Heart Attack Prediction by Machine Learning

Shoaib N. Ghodimar[1], Rutika R. Nangare[2], Tanishka K. Kamble[3], Pranali P. Lohar[4] Mrs. Swati More[5]
[1,2,3,4,5]*Department of Computer Science and Engineering, D. Y. Patil Education Society's D. Y. Patil Technical Campus, Faculty of Engineering & Faculty of Management, Talsande, Kolhapur, India*

***Abstract*—One of the main causes of death globally, heart disease, makes early and precise forecasting necessary for good prevention and treatment. Using demographic and clinical data, this research suggests a machine learning-based approach to forecasting heart disease. Among other medical factors, the dataset contains crucial health-related characteristics, including age, cholesterol level, and blood pressure. To enhance the data quality and model performance in this study, data preprocessing methods, including dealing with missing values, categorical feature encoding, and feature scaling, were employed. K-Nearest Neighbors (KNN), Decision Tree, and Random Forest, among several supervised machine learning algorithms, are executed and evaluated. Cross-validation approaches are used to evaluate the models to guarantee generalization and dependability. The experimental findings show that ensemble-based techniques, especially random forest, provide more stable and precise predictions than other models. After adjusting the parameters, the optimized KNN model also showed competitive performance. By offering a quick and data-driven approach for the early identification of heart disease, the suggested system can help healthcare professionals. In the future, this might help patients' results and prompt medical care.**

***Index Terms*—Machine Learning, K Neighbors Classifier, Decision Tree Classifier, Random Forest Classifier, HTML, CSS, Python**

I. INTRODUCTION

One of the main causes of death worldwide, cardiovascular diseases (CVDs)—especially heart disease—challenge healthcare systems considerably. Early identification and correct diagnosis are absolutely necessary to lower mortality rates and raise patient outcomes. Traditional diagnostic techniques, meanwhile, depend on manual analysis and clinical judgment, hence making them error-prone and time-consuming. As healthcare information becomes more accessible, there is more opportunity to use machine learning tools for accurate and dependable disease forecasts.

Because it can detect sophisticated patterns and correlations in large datasets, machine learning has become a strong instrument in medical diagnostics. Machine learning systems can help forecast the probability of diseases by studying past patient data, hence assisting doctors in making wise decisions. Predictive models can assess several risk indicators in the setting of heart disease—such as age, blood pressure, cholesterol levels, and other clinical characteristics—to estimate the likelihood of events.

Utilizing supervised machine learning algorithms, this work seeks to create a heart disease prediction model. To guarantee data quality and consistency, the investigation included data preprocessing methods such data cleaning, encoding of categorical variables, and feature scaling. To examine and forecast heart disease occurrence, several classification methods have been used, including Decision Trees, Random Forest, and K-Nearest Neighbors. Performance measures and cross-validation methods were used to assess the efficacy of the suggested system and guarantee the reliability and generalizability of the models. This work aimed at developing a predictive system capable of real-world uses that is not only precise but also effective and scalable. For healthcare practitioners, such a system can be a decision-support tool that allows for early diagnosis and appropriate treatment hence lowering the healthcare load and improving patient care.

II. LITERATURE REVIEW

Using data mining and machine learning methods, prediction of heart disease has been extensively studied. Fast medical care data development has led scientists to investigate several and many

computational approaches to help medical professionals in the first detection of cardiovascular illnesses as well as to increase diagnostic precision.

1. Paper 1: Heart Attack Classification Based on Random Forest:

Focus of this research is on using clinical and biochemical information to predict first heart attack via the Random Forest method. The patient characteristics age, blood pressure, glucose level, and cardiac biomarkers made up the dataset. The authors enhanced the model performance by means of preprocessing methods including label encoding, data balance, SMOTE application, and feature scaling. The model was assessed using a 10-fold cross-validation technique to ensure robustness and generalization. With some folds reaching nearly-perfect precision, accuracy, recall, and F1-score, the model achieved very high performance. Variable relevance analysis showed that troponin, potassium levels, and age clearly affect prediction.

