iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: VI Month of publication: June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83832

Early Heart Disease Prediction Using Machine Learning Algorithms with Power-Bi Visualization

Elakia K[1], Dr. R.G. Suresh Kumar [2], Preethi G[3]
[1]*Assistant Professor, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*
[1]*Professor and Head of the Department, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*
[3]*B.Tech, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*

Abstract-Artificial Intelligence enhances predictive healthcare by enabling faster and more accurate cardiovascular disease risk estimation. By integrating AI with Machine learning and Data science, cardiology has been transformed through reliable, accessible, and timely diagnostics. Artificial Intelligence optimizes cardiovascular disease detection by supporting informed clinical decisionmaking, enabling real-time patient monitoring, promoting early detection and personalized treatment planning. Moreover, AI-driven predictive analytics and visualization techniques improve cardiovascular risk prediction and management. The existing studies uses classifiers such as Logistic Regression, Decision Trees, Random Forests, KNN, and SVM, including Adaptive Boosting, Bagging, Voting, and Extreme Gradient Boosting to improve accuracy. Data imbalance is addressed using SMOTE, while feature selection techniques like chisquare, ANOVA, and Mutual Information refine predictors. Explainable AI, through SHAP, increases model transparency and is incorporated in mobile applications. Eventhough advancements exist, AIdriven cardiovascular prediction faces significant challenges, including limited and imbalanced datasets that increase over-fitting, high implementation costs, and inadequate evaluation of mobile usability. Labor-intensive diagnostic procedures like CT scans, ECGs, and MRIs impede timely cardiovascular disease detection. To overcome these problems, I propose collecting a balanced dataset, applying multiple machine learning algorithms, Box-Cox for normalization and Kaplan-Meier analysis for survival analysis, enhancing feature selection, and deploying interactive real-time visualizations through PowerBI to build a scalable, personalized cardiovascular decision support system.
Keywords: Artificial Intelligence, Machine learning, Data science, Cardiovascular Disease, Predictive analytics, SMOTE, Feature selection, Box-Cox, Kaplan-Meier, Power BI.

I. INTRODUCTION

The heart is a crucial muscular organ that plays a fundamental role in the human circulatory system by maintaining continuous blood flow throughout the body. It ensures the delivery of oxygen and essential nutrients to tissues while removing metabolic waste, thereby helping to maintain physiological balance. The cardiovascular system, which includes the heart and an extensive network of blood vessels, is essential for overall human health. Cardiovascular diseases (CVDs) refer to a group of disorders that affect the heart and blood vessels, often resulting from impaired blood circulation. These include conditions such as coronary artery disease, heart failure, and stroke, which collectively contribute to high rates of morbidity and mortality worldwide [1].

According to the World Health Organization (WHO), CVDs are responsible for approximately 17.9 million deaths annually, with many occurring prematurely in individuals below 70 years of age [2]. Common symptoms include chest pain, shortness of breath, fatigue, and irregular heartbeat, however, they are often not recognized until the disease reaches an advanced stage. Major risk factors include smoking, hypertension, diabetes, obesity, high cholesterol, physical inactivity, aging, and genetic predisposition. Many of these risk factors can be reduced through appropriate lifestyle modifications [3].

In India, the prevalence of CVDs has increased significantly due to urbanization, unhealthy dietary habits, and sedentary lifestyles. This rising trend has placed a substantial burden on the healthcare system, highlighting the need for effective strategies for early detection and prevention. As illustrated in Fig. 1, heart attack-related deaths in India increased steadily between 2018 and 2022, emphasizing the need for advanced predictive models and intelligent healthcare solutions to enable timely diagnosis and reduce mortality rates [4].

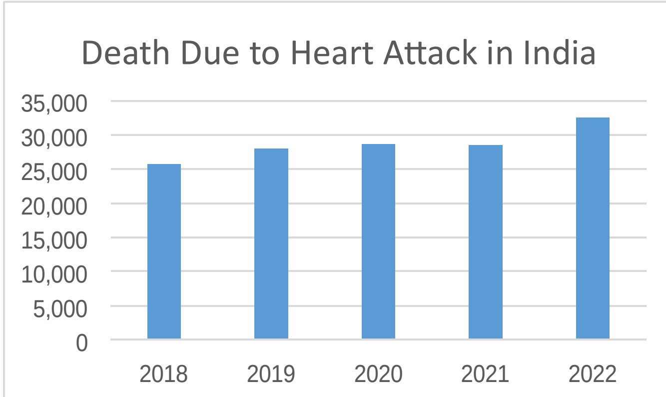


Fig. 1. Heart Attack Deaths in India (2018–2022)

Traditional diagnostic methods for cardiovascular diseases include clinical assessments, electrocardiograms (ECG), echocardiography, magnetic resonance imaging (MRI), and laboratory tests. In addition, treatment methods such as pharmacological therapy, angioplasty, coronary artery bypass grafting, and implantable devices like pacemakers are widely used [5]. However, these approaches are often resource-intensive, expensive, and require specialized expertise, limiting their accessibility in large-scale healthcare settings.

Despite these advancements, early and accurate prediction of cardiovascular diseases remains a significant challenge. Existing predictive models frequently suffer from issues such as class imbalance, overfitting, and limited generalizability across diverse populations. In addition, a strong focus on improving accuracy often leads to reduced attention to interpretability, usability, and real-time applicability. Furthermore, the lack of integration of advanced analytical techniques, such as survival analysis, and the absence of interactive visualization tools limit effective clinical interpretation and realworld implementation [6].

The objective of this study is to develop an Artificial Intelligence-driven machine learning framework for the early and accurate prediction of cardiovascular diseases (CVDs) by evaluating multiple algorithms to identify the most reliable model. The proposed approach enhances data quality and model performance through preprocessing techniques such as data balancing, normalization, and feature selection, while improving interpretability using Explainable Artificial Intelligence (XAI) methods [7]. The framework also integrates data visualization and business

intelligence tools, such as Power BI, to generate realtime insights through interactive dashboards for clinicians and data analysts. Overall, the proposed system aims to provide a scalable, interpretable, and user-oriented decision-support solution for improving cardiovascular disease prediction and supporting clinical decision-making.

II. RELATED WORK

This section reviews existing research on the early prediction of cardiovascular diseases using machine learning and deep learning techniques. Recent studies focus on developing intelligent decisionsupport systems using structured clinical datasets to identify complex patterns beyond traditional methods. The survey further examines the datasets, algorithms, and preprocessing techniques employed to enhance model performance and reliability.

The study by Ogunpola et al. (2024), titled “Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases,” utilized datasets from Mendeley and the Cleveland dataset from Kaggle. The study applied preprocessing techniques such as missing value handling, normalization, and feature scaling, along with exploratory data analysis (EDA). Multiple algorithms, including KNN, SVM, Logistic Regression, CNN, Gradient Boosting, XGBoost, and Random Forest, were implemented. Model performance was evaluated using accuracy, precision, recall, and F1-score, with ensemble methods demonstrating superior performance [8].

