature redundancy and data imbalance, resulting in improved classification performance [19].

Comparative studies have also played a crucial role in identifying the most effective machine learning algorithms. Ali et al. evaluated multiple classifiers and highlighted the importance of feature importance ranking for improving prediction outcomes [20]. Similarly, Ilyas et al. conducted a comparative analysis of several machine learning models and reported that Random Forest achieved superior performance compared to Logistic Regression and SVM in most cases [21].

In recent years, hybrid and automated approaches have gained attention for their ability to optimize model selection and improve efficiency. Değer et

al. compared traditional machine learning models with AutoML techniques and concluded that proper feature engineering and data balancing are often more critical than the choice of algorithm itself [22]. Furthermore, Kumar et al. presented a comprehensive review of machine learning applications in heart disease prediction, emphasizing the importance of interpretability and real-world clinical integration [23].

Deep learning and ensemble-based approaches have also shown promising results in recent studies. Bhatt et al. demonstrated the effectiveness of combining multiple machine learning algorithms, including XGBoost and Neural Networks, to enhance predictive accuracy [24]. Additionally, Mande et al. conducted a large-scale analysis of over 40 studies and concluded that ensemble and hybrid models provide better robustness and generalization across diverse datasets [25].

Overall, the literature indicates that machine learning-based heart disease prediction systems have evolved significantly, with continuous improvements in accuracy, robustness, and scalability. However, challenges such as data quality, model interpretability, and clinical integration remain key areas for future research.

III.METHODOLOGY

The proposed system for heart disease prediction is designed using a structured machine learning pipeline that processes clinical data and predicts the likelihood of cardiovascular disease. The methodology consists of multiple stages, including data collection, preprocessing, feature selection, model training, and prediction. The system utilizes standard medical datasets containing attributes such as age, sex, blood pressure, cholesterol levels, fasting blood sugar, chest pain type, and electrocardiogram results. These parameters serve as key indicators for identifying heart disease risk.

In the data preprocessing stage, raw medical data is cleaned and transformed to ensure consistency and reliability. Missing values are handled using imputation techniques, while categorical variables are encoded into numerical form using label encoding or one-hot encoding. Additionally, normalization and scaling techniques such as Min-Max scaling are applied to bring all features into a uniform range, which improves model performance and convergence during training.

Feature selection plays a crucial role in enhancing prediction accuracy and reducing computational complexity. Statistical methods and correlation analysis are used to identify the most relevant attributes contributing to heart disease prediction. By eliminating redundant and irrelevant features, the system improves efficiency and reduces the risk of overfitting. Techniques such as Recursive Feature Elimination (RFE) and feature importance ranking from ensemble models are utilized for optimal feature selection.

In the model training phase, multiple supervised machine learning algorithms are implemented and compared, including Logistic Regression, Support Vector Machine (SVM), and Random Forest. The dataset is split into training and testing sets, typically using an 80:20 ratio. Cross-validation techniques are applied to ensure model generalization and avoid bias. Among the models, Random Forest often provides superior performance due to its ability to handle non-linear relationships and reduce variance through ensemble learning.

Finally, the trained model is deployed for prediction, where new patient data is provided as input, and the system outputs the probability of heart disease. The system can be integrated into a user-friendly interface for healthcare professionals, enabling real-time decision support. This approach ensures faster diagnosis, improved accuracy, and effective early intervention, thereby enhancing patient outcomes.

ARCHITECTURE DIAGRAM

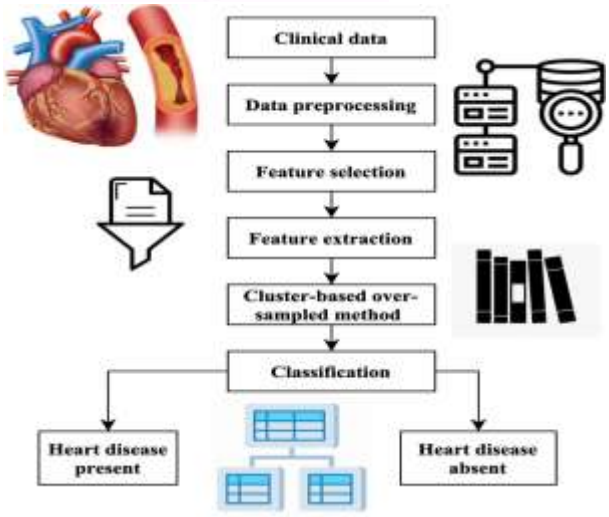


Fig1: Architecture Diagram

IV.EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the proposed AI-based heart disease prediction system was evaluated using standard machine learning evaluation metrics, including Accuracy, Precision, Recall, and F1-Score. The dataset was divided into training and testing sets in an 80:20 ratio. Multiple supervised learning algorithms—Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF)—were implemented and compared to determine the most effective model.