2. Paper 2: Heart Disease Diagnosis Using Machine Learning and Data Mining:

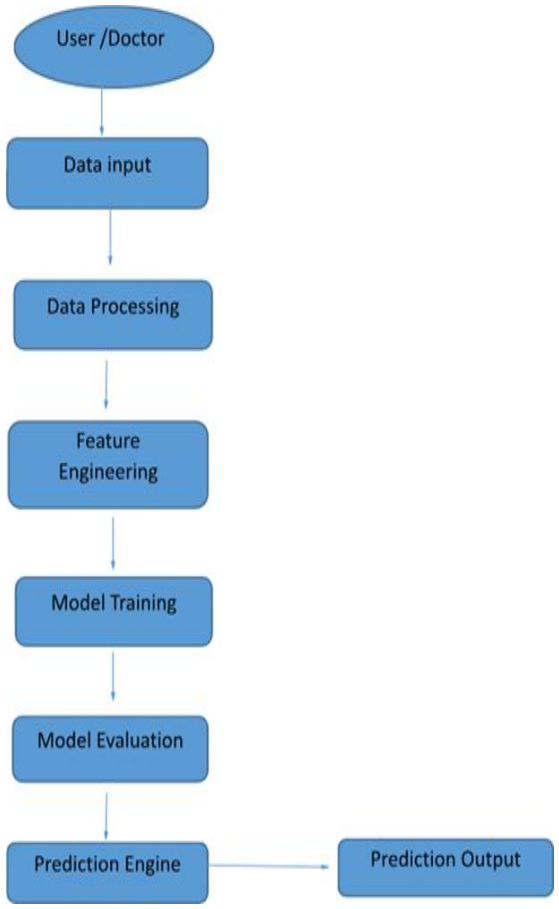
Using datasets like the UCI repository, this work examines how machine learning and data mining methods may help to predict heart disease. Several approaches combining Support Vector Machines, Decision Trees, Naive Bayes, and Random Forest were studied. The research shows how machine learning can find hidden patterns in huge and significant medical care data sets, hence enabling early detection and reducing misdiagnosis. Among the tested models, Random Forest achieved around 89.4% accuracy, demonstrating strong predictive ability.

3. Paper 3: Machine learning early heart attack prediction:

Utilizing machine learning models, this research emphasizes how crucial first identification of heart attacks is. Demographic, lifestyle, and medical characteristics make up a large and significant dataset. The data was divided such training and testing sets; sophisticated optimization methods were used to execute hyperparameter tuning. While machine learning models offer faster and more affordable and cost-effective predictions, the study points out that traditional and customary diagnostic techniques are pricey and time-consuming. For binary classification tasks, Light GBM and ensemble methods demonstrated encouraging and positive results among several and varied algorithms.

Research Paper	Year	Authors	Research Gap
Random Forest-Based Heart Attack Classification	June 2025	Suleyman Dal and Necmettin Sezgin	Study demonstrates high accuracy using the Random Forest algorithm, it primarily focuses on a single model without comparing its performance with other machine learning algorithms.
Machine Learning & Data Mining for Heart Disease Diagnosis	December 2022	Intisar Ahmed	This study explores multiple machine learning algorithms; however, it lacks a standardized preprocessing pipeline, which can affect the fairness of model comparison. The dataset used is relatively small and may not represent diverse patient populations.
Early Heart Attack Prediction Using Machine Learning	February-April 2025	Sarang Ali, Abuzar Shahid , Gohar Mumtaz	Emphasizes model performance without sufficient focus on interpretability, which is critical in medical applications.

III. SYSTEM ARCHITECTURE



IV WORKING PROCESS

1. Loading of dataset:

The system first loads the heart disease dataset using data handling libraries. There are many medical characteristics needed for prediction among the data. Utilizing: CSV file loading of datasets.

2. Data knowledge and analysis:

First analysis is done to grasp the structure of the dataset, including:

1. Feature details and data kinds 2. Statistical summary including mean, std, etc. 3. Characteristics correlation Patterns and correlations are found by visual methods including histograms and heatmaps.