Similarly, Bhatt et al. (2023), in “Effective Heart

Disease Prediction Using Machine Learning Techniques,” utilized the Kaggle cardiovascular dataset. The study performed data preprocessing and EDA to clean and analyze the dataset before applying classification algorithms such as K-Nearest Neighbors, Naïve Bayes, Random Forest, Decision Tree, Multi-Layer Perceptron (MLP), and XGBoost. Model evaluation was conducted primarily using accuracy, and the study highlighted that ensemblebased models consistently achieve higher predictive performance compared to individual classifiers [9].

In the study by Sree et al. (2023), titled "An Analysis of Heart Disease Prediction using Machine Learning," the Heart Disease dataset was used. Preprocessing techniques such as normalization and missing value imputation were applied, along with exploratory data analysis. The performance of Decision Tree, SVM, Logistic Regression, KNN, Random Forest, and Neural Networks was evaluated using metrics including accuracy, precision, recall, and F1-score, with SVM demonstrating stable and competitive performance [10].

Authors	Year	Dataset Used	Algorithms used	Results
Adedayo Ogunpola	2024	Cardiovascular Heart Disease Dataset - Mendeley database	KNN, SVM, LR, CNN, GB, XGB, RF	Accuracy: 98.50%, Precision: 99.14%, Recall: 98.29%, F1 score:98.71%
Chintan M. Bhatt	2023	Kaggle Cardiovascular Dataset	RF, DT, MLP, XGB	Accuracy: 87.28%
Saladi Novya Sree	2023	Heart Disease Dataset	KNN, DT, RF, SVM, LR, NN	Accuracy:82%, Precision: 82%, Recall: 80%, F1 score:81%
Niloy Biswas	2023	UCI Cleveland Dataset	LR, SVM, KNN, RF, NB, DT	Accuracy: 94.51%, Sensitivity: 94.87%,Specificity: 94.23%
Zeinab Noroozi	2023	Cleveland Dataset	Bayes net, MLM, SVM, Logit Boost, j48, RF	Accuracy: 85.5%

Table 1. Summary of Related Work on Heart Disease Prediction

Furthermore, the study by Biswas et al. (2023), titled "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," utilized the UCI Cleveland dataset. The study emphasized feature selection methods such as Chi-square and correlation-based techniques to reduce dimensionality. The results indicated that effective feature selection improves model accuracy while reducing computational complexity [11].

Additionally, the study by Noroozi et al. (2023), titled "Analyzing the Impact of Feature Selection Methods on Machine Learning Algorithms for Heart Disease Prediction," utilized the Cleveland dataset. The study applied preprocessing techniques and exploratory data analysis (EDA) before implementing feature selection methods such as Mutual Information, ANOVA, and Recursive Feature Elimination (RFE). Machine learning models including Bayesian Networks, Multilayer Models, SVM, LogitBoost, J48, and Random Forest were evaluated using accuracy as the primary metric. The findings demonstrated that feature selection enhances classification accuracy, reduces overfitting, and improves model generalization while lowering computational cost [12].

III. CLASSIFICATION TECHNIQUES FOR PREDICTION

The classification techniques of machine learning have been widely used for predicting cardiovascular disease (CVD) across various datasets. This section aims to discuss existing approaches and related studies on machine learning-based classification methods for prediction and evaluates multiple classifiers to identify important patterns that improve the accuracy and effectiveness of CVD prediction.

1) Logistic Regression (LR): Logistic Regression is a widely used linear classification technique for cardiovascular disease prediction. In this study, LR achieved an accuracy of 86%, indicating moderate performance in distinguishing between healthy and diseased cases. However, its linear nature limits its ability to capture complex nonlinear relationships in medical datasets [13].
2) Random Forest (RF): Random Forest is an ensemble learning technique based on multiple decision trees. It achieved an accuracy of 84%, demonstrating strong predictive capability due to its ability to reduce over-fitting and handle feature interactions effectively [14].
3) K-Nearest Neighbor (KNN): KNN is a distance-based classification algorithm used for CVD prediction. In this study, it achieved an accuracy of 85%, showing strong performance in pattern recognition when similar instances exist in the dataset [15].

4) Decision Tree (DT): Decision Tree is a simple yet interpretable classification technique. It achieved an accuracy of 87%, indicating good performance in identifying decision boundaries, although it may suffer from over-fitting in complex datasets [16].
5) Naïve Bayes (NB): Naïve Bayes is a probabilistic classifier based on Bayes' theorem. It achieved an accuracy of 83%, showing comparatively lower performance due to its strong assumption of feature independence, which may not hold in medical datasets [17].
6) Multi-Layer Perceptron (MLP): MLP is a feedforward artificial neural network capable of learning nonlinear relationships. It achieved an accuracy of 93%, demonstrating strong predictive capability in capturing complex feature interactions in cardiovascular data [18].
7) XGBoost: XGBoost is an advanced gradient boosting ensemble technique known for efficiency. It achieved 91% accuracy, indicating strong robustness and improved generalization through boosting optimization [19].
8) Support Vector Machine (SVM): SVM is a powerful classification algorithm effective in highdimensional spaces. It achieved the highest accuracy of 95%, demonstrating superior capability in separating complex decision boundaries in cardiovascular datasets [20].

Machine learning classification techniques are widely used for predicting cardiovascular disease (CVD), and the classifiers evaluated in this study demonstrate promising results in identifying CVD risk patterns. Among the models, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR) exhibit strong classification performance, while ensemble and advanced methods such as Random Forest (RF), XGBoost, and Multi-Layer Perceptron (MLP) further enhance predictive capability. Decision Tree (DT) and Naïve Bayes (NB) also contribute to classification with moderate effectiveness. Overall, ensemble and advanced approaches outperform traditional models in terms of accuracy and robustness. These findings indicate that integrating multiple machine learning techniques can further improve prediction accuracy and enhance the reliability and effectiveness of cardiovascular disease diagnosis [21].

IV. METHODOLOGY

A. *Dataset information*

The dataset utilized in this study was obtained from the Kaggle platform and the UCI Machine Learning Repository. It comprises 1,025 patient records with 14 distinct attributes representing various clinical and diagnostic features relevant to cardiovascular disease prediction. The target variable, labeled "target," indicates the presence of heart disease, where a value of 1 represents the presence of cardiovascular disease and 0 represents its absence. This dataset is widely used for binary classification tasks in the healthcare domain, particularly for predictive analysis of heart disease. The dataset is structured and contains no missing values, ensuring data completeness and reliability for further analysis and model development. A summary of the dataset characteristics is presented in Table 2.

B. *Statistical Summary Of Dataset*

A statistical analysis was conducted to understand the structure and distribution of the dataset prior to model development. Initial exploration involved inspecting sample records to verify data integrity and ensure correct attribute loading. Descriptive statistics, including mean, median, standard deviation, minimum, and maximum, were computed for numerical variables, while categorical variables were analyzed using frequency distributions.

Fig. 2. Distribution of Target Variable

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 VI June 2026- Available at www.ijraset.com

The target variable exhibited a binary distribution, where 1 indicates the presence and 0 indicates the absence of heart disease. As shown in Fig. 2, the dataset contains 499 instances of class 0 and 526 instances of class 1, indicating a relatively balanced distribution. Missing value analysis confirmed the absence of null values, ensuring data completeness and suitability for further preprocessing and modeling.