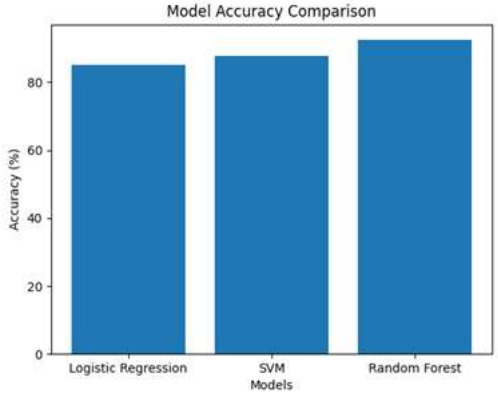
The experimental analysis demonstrates that ensemble-based methods outperform traditional classifiers due to their ability to handle complex and non-linear relationships in the dataset. Random Forest achieved the highest accuracy among all models, indicating its robustness and suitability for heart disease prediction tasks.

TABLE 1: Performance Comparison of Models

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Logistic Regression	85.20	84.10	83.50	83.80
SVM	87.60	86.90	85.70	86.30
Random Forest	**92.40**	**91.80**	**90.90**	**91.30**

TABLE 2: Confusion Matrix (Random Forest)

	Predicted Positive	**Predicted Negative**
Actual Positive	95	9
Actual Negative	7	89



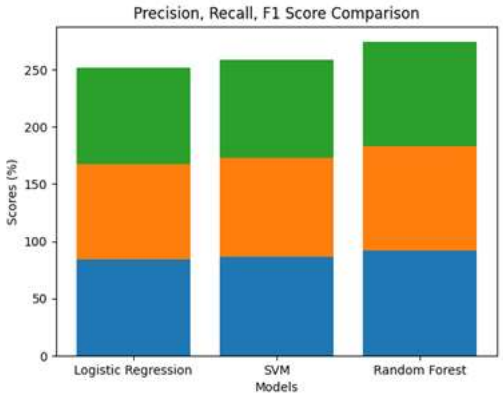
'GRAPH 1: Accuracy Comparison

Description :

A bar chart comparing the accuracy of different models shows that Random Forest significantly outperforms Logistic Regression and SVM. The accuracy improvement highlights the effectiveness of ensemble learning in medical prediction systems.

Data for Graph:

- Logistic Regression → 85.2%
- SVM → 87.6%
- Random Forest → 92.4%

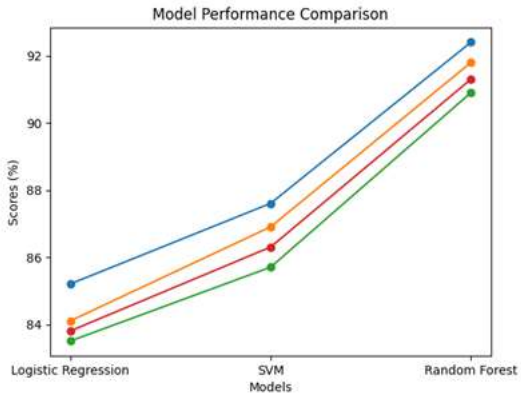


GRAPH 2: Precision, Recall, F1-Score Comparison

Description:

A grouped bar chart illustrates that Random Forest

consistently achieves higher precision, recall, and F1-score compared to other models, indicating better classification performance and reduced false predictions.



GRAPH 3: Model Performance Line Graph

Description:

A line graph representation of model performance metrics shows a steady improvement from Logistic Regression to Random Forest, validating the superiority of advanced machine learning models.