3. Information Treatment:

Several preprocessing phases are used to get the dataset ready for machine learning models:

1. Categorical data are changed into a numerical form by dummy variables.
2. Feature Scaling: Normalizing feature values calls for standardizing, which is especially crucial for distance-based models like KNN.

4. Feature Selection and Splitting:

The dataset is divided into:1. Input Features (X) → Independent variables 2. Target Variable (y) → Output (heart disease presence)

5. Train Modeling

1. K-Nearest Neighbors (KNN): i. Several K values are tested. ii. For every value, precision is determined. iii. The best K value is chosen.
2. Random Forest Classifier: Trained utilizing ensemble learning; handles challenging feature interactions well.
3. Decision Tree:- Used as a baseline model for comparison

6. Evaluation of Models:

Evaluation is done using: to guarantee model dependability. I. Cross-validation (ten times) ii. Comparison of model accuracy. This measure helps to prevent overfitting and guarantees that the model

7. works well on previously unknown data.

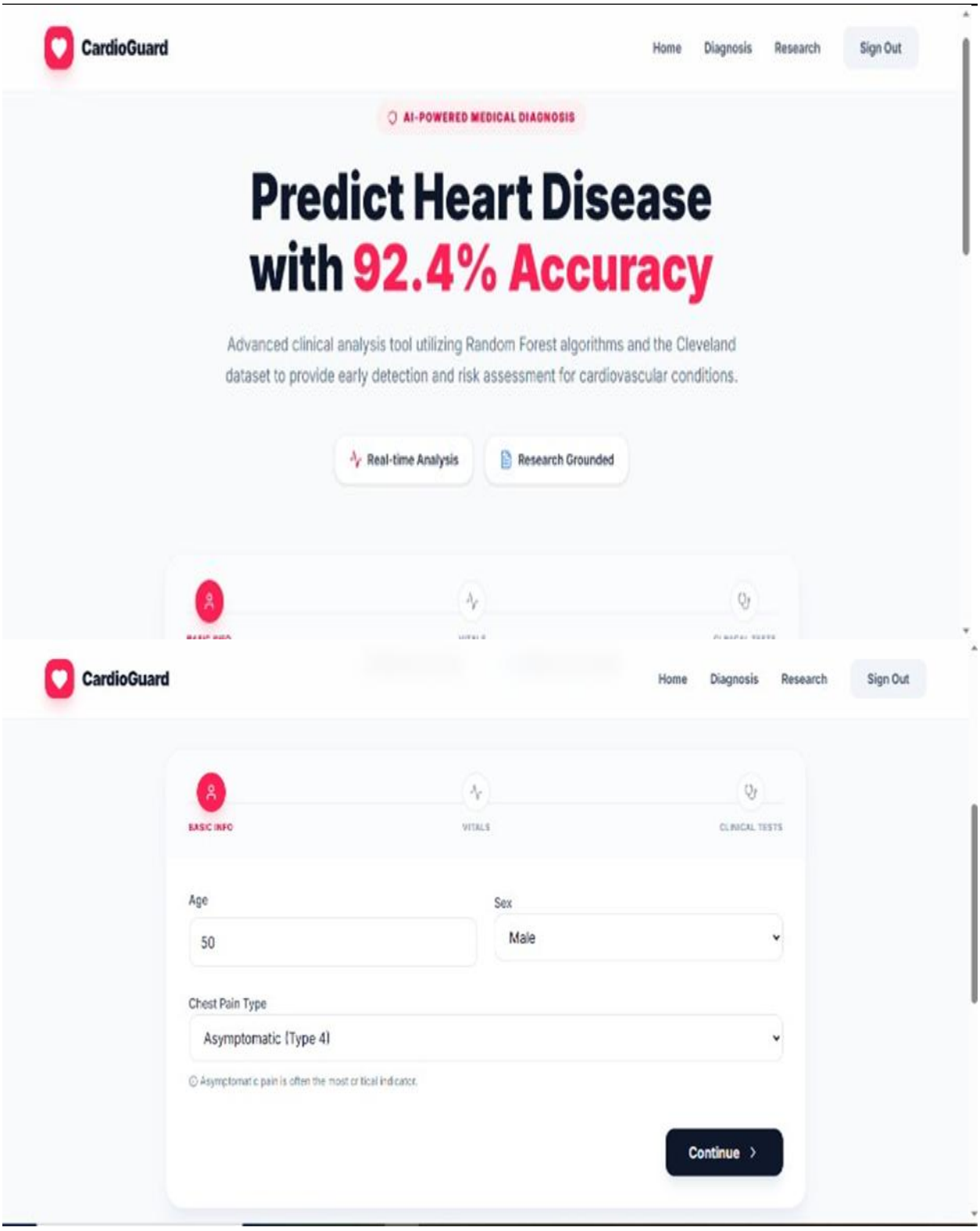
8. Results Visualizations:

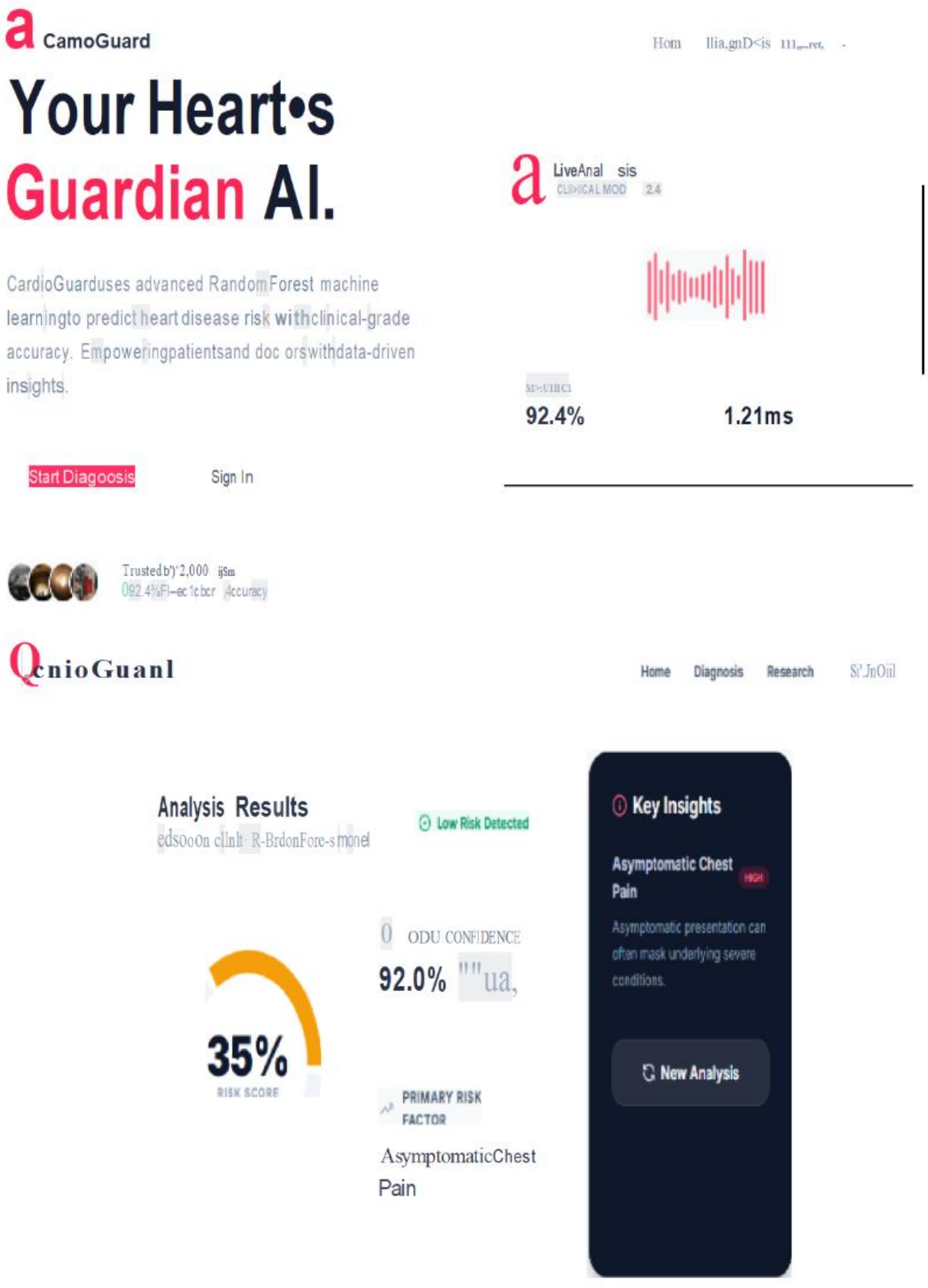
Graphs are plotted to: i. Display KNN accuracy against the number of neighbors. ii. Compare model performance so that one might choose the best-performing one.