C. Exploratory Data Analysis

In this research, Exploratory Data Analysis (EDA) was performed to understand the dataset characteristics. Univariate analysis examined the distribution of numerical and categorical variables, while bivariate analysis explored relationships between features and the target variable. These analyses helped identify patterns, feature relevance, and potential inconsistencies. Correlation analysis was also conducted to evaluate relationships among variables. Furthermore, Kaplan–Meier survival analysis was performed using age as a surrogate time variable. As shown in Fig. 3, the survival curve indicates a gradual decline in event-free probability with increasing age, with a more pronounced decrease after middle age, highlighting the elevated risk of cardiovascular disease in older individuals.

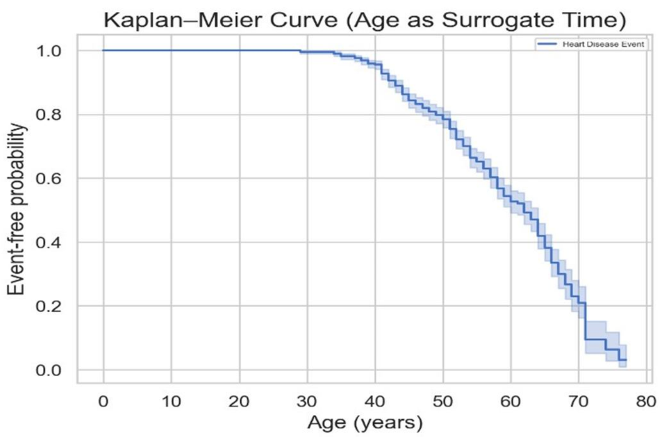


Fig. 3. Kaplan Meier Curve Survival Analysis

Table 2. Heart Disease Dataset Description

Sl no.	Feature name	Feature id	Description	Value type
1	Age	age	Age of the patient in years	Number of years
2	Gender	sex	Indicates patient gender	Male = 0, Female = 1
3	Chest Pain Type	cp	Type of chest pain experienced	Typical angina = 0, Atypical angina = 1, Non-anginal pain = 2, Asymptomatic = 3
4	Resting Blood Pressure	trestbps	Blood pressure measured at rest	Hg
5	Cholesterol	chol	Evaluation of a patient's cholesterol levels	mg/dl
6	Fasting Blood Sugar Levels	fbs	Blood sugar level after fasting	< 120 mg/dl = 0 , > 120 mg/dl = 1
7	Resting Electrocardiographic Results	restecg	Electrocardiographic results at rest	= Normal, 1 = Having ST-T wave abnormality, 2= definite left ventricular hypertrophy
8	Maximum heart rate	thalach	Maximum heart rate achieved during testing	Continuous
9	Exercise Induced Angina	exang	Presence of angina during exercise	= Yes , 0 = No
10	ST Depression	oldpeak	ST depression induced by exercise relative to rest	Continuous
11	Slope	slope	Peak exercise ST segment slope	= Upsloping, 1 = Flat, 2 = Downsloping
12	Number of Major Vessels	ca	Count of main vessels colored by fluroscopy	0-4
13	Thalassemia	thal	Thallium stress	= Normal, 1 = Fixed defect, 2 = Reversible defect, 3 = Not described
14	Heart Disease Status	target	diagnosis of heart disease using the angiographic disease status	= Presence of Heart disease, 0 = No Heart Disease

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 VI June 2026- Available at www.ijraset.com

D. Data Preprocessing

Data preprocessing was performed to enhance data quality and ensure compatibility with machine learning models. The dataset was examined for missing values and inconsistencies, and necessary corrections were applied. Categorical variables were converted into numerical form using one-hot encoding. Feature scaling was performed using StandardScaler to normalize the data to zero mean and unit variance. Additionally, the Box–Cox transformation was applied to selected numerical features to reduce skewness and improve data distribution.

E. Model Development And Training

Model development and training were performed using supervised machine learning techniques to predict the presence of heart disease. The dataset was divided into training and testing sets, and the model was trained to learn the relationship between input features and the target variable. Baseline performance was established using default hyperparameters. Hyperparameter optimization was performed using a grid search approach to identify the optimal configuration. The model was then retrained using the optimal hyperparameters, improving predictive performance and generalization. Finally, the optimized model was evaluated on the test dataset to assess its effectiveness on unseen data.

F. Model Evaluation

Model evaluation was conducted to assess the performance and generalization ability of the developed model. The trained model was tested on unseen data using metrics such as accuracy, precision, recall, and F1-score. A confusion matrix was used to analyze classification outcomes, including true positives, true negatives, false positives, and false negatives. These measures provide a comprehensive assessment of the model's effectiveness in heart disease prediction.

G. Deployment

A simple deployment setup was developed to interact with the trained model and validate its predictions. The interface was implemented using HTML and ipywidgets, enabling user input of feature values and real-time prediction generation. This setup was primarily designed for testing and validation purposes, facilitating the verification of the model's predictive performance.

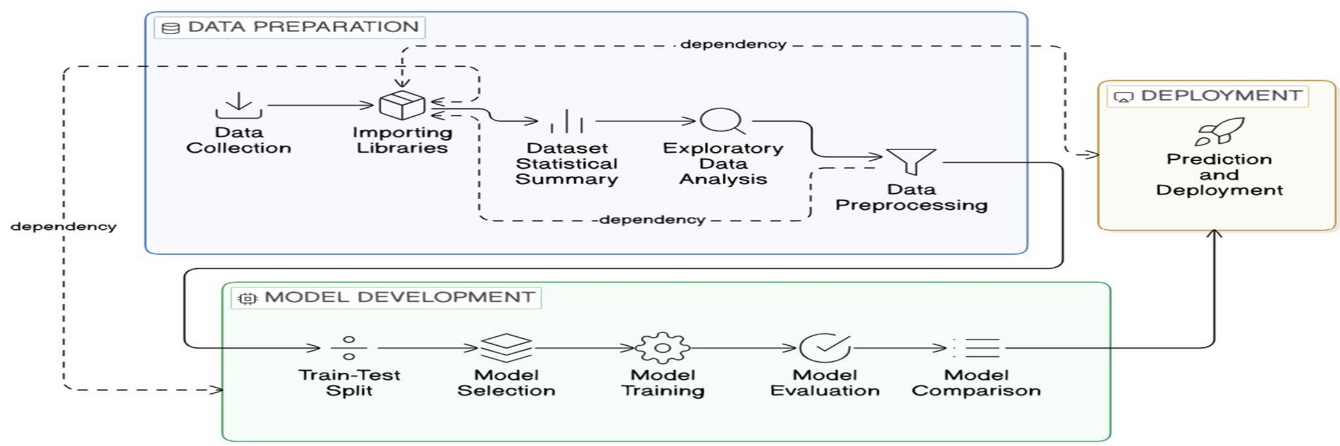


Fig. 4. Proposed System Architecture

H. Data Transformation

Microsoft Excel was used for data preparation and transformation prior to developing the Power BI dashboards. The dataset was reviewed to ensure structural consistency, proper attribute formatting, and the removal of duplicate entries, thereby improving data quality for visualization. As part of feature engineering, a Patient ID was included for the unique identification of records, and age binning was performed to convert continuous age values into categorical groups for better visualization of agewise trends. Additionally, selected numerical attributes were transformed into categorical variables using Excel formulas and conditional logic (IF statements) based on predefined threshold values. These transformations improved interpretability and enabled meaningful data grouping and segmentation. Excel functions such as filtering, sorting, and data validation were also applied to refine the dataset. The final dataset was structured for seamless integration with Power BI.