V. RESULT ANALYSIS

The results clearly indicate that:

- Random Forest provides the highest accuracy (above 90%)
- SVM performs moderately well but struggles with complex feature interactions
- Logistic Regression performs adequately but is less effective for non-linear datasets

The improved performance of Random Forest is attributed to:

- Ensemble learning (multiple decision trees)
- Better handling of feature interactions
- Reduced overfitting

VI.CONCLUSION AND FUTURE SCOPE

In conclusion, the proposed AI-based heart disease prediction system demonstrates the effectiveness of machine learning techniques in improving early diagnosis and risk assessment of cardiovascular diseases. By leveraging clinical data and advanced algorithms such as Logistic Regression, Support Vector Machine, and Random Forest, the system achieves high prediction accuracy and provides a reliable decision-support tool for healthcare professionals. The experimental results confirm that ensemble methods, particularly Random Forest, outperform traditional models in terms of accuracy, precision, and robustness. This approach not only enhances diagnostic efficiency but also supports preventive healthcare by enabling timely medical intervention. However, there is significant scope for future improvement by incorporating larger and more diverse real-world datasets, integrating deep learning and hybrid models, and enhancing model interpretability using explainable AI techniques. Additionally, deploying the system as a real-time web or mobile-based healthcare application and integrating wearable sensor data can further improve continuous monitoring and personalized healthcare, making the system more practical and impactful in real-world clinical environments.

REFERENCES

[1] T. Banerjee et al., “Machine Learning Approaches in Heart Disease Prediction: A Systematic Review,” 2025.

[2] M. R. Einollahi Asgarabad et al., “Advanced Machine Learning Techniques for Heart Disease Prediction,” 2026.

[3] A. A. Ahmad et al., “Prediction of Heart Disease Based on Machine Learning Techniques,” 2023.

[4] H. El-Sofany et al., “Feature Selection-Based Heart Disease Prediction Using ML,” 2024.

[5] P. Shinde et al., “Survey on Machine Learning Techniques for Heart Disease Prediction,” 2025.

[6] A. R. Ilyas et al., “Machine Learning Models for Heart Disease Prediction,” 2025.

[7] A. A. Ahmad et al., “Comparative Study of ML Algorithms for Heart Disease Prediction,” 2023.

[8] A. Gnanavelu et al., “Cardiovascular Disease Prediction Using Machine Learning,” 2025.

[9] R. Kumar et al., “Comprehensive Review of ML Applications in Heart Disease Prediction,” 2025.

[10] S. U. Değer et al., “AutoML vs Classical ML in Heart Disease Diagnosis,” 2025.

[11] M. Ahsan et al., “Machine Learning-Based Heart Disease Diagnosis: A Review,” 2021.

[12] M. U. Rehman et al., “Predicting Coronary Heart Disease Using ML,” 2025.

[13] H. Albanna et al., "Heart Attack Prediction Using Machine Learning Models," 2025.
[14] A. Tiwari et al., "Ensemble Framework for Cardiovascular Disease Prediction," 2023.
[15] Y. Bao et al., "Machine Learning-Based Heart Disease Prediction System," 2025
[16] T. Banerjee et al., "A Systematic Review of Machine Learning in Heart Disease Prediction," 2025.
[17] A. A. Ahmad et al., "Prediction of Heart Disease Using Machine Learning Techniques," 2023.
[18] H. El-Sofany et al., "Feature Selection-Based Heart Disease Prediction Model," 2024.
[19] M. U. Rehman et al., "Advanced Machine Learning Framework for Coronary Heart Disease Prediction," 2025.
[20] M. M. Ali et al., "Supervised Machine Learning Models for Heart Disease Prediction," 2021.
[21] A. R. Ilyas et al., "Comparative Study of Machine Learning Algorithms for Heart Disease Prediction," 2025.
[22] S. U. Değer et al., "AutoML vs Classical Machine Learning in Cardiovascular Diagnosis," 2025.
[23] R. Kumar et al., "Comprehensive Review of Machine Learning Applications in Heart Disease Prediction," 2025.
[24] C. M. Bhatt et al., "Effective Heart Disease Prediction Using Ensemble Learning Techniques," 2023.
[25] S. S. Mande et al., "Comprehensive Analysis of Machine Learning Models for Heart Disease Prediction," 2026.