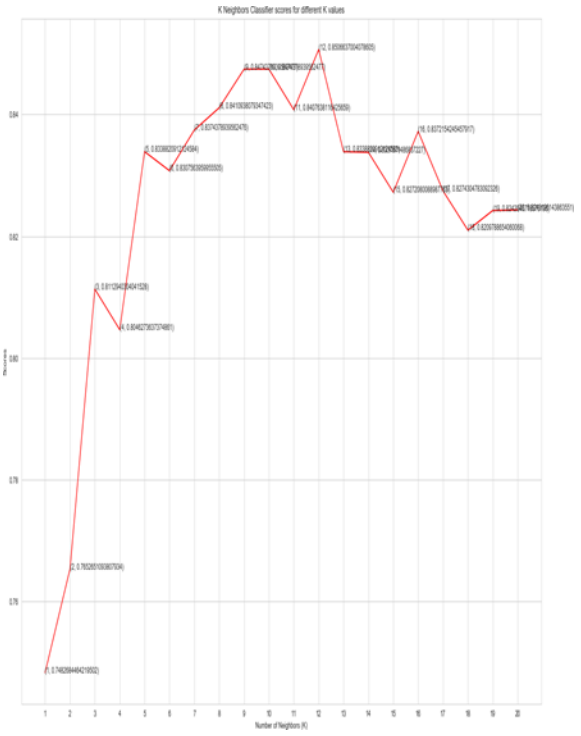
9. Predictative Stage:

Following instruction, the ultimate model predicts outcomes: i. Fresh input data is given. ii. Model processes the input. iii. Output results from: a. 1: Heart Disease Present; b. 0: No Heart Disease.

V. RESULT


CardioGuard
Home
Diagnosis
Research
Sign Out
AI-POWERED MEDICAL DIAGNOSIS
Predict Heart Disease with 92.4% Accuracy
Advanced clinical analysis tool utilizing Random Forest algorithms and the Cleveland dataset to provide early detection and risk assessment for cardiovascular conditions.
Real-time Analysis
Research Grounded
CardioGuard
Home
Diagnosis
Research
Sign Out
BASIC INFO
VITALS
CLINICAL TESTS
Age
50
Sex
Male
Chest Pain Type
Asymptomatic (Type 4)
Asymptomatic pain is often the most critical indicator.
Continue


CamoGuard
Your Heart•s
Guardian AI.
CardioGuarduses advanced RandomForest machine learningto predict heart disease risk withclinical-grade accuracy. Empoweringpatientsand doc orswithdata-driven insights.
Start Diagoosis
Sign In
LiveAnal sis
92.4%
1.21ms
Trusted b)'2,000
Home
Diagnosis
Research
Analysis Results
Low Risk Detected
35%
RISK SCORE
92.0%
PRIMARY RISK FACTOR
AsymptomaticChest Pain
Key Insights
Asymptomatic Chest Pain
Asymptomatic presentation can often mask underlying severe conditions.
New Analysis