I. Data Visualization Using Power-Bi

Interactive dashboards were developed using Microsoft Power BI to enhance data visualization and interpretability. The preprocessed dataset was used to create visualizations such as bar charts, line graphs, and distribution plots to represent relationships among key variables, including age, gender, chest pain type, cholesterol, resting blood pressure, ST depression, maximum heart rate (thalach), and the number of major vessels (ca).

Two dashboards were developed: a clinical dashboard for patient monitoring and an analytical dashboard for feature exploration and relationship analysis. Interactive filters and slicers were included to enable dynamic exploration based on parameters such as gender, blood sugar level, and ECG results.

Overall, the dashboards provide an effective visualization layer that complements the machine learning framework and enhances understanding of the dataset.

V. EXPERIMENTS AND RESULTS

This section presents the evaluation of machine learning models for heart disease prediction using a clinical dataset, with the objective of identifying the most effective model based on performance metrics and feature analysis. Exploratory Data Analysis (EDA), including statistical analysis, visualization, and Kaplan–Meier survival curves, was performed to extract insights. Data preprocessing techniques such as scaling, encoding, feature transformation, and Box–Cox transformation were applied to improve data quality and model performance [22].

Additionally, interactive Power BI dashboards were developed to enhance interpretability. A clinical dashboard supports patient monitoring and risk identification, while an analytical dashboard enables exploration of feature relationships and patterns. Business intelligence tools such as Power BI are widely used in healthcare analytics for visualization and decision support [23].

A. Exploratory Data Analysis Results

To effectively analyze the heart disease dataset for machine learning applications, Exploratory Data Analysis (EDA) was performed to understand feature distributions, relationships, and their influence on the target variable. Both numerical and categorical features were examined. Continuous variables such as age, trestbps, chol, and thalach showed normal or moderately skewed distributions, as shown in Fig. 5, while oldpeak was right-skewed, indicating the need for normalization and scaling. Patients with heart disease generally showed higher thalach and lower oldpeak values, while age, cholesterol, and blood pressure exhibited moderate variation across classes.

Categorical analysis showed that chest pain type (cp) categories 1–3 were more associated with heart disease, while most patients had normal fasting blood sugar (fbs) and no exercise-induced angina (exang). The dataset also had a higher proportion of male patients and a relatively balanced target variable. Correlation analysis using a heatmap, as shown in Fig. 6 , indicated that thalach was positively correlated with the target, whereas oldpeak and exang were negatively correlated. Cp showed moderate positive correlation, while chol and fbs had weak influence. Overall, EDA helped identify key predictive features such as cp, thalach, exang, and oldpeak, supporting effective preprocessing and improving model performance.

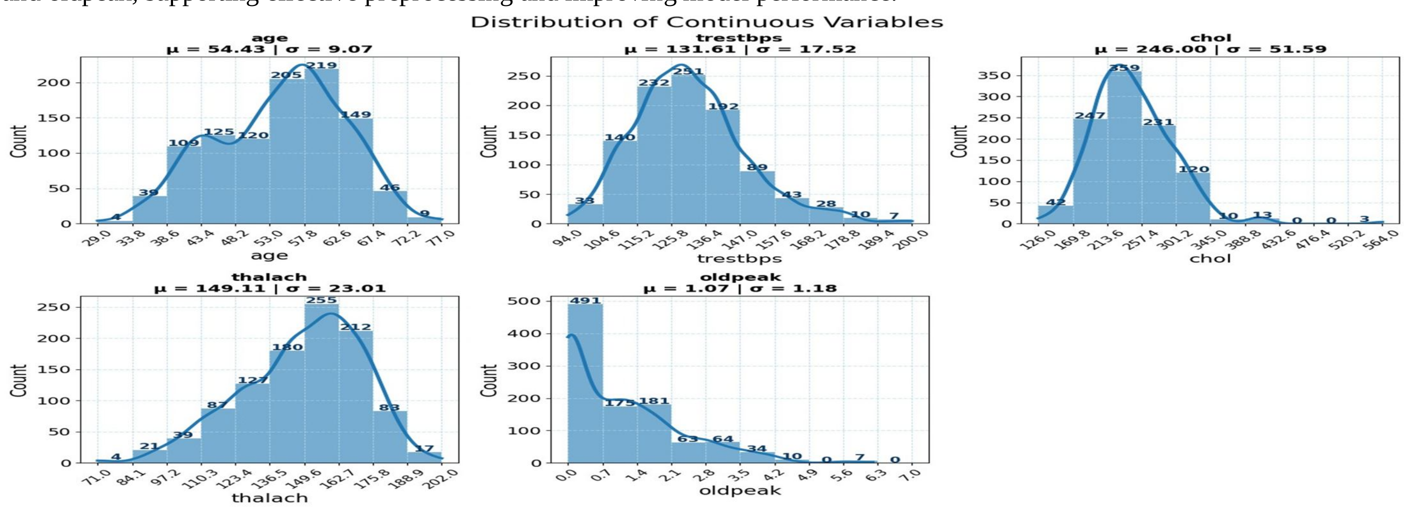


Fig. 5. Distribution of Continuous Variables

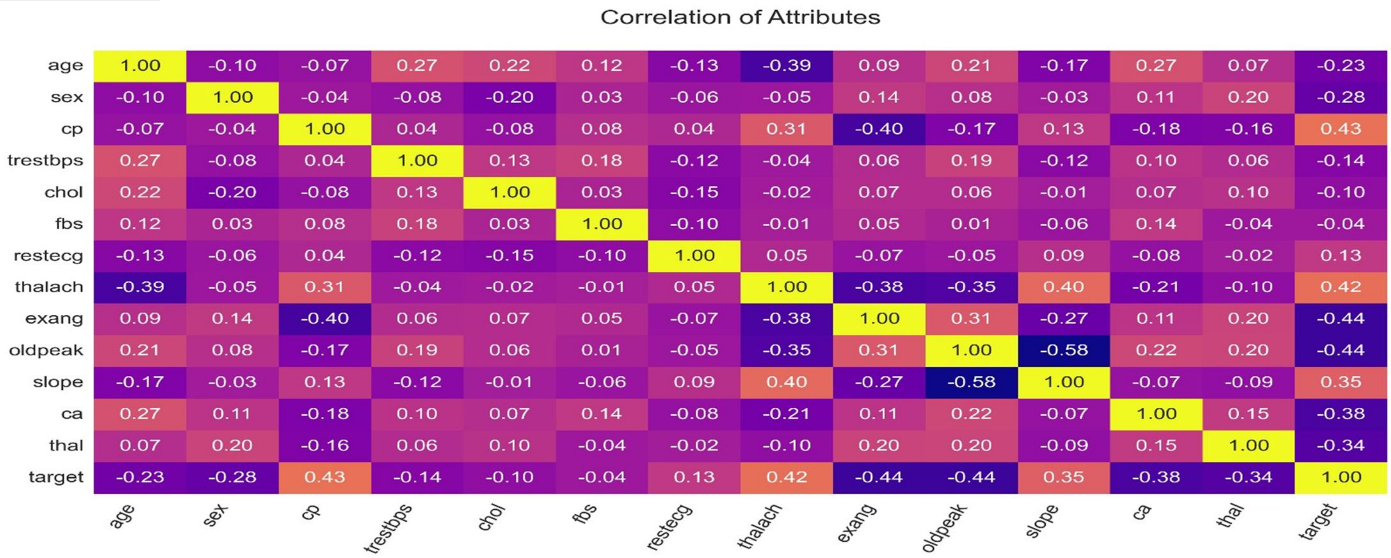


Fig. 6. Feature Correlation Heatmap

B. Data preprocessing results

Data preprocessing was performed to improve the applied to continuous variables such as age, trestbps, quality and suitability of the dataset for machine chol, thalach, and oldpeak to reduce skewness and normalize their distributions. As shown in Fig. 7, the learning models. Transformation techniques were original variables exhibited varying degrees of skewness, with oldpeak showing strong right skewness and trestbps and chol showing moderate irregularities. After transformation, the distributions became more symmetric and well-balanced, with improved spread across features. Age and thalach remained relatively stable with minor refinements, contributing to improved feature consistency and scaling. Overall, these preprocessing steps reduced the impact of outliers and enhanced data distribution, thereby improving model performance.

C. Performance Evaluation Results

In this section, the performance of the developed machine learning models is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics are computed from the confusion matrix, which comprises true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), providing a comprehensive assessment of model performance. balanced precision, recall, and F1-score of approximately 0.95.

A class-wise evaluation shows consistent performance across both classes, indicating no significant bias. MLP and XGBoost also performed well with accuracies of 93% and 91%, respectively, while Decision Tree, Random Forest, and Logistic Regression showed moderate performance (84– 87%). KNN and Naïve Bayes recorded lower

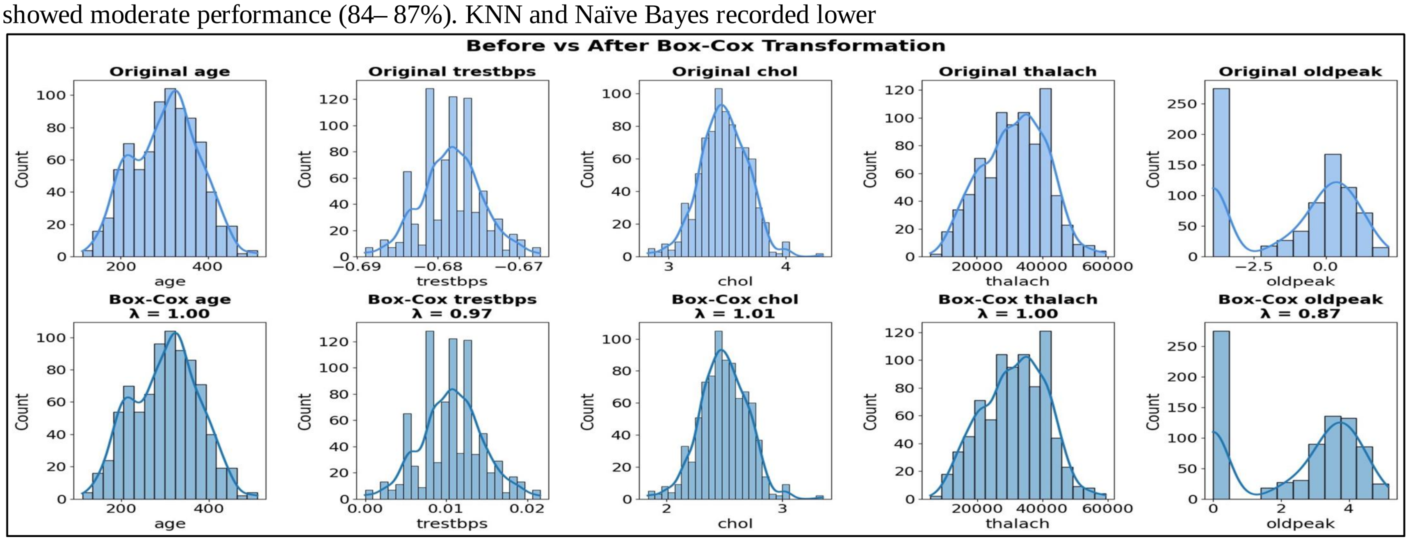


Fig. 7. Box-cox Transformation Analysis

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 VI June 2026- Available at www.ijraset.com

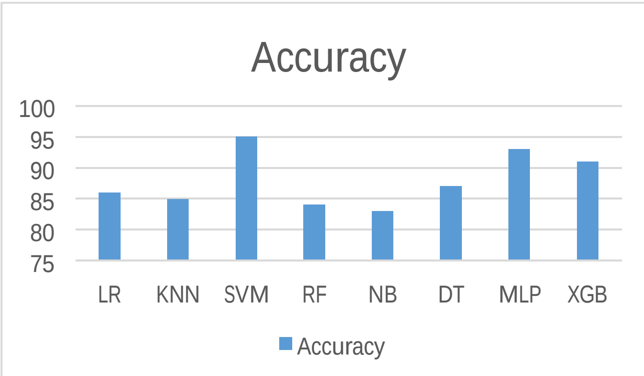


Fig. 8. Comparative Analysis of Models

The comparative performane is also illustrated in Fig. 8, highlight ing accuracy differences across models. Among all models, SVM achieved the best performance with an accuracy of 95%, along with

accuracy of around 83%.

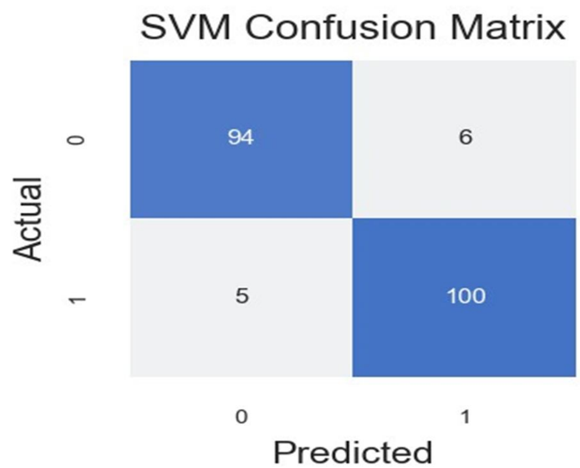


Fig. 9. Confusion Matrix of SVM

ML Algorithms	precision_0	precision_1	recall_0	recall_1	f1_0	f1_1	accuracy
Logistic Regression	0.87	0.85	0.84	0.88	0.85	0.86	0.86
K-Nearest Neighbour	0.85	0.85	0.84	0.86	0.84	0.85	0.85
Support Vector Machine	0.95	0.95	0.95	0.95	0.94	0.95	0.95
Random Forest	0.89	0.81	0.78	0.90	0.83	0.86	0.84
Naïve Bayes	0.81	0.86	0.0.86	0.81	0.83	0.83	0.83
Decision Tree	0.89	0.86	0.85	0.90	0.87	0.88	0.87
Multi Layer Perceptron	0.94	0.92	0.91	0.94	0.92	0.93	0.93
XGBoost	0.90	0.91	0.91	0.90	0.91	0.91	0.91

Table 3. Comparative Summary of Machine Learning Models

minimal error, which is critical in medical diagnosis where false negatives must be minimized. The ROC curve in Fig. 10 shows excellent class separability with an AUC of 0.99, confirming strong discriminative performance.