VI. FUTURE WORK

1. Including larger and more varied datasets with real-time clinical data from many populations would improve the model. This would increase the system's generalizability and dependability in actual healthcare settings. Adding more sophisticated machine learning and ensemble approaches might also increase the robustness and accuracy of predictions. 2. Another significant direction is the inclusion of feature selection and dimensionality reduction approaches to find the most pertinent medical qualities and minimize computing complexity. This would preserve accuracy while also boosting model efficiency. 3. Making the prediction model readily available to medical professionals and patients, developing a real-time application interface like a web or mobile-based platform can also help to expand the system. This would allow for utility in clinical environments and operational deployment. 4. In medical uses especially, then, enhancing the interpretability of the model is quite important. Future research might concentrate on using interpretable machine learning approaches to provide clear insights into how predictions are produced, therefore strengthening medical practitioners' trust.

VII. CONCLUSION

Using machine learning techniques on clinical and demographic data, this research provides a way to forecast heart disease. Review of the available studies reveals that different machine learning techniques, including Random Forest, K-Nearest Neighbors (KNN), and other ensemble approaches, have been effectively used to boost prediction accuracy and help early diagnosis of cardiovascular diseases. Data preprocessing, feature engineering, model training, and evaluation were all part of a structured pipeline used in the suggested system. To improve model performance, techniques including characteristic scaling and encoding of categorical variables were used. Cross-validating several machine learning models—including K-Nearest Neighbors and Random Forest—was done to guarantee reliability and generality.

According to experimental data, model performance depends on algorithm selection and parameter tuning. While the Random Forest model's ensemble character guaranteed strong and reliable results, the KNN model showed better performance following fine-tuning of the parameter K. These results fit with earlier research showing that ensemble techniques usually surpass single models in terms of stability and accuracy [1]. Although performing well, some constraints persist, including reliance on data source

quality and poor real-world validation. Still, the suggested approach shows how early heart disease prediction and support for clinical decision-making may be greatly aided by machine learning methods.

REFERENCES

[1] S. Ali, A. Shahid, and G. Mumtaz, "Early Heart Attack Prediction by Using Machine Learning,"International Journal of Social Sciences Bulletin, vol. 1, no. 1, 2025. https://www.researchgate.net/publication/392842832_EARLY_HEART_ATTACK_PREDICTION_BY_USING_MACHINE_LEARNING

[2] Intisar Ahmed "A Study of Heart Disease Diagnosis Using Machine Learning and Data Mining Techniques,"California State University, San Bernardino (CSUSB), Master's Project, 2022. https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2748&context=etd

[3] Suleyman Dal and Necmettin Sezgin , "Heart Attack Classification with a Machine Learning Approach Based on the Random Forest Algorithm,"ResearchGate Publication, 2025. https://www.researchgate.net/publication/393621267_Heart_Attack_Classification_with_a_Machine_Learning_Approach_Based_on_the_Random_Forest_Algorithm

[4] "Dataset" https://www.kaggle.com/datasets/ronitf/heart-disease-uci "Heart Disease Prediction Using Machine Learning Algorithms" Harshit Jindal et al., 2021 https://www.researchgate.net/publication/348604625_Heart_disease_prediction_using_machine_learning_algorithms?utm_source=chatgpt.com

[5] "Heart Disease Prediction Using Machine Learning" Baban U. Rindhe et al., 2021 https://www.researchgate.net/publication/351545128_Heart_Disease_Prediction_Using_Machine_Learning?utm_source=chatgpt.com

[6] Prediction of Heart Disease: A Comparative Study of Machine Learning Algorithms" Reham Ali et al., 2025 https://www.researchgate.net/publication/390877685_Prediction_of_Heart_Disease_A_Comparative_Study_of_Machine_Learning_Algorithms?utm_source=chatgpt.com

[7] Pattern Recognition and Machine Learning Author: Christopher M. Bishop

[8] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Author: Aurélien Géron