Overall, SVM outperforms other models in terms of accuracy and stability, making it the most suitable model for heart disease prediction in this study.

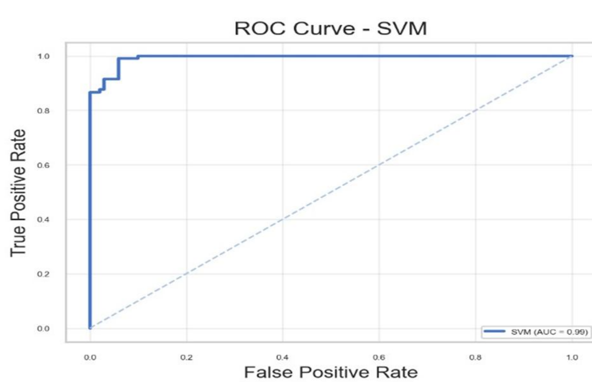


Fig. 10. AUC-ROC Curve of SVM

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 VI June 2026- Available at www.ijraset.com

D. *Deployment Results*

The final phase of this study involves the deployment of the developed heart disease prediction model using an interactive interface built with Python's ipywidgets library, demonstrating its practical applicability in a real-time clinical decision support system.

The Table 3 presents the comparative results of various models, including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors, Naïve Bayes, Multi-Layer Perceptron, and XGBoost. The confusion matrix of the SVM model, illustrated in Fig. 9 further validates its robustness, with 100 true positives and 94 true negatives, along with only 6 false positives and 5 false negatives.

This reflects the model's ability to correctly classify both positive and negative cases with

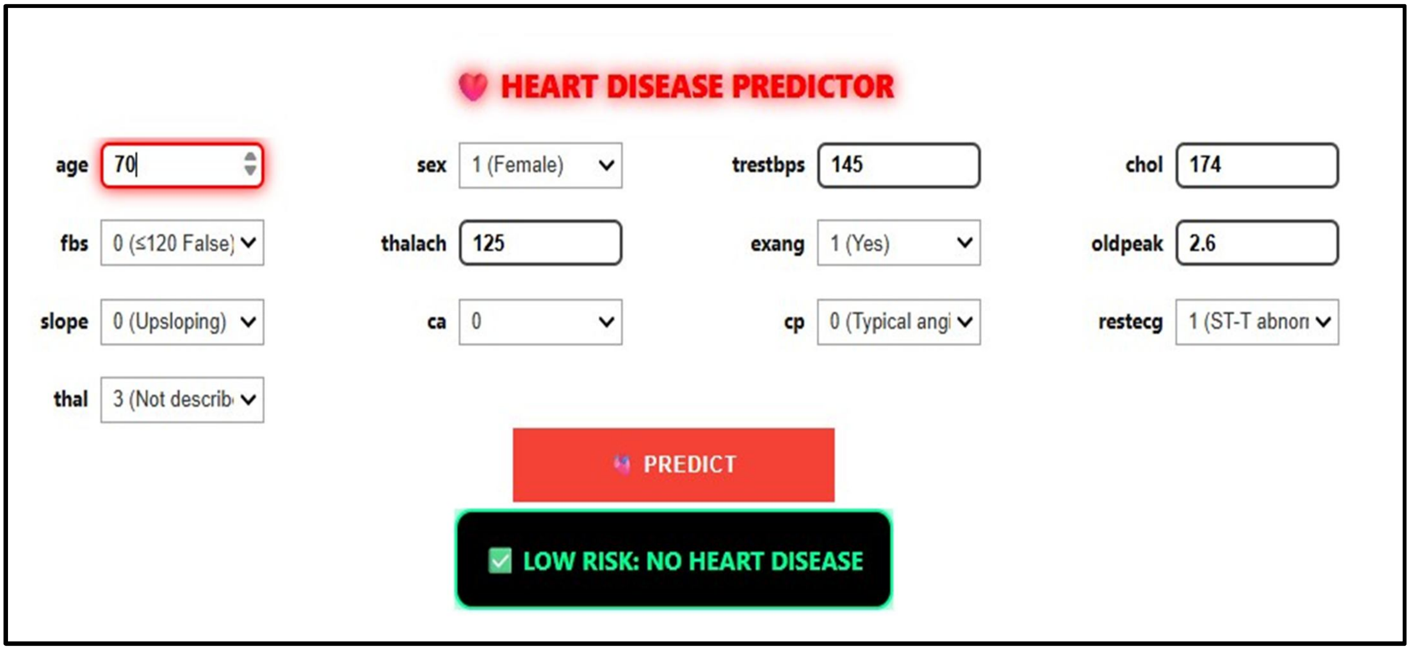


Fig. 11. Deployment of Heart Disease Prediction System

The system accepts clinically relevant inputs such as age, sex, trestbps, chol, fbs, thalach, exang, oldpeak, slope, ca, cp, restecg, and thal, which were identified as important features during exploratory data analysis. Inputs are provided through structured fields and dropdown menus, while categorical variables are internally encoded to match the trained model.

Upon submission, the model generates real-time predictions, classifying patients as high risk (heart disease detected) or low risk (no heart disease). The results are displayed using a color-coded interface for easy interpretation, with red indicating high risk and green indicating low risk. The Fig. 11 illustrates the deployed prediction interface.

Overall, the deployment confirms that the proposed model can be effectively integrated into a clinical decision support system for real-time heart disease prediction.

E. *Power-Bi Dashboard Results*

To enhance interpretability, interactive dashboards were developed using Microsoft Power BI, providing visual insights into feature relationships from both clinical and analytical perspectives. The clinical analysis dashboard, shown in Fig. 12, highlights key risk patterns, indicating higher heart disease prevalence in the 50–70 age group, while younger individuals show lower occurrence. Analysis of chest pain type and exerciseinduced angina suggests a stronger association of asymptomatic and atypical cases with heart disease. Higher ST depression values, particularly in flat and downsloping categories, are strongly linked to disease presence, along with variations in cholesterol, thalach, and resting blood pressure.

The population-level analytics dashboard, shown in Fig. 13, provides an overview of feature distributions and associations. It shows a higher proportion of heart disease cases among male patients, while an increased number of major vessels (ca) is strongly associated with higher risk. ST depression also shows elevated values in middleaged individuals. The target distribution remains relatively balanced, supporting reliable model evaluation.

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 VI June 2026- Available at www.ijraset.com

Overall, the Power BI dashboards reinforce the importance of key predictors such as age, chest pain type, ST depression, vessel count, and exerciseinduced angina, complementing the machine learning results and improving interpretability.

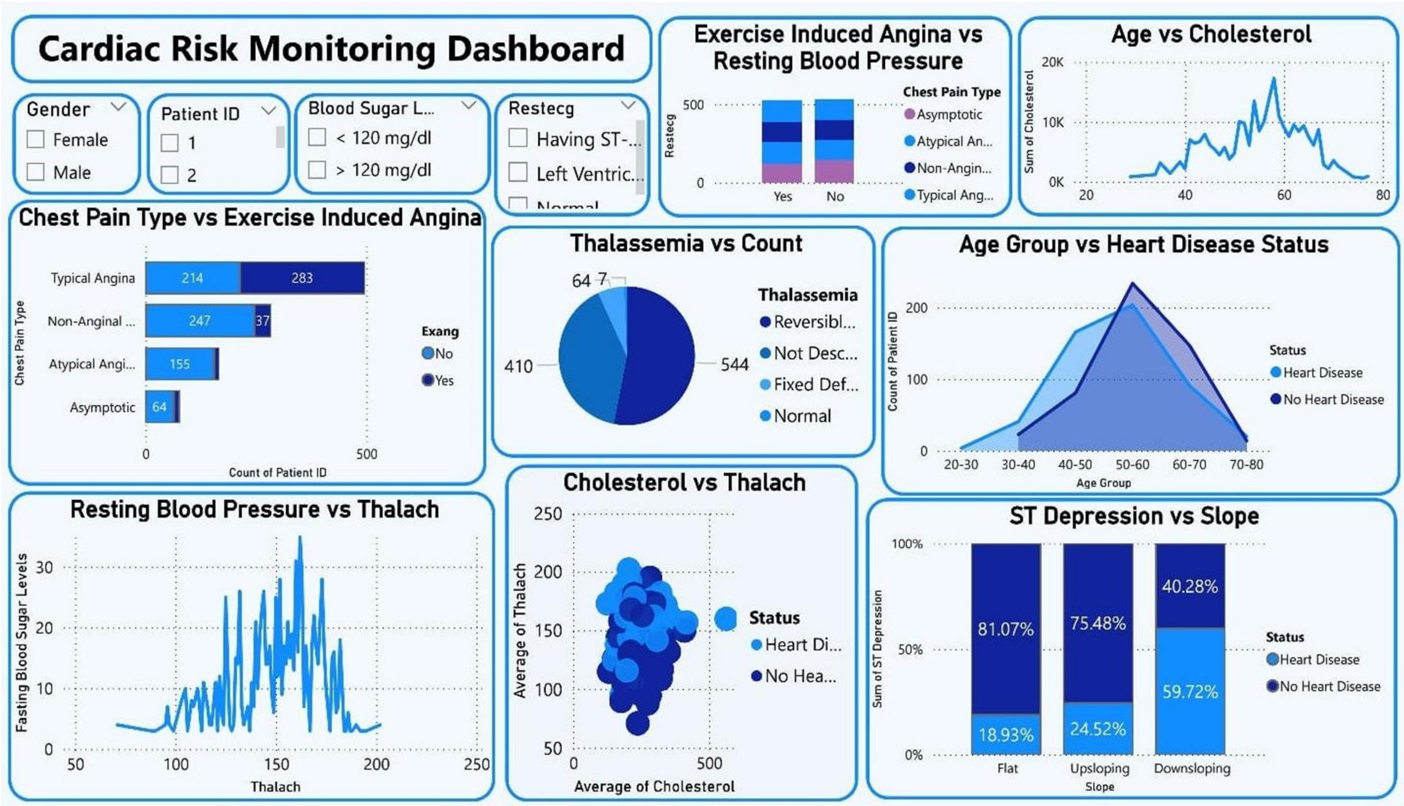


Fig. 12. Cardiac Risk Monitoring Dashboard

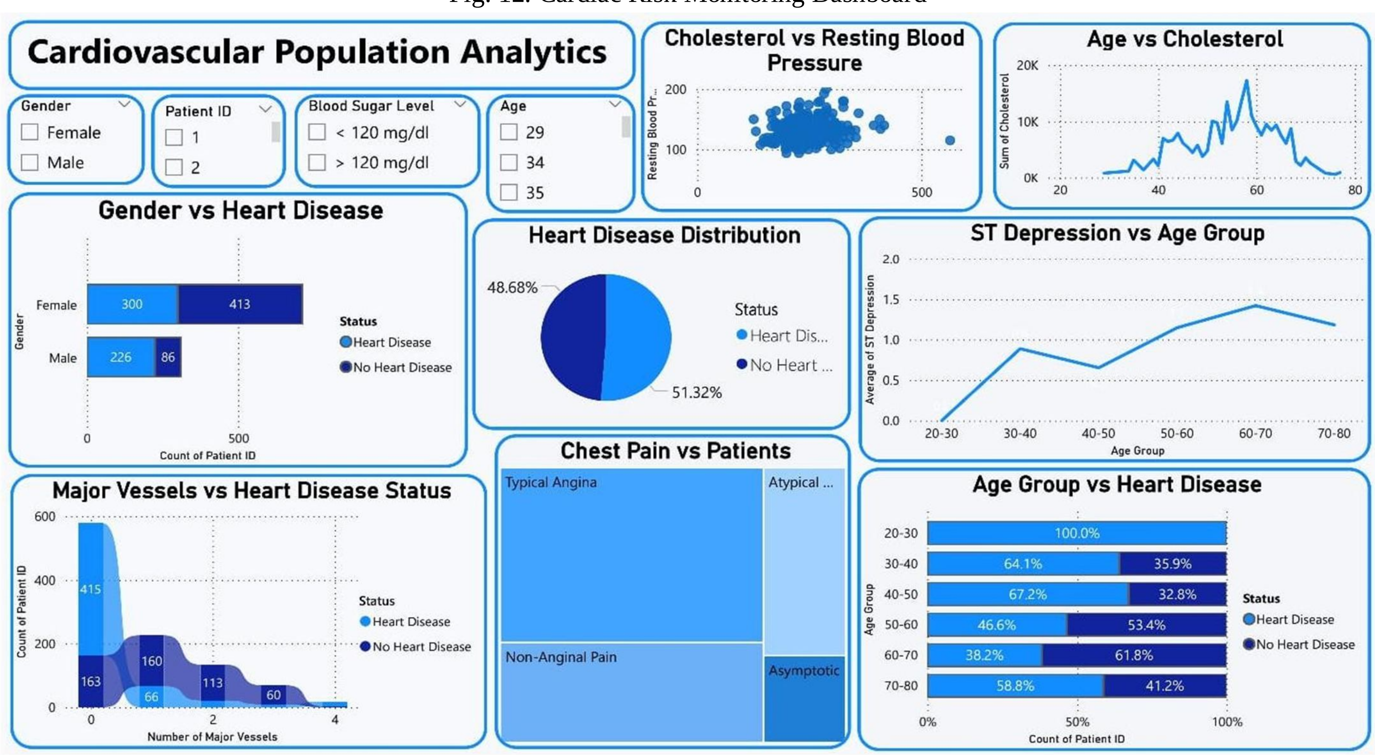


Fig. 13. Cardiovascular Population Analytics Dashboard

VI. CONCLUSION

The proposed architecture overcomes the limitations of existing heart disease prediction systems by integrating effective preprocessing, advanced machine learning techniques, and interactive visualization to improve accuracy, interpretability, and usability. Additionally, survival analysis indicates a decline in survival probability with increasing age, emphasizing the importance of early diagnosis. Overall, the proposed system provides an accurate and interpretable framework for heart disease prediction, making it suitable for clinical decision support applications.

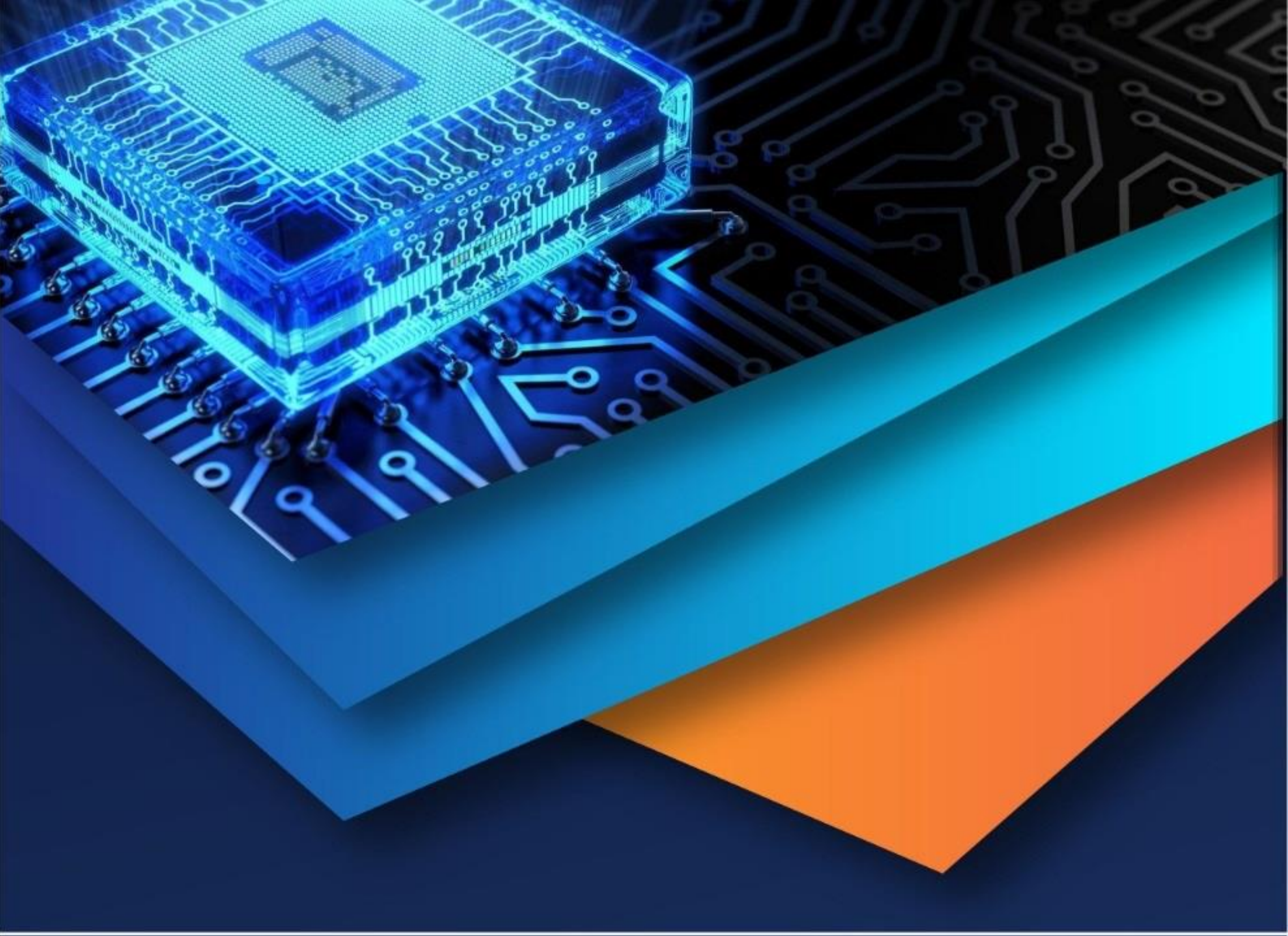
The system executes trained machine learning models on clinical data in a structured and optimized manner. Since the data is processed through a welldefined pipeline, inconsistencies are minimized and prediction reliability is improved. The use of visualization tools such as heatmaps, confusion matrices, and ROC curves ensures transparency in model evaluation, enabling better interpretability for clinical decision-making. The architecture also ensures that data is handled in a controlled and structured manner, maintaining consistency throughout the prediction process.

Further, multiple machine learning algorithms are evaluated to enable comparative analysis and the selection of the best-performing model, thereby improving classification accuracy and robustness across varying data patterns. Preprocessing techniques such as scaling and Box–Cox transformation enhance data distribution and reduce the impact of outliers, resulting in more stable predictions. In addition, interactive Power BI dashboards provide real-time visual insights into risk factors and disease patterns, improving usability and decision support.

It is concluded that high accuracy, reliability, and interpretability are achieved through optimized model performance and effective feature analysis. Thus, the proposed system fulfills all objectives of the study and overcomes the limitations of existing approaches. Future enhancements include deploying the system as a cloud-based application to enable wider accessibility and real-time usage across healthcare platforms.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO Fact Sheets, 2023.
[2] World Health Organization, Global Health Estimates: Leading Causes of Death, WHO, Geneva, 2022.
[3] Centers for Disease Control and Prevention (CDC), "Heart Disease Facts," CDC, 2023.
[4] Ministry of Health and Family Welfare,
[5] Government of India, "National Health Profile and CVD Statistics," 2022.
[6] American Heart Association, "Cardiovascular Disease Diagnosis and Treatment Guidelines," Circulation Journal, 2021.
[7] D. K. Choubey et al., "Challenges in Machine Learning-Based Cardiovascular Disease Prediction: A Review," Computational Intelligence and Neuroscience, 2022.
[8] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions (SHAP)," Advances in Neural Information Processing Systems, 2017.
[9] A. Ogunpola et al., "Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases," Mendeley Data / Kaggle Dataset, 2024.
[10] C. M. Bhatt et al., "Effective Heart Disease Prediction Using Machine Learning Techniques," Kaggle Cardiovascular Dataset Study, 2023.
[11] S. N. Sree et al., "An Analysis of Heart Disease Prediction Using Machine Learning," Heart Disease Dataset Analysis, 2023.
[12] N. Biswas et al., "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," UCI Cleveland Dataset Study, 2023.
[13] Z. Noroozi et al., "Analyzing the Impact of Feature Selection Methods on Machine Learning
[14] Algorithms for Heart Disease Prediction," Cleveland Dataset Study, 2023.
[15] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Wiley, 2013.
[16] L. Breiman, "Random Forests," Machine
[17] Learning, vol. 45, no. 1, pp. 5–32, 2001.
[18] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
[19] J. R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986.
[20] P. M. Domingos and M. J. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," Machine Learning, 1997.
[21] D. E. Rumelhart, G. E. Hinton, and R. J.
[22] Williams, "Learning Representations by BackPropagating Errors," Nature, vol. 323, pp. 533–536, 1986.
[23] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. KDD, 2016, pp. 785–794.
[24] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[25] S. M. Odeh et al., "A Review of Machine Learning Techniques for Cardiovascular Disease Prediction," Journal of Healthcare Engineering, vol. 2022, Article ID 1234567, 2022.
[26] D. R. Cox and D. Oakes, Analysis of Survival Data, Chapman & Hall/CRC, 1984.
[27] S. T. Alexander, "Data Visualization and Business Intelligence in Healthcare Analytics Using
[28] Power BI," International Journal of Computer Applications, vol. 182, no. 35, pp. 1–6, 2019.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY