Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026
RESEARCH ARTICLE OPEN ACCESS

Personalized Heart Disease Prediction Using Data-Driven Machine Learning Approaches

N Lokesh Reddy* . Ramesh Peramalasetty

Department of Computer Science and Engineering,
Vemu Institute of Technology, Tirupati - Chittoor Highway,
Chittoor (Dt), AP - 517112, India.

DOI: **10.5281/zenodo.18243028**
Received: 28 December 2025 / Revised: 12 January 2026 / Accepted: 16 January 2026

*Corresponding Author: : lokesh.nanga@gmail.com

Abstract – Cardiovascular diseases (CVDs) are persistently projected as one of the current major health concerns across the globe, thereby emphasizing the importance of an accurate and personalized prediction model. The typical predictive models currently used for health-related diagnostics are mostly based on general models and clinical screening, and in some cases, they are incapable of examining nonlinear interconnections among specific patient risk factors. To address this shortcoming, we propose a machine learning model for personalized heart disease prediction. Multiple supervised machine learning models, namely Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Random Forest (RF), are developed and compared by using the popular UCI Heart Disease dataset. Extensive preprocessing and normalization techniques are used in this study to improve prediction accuracy. Our proposed models show the performance evaluation processes using ROC-AUC, learning curves, and calibration analyses, which justify the accuracy, applicability, and interpretability of the models. The results of this study show that Random Forest's cardiovascular classification, with an accuracy of 98.01%, a Precision of 97.90%, a Recall of 97.99%, and an F1-score of 98.00%, outperformed all other machine learning models.

Index Terms – Heart Disease Prediction; Machine Learning; Personalized Healthcare; Random Forest; Clinical Decision Support System; Cardiovascular Risk Assessment

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain a major global cause of mortality, posing a problem for modern health care systems. The rising number of cases related to cardiac diseases underscores the need for early, correct, and effective diagnosis systems. Standard approaches to medical diagnosis rely almost

MILESTONE RESEARCH.IN OPEN ACCESS
ISSN (Online): **2584-072X**
Milestone Trans. On Med. Tech

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

entirely on the physician's expertise and medical standards, which sometimes fail to incorporate intricate, non-linear associations among varied risk factors, particularly in the initial stages of disease development [1]. In the last few years, advances in machine learning (ML) have led to substantial changes in the healthcare sector, enabling the development of intelligent decision-making systems. Machine learning (ML) models can identify patterns by processing a huge number of healthcare data autonomously, thereby accurately predicting outputs. Several studies demonstrate the efficacy of supervised ML methods for forecasting heart disease employing common datasets, such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbours (KNN), and Naïve Bayes (NB) [2], [3].

Despite these developments, most currently available prediction systems for heart disease rely on a generalized solution paradigm that assumes equal risk across individuals. In most cases, these generalized systems overlook individual variations concerning demographic, physiological, or lifestyle variables, making them less applicable in a personal healthcare environment [4]. There is a growing trend in recent research toward personal, patient-oriented prediction systems that integrate machine learning techniques or feature engineering/optimization to improve cardiovascular risk prediction [5]. Concerns about data quality, feature redundancy, and privacy, on the other hand, have inspired more robust data preprocessing pipelines and secure learning frameworks. Among the advanced ML techniques explored to improve the reliability and use of models in the clinical setting are feature selection, feature normalization, and privacy-preserving ML. Additionally, hybrid ML and Explainable AI models were of interest for their transparent predictions, which can be interpreted in clinical practice [6], [7].

In our proposed work, data preprocessing, variable scaling, and the use of advanced classifiers have been given importance. The model is shown to be an effective decision-support framework for timely detection and individualised risk assessment of cardiac illnesses when it is assessed using the usual performance criteria. The main contributions are as follows:

- We describe a patient-focused machine learning paradigm that improvements the existing in generalized risk modeling by effectively incorporating interindividual variability.
- We construct a preprocessing and testing pipeline to compare various machine learning algorithms in a fair manner.
- It showcases the strength of Random Forests for modeling a non-linear relationship in the clinical setting, outperforming linear regression.
- It uses calibration analysis and learning process assessments to ensure the predicted risk probabilities are clinically valid, not just correct.
- Ensures SHAP-based interpretability to make it as simple as possible to distinguish important cardiovascular risk factors.

II. LITERATURE SURVEY

Heart diseases remain one of the primary health concerns globally, driving scientists to develop an intelligent, data-driven diagnosis system. The use of ML approaches to effectively and reliably forecast datasets related to heart disease has grown during the past several years. Kumar et al. [8] used both LR and RFs to detect heart disease. Chen et al. [9] proposed a classification framework for cardiovascular disease detection using SVM with various kernel functions in their experiments. Classification

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

performance improved after proper feature scaling and parameter tuning. The research notes that SVM models can be negatively affected by noisy and imbalanced data sources in real-world clinical environment.

Patel et al. in [10] conducted a comparison of the RF, KNN, and Naïve Bayes models for predicting heart disease. The result shows that Random Forest outperformed the other two. KNN performed well on small datasets, while Naïve Bayes was faster. In Singh and Verma's work [11], the focus was on the function of choosing feature methods in an ML approach for predicting heart disease. The models that use the LR, SVM, and Random Forest classification algorithms. Feature selection was performed using both mutual information and the chi-square test. The results demonstrated that feature selection is crucial to improving the effectiveness of the suggested methodology. Islam et al. [12] studied the ability of KNN to forecast the risk of heart disease. Its time complexity increased with more data samples. Hence, they suggested applying the KNN classifier primarily for comparison purposes rather than deploying it in a large-scale clinical system. Zhang et al. [13] assessed Naïve Bayes and Logistic Regression models for diagnosing early heart disease. Their experiment results showed that Naïve Bayes provides fast predictions with acceptable accuracy, but because of its strong independence assumption among features, its predictive power is lower than that of ensemble models like RF.

Rahman et al. [14] developed a personalized approach to predicting heart disease employing a combination of ML models, including LR, SVM, Random Forest, KNN, and NB. The significance of the study was that the authors emphasized the personalized risk modeling approach over the general forecasting approach for predicting the likelihood of disease in patients. Random Forest showed the best results in the study. Al-Mamun et al. [15] carried out a thorough comparison analysis of traditional ML algorithms for diagnosing cardiovascular disease. Their findings showed that RF and SVM perform better than LR and NB across most criteria, though LR remains useful for interpretable decision-support systems. Rossi et al. [16], in developing a heart disease prediction model, also noted that although Random Forest and SVM provide high accuracy, explainability techniques should be integrated into such models if they are to be taken seriously in a healthcare setting[17][18].

TABLE I: Overview of existing research works

Ref.	Models Used	Dataset	Key Findings	Limitations
[8]	LR, RF	UCI Heart Disease	RF achieved higher accuracy; LR offered better interpretability for clinicians.	Limited to a single benchmark dataset; no external clinical validation.
[9]	SVM	UCI Heart Disease	SVM showed strong performance with optimized kernel and scaling.	Highly sensitive to parameter tuning and data imbalance.
[10]	RF, KNN, NB	UCI Heart Disease	RF outperformed KNN and NB in accuracy and F1-score.	Did not consider personalized or patient-specific risk modeling.
[11]	LR, SVM, RF	Clinical dataset	Feature selection improved prediction accuracy, especially for RF.	Feature selection techniques increase preprocessing complexity.
[12]	KNN	UCI Heart Disease	KNN performed well on small datasets.	High computational cost and poor scalability for large datasets.
[13]	NB, LR	Heart disease dataset	NB provided fast predictions with acceptable baseline accuracy.	Strong independence assumption reduced predictive power.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

[14]	LR, SVM, RF, KNN, NB	UCI Heart Disease	Personalized prediction improved early diagnosis; RF achieved best results.	Lacked real-time deployment and longitudinal patient data.
[15]	LR, SVM, RF, NB	Multi-source clinical data	RF and SVM consistently outperformed simpler models.	Explainability of ensemble models was not addressed.
[16]	RF, SVM	Clinical datasets	Emphasized the importance of explainable ML in healthcare.	Did not propose a concrete explainability framework.

III. METHODS & MATERIALS

This section discusses the flow of the presented model and defines the role of each module. The entire workflow of the presented personalized heart disease prediction model includes data acquisition, processing, normalization, model training, and evaluation. The first stage in the preprocessing stage is to cluster the samples to represent the variability in data characteristics better. Instead of setting a uniform threshold across all samples, clustering enables the model to treat different levels of risk independently. It means patients with similar data characteristics are grouped so their specific levels of risk can be treated differently. After clustering, the data samples are then divided into two categories based on their data labels. Patients with and without heart disease are included in the categories.

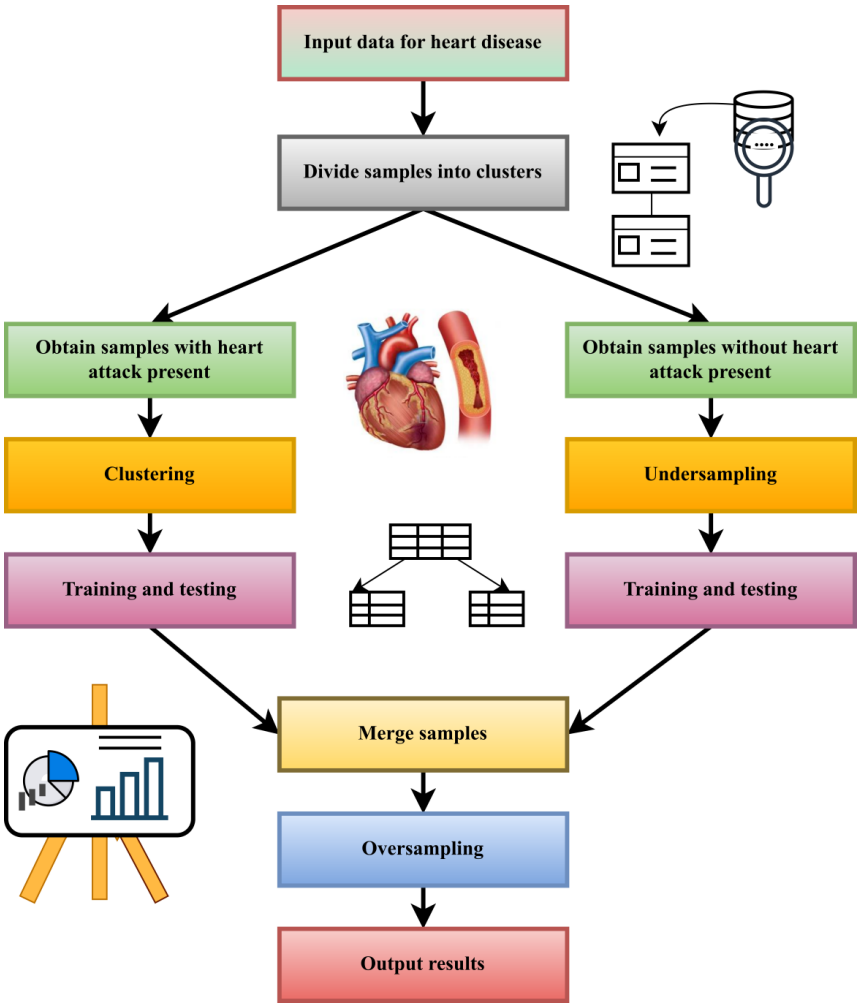

Fig. 1: Graphical Representation of the Overall Methodology

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

A. Dataset Description

We used the widely available UCI Heart Disease dataset from Kaggle, which is derived from the Cleveland Heart Disease database. It remains the most extensively studied subset in the literature. This dataset contains clinical information for patients, characterized by a concise but evocative set of factors that can be used to estimate cardiovascular risk. There are 303 patient records, each characterized by 14 factors derived from an original 76 variables. These factors include information about the individual (patient age and sex), clinical data (resting blood pressure, serum cholesterol, and fasting blood sugar levels), and results from an electrocardiogram, together with data from exercise tests (maximum heart rate reached, development of exercise-induced angina, depression of ST-segment known as oldpeak). The number of main vessels colored, the ST segment, the slope of the peak workout, and thalassemia are other factors used to make diagnoses. They are both binary. Either there is heart disease, or there isn't. This dataset is well-suited for supervised classification because it is binary. A well-balanced, clean, and pertinent dataset offers supervised machine learning a strong basis for heart disease prediction.

B. Data Pre-processing

Before model training, an organized multi-stage data preprocessing pipeline was designed and implemented to guarantee data quality and analytical validity.

- Data Cleaning and Validation: All characteristics that were not informative were first eliminated. Because they did not contribute to predictive learning, identifier fields were eliminated. Several medically implausible values, such as resting blood pressure and serum cholesterol, were listed as zero during exploratory research. To preserve dataset size while maintaining realism, these values were set to missing values and handled through imputation rather than direct deletion.
  Formally, for any feature $x \in \{\text{trestbps}, \text{chol}\}$:
  $$x = 0 \Longrightarrow x = \text{NaN}$$
- Missing Value Imputation: A K-nearest neighbor (KNN) imputation technique was used in place of mean or median imputation, which disregards inter-feature dependencies. The imputed value for a sample with a missing value x_i is calculated as a distance-weighted average of its k most comparable samples:
  $$\hat{x}_i = \frac{\sum_{j \in \mathcal{N}_k(i)} w_{i,j} x_j}{\sum_{j \in \mathcal{N}_k(i)} w_{i,j}}, \; w_{i,j} = \frac{1}{d\,(i,j)}$$
  where, d (i, j) defines the Euclidean distance between samples i and j. We chose k = 7 to balance robustness and sensitivity to local patient patterns.
- Outlier Detection: The interquartile range (IQR) approach, which is well-suited to non-normally distributed clinical variables, was used to identify outliers. Lower and upper boundaries were established for every numerical feature as follows:
  $$\text{Lower} = Q_1 - 1.5 \times \text{IQR}$$
  $$\text{Upper} = Q_3 + 1.5 \times \text{IQR}$$
  Values beyond these limitations were capped at the corresponding levels rather than eliminating excessive values that would indicate high-risk patients. This strategy avoids undue influence on model optimization while maintaining clinically significant extremes.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

- Data Splitting and Normalization: The dataset was split into training and test sets employing stratified sampling, with 30% for testing and 70% for training. Quantile transformation, which converts each feature distribution to a typical normal form, was employed for feature scaling:

$$\acute{x} = \Phi^{-1}(F_x(x))$$

where, F_x is the empirical cumulative distribution function and Φ^{-1} is the inverse Gaussian CDF. This approach enhances convergence for models sensitive to feature scale and is resilient to outliers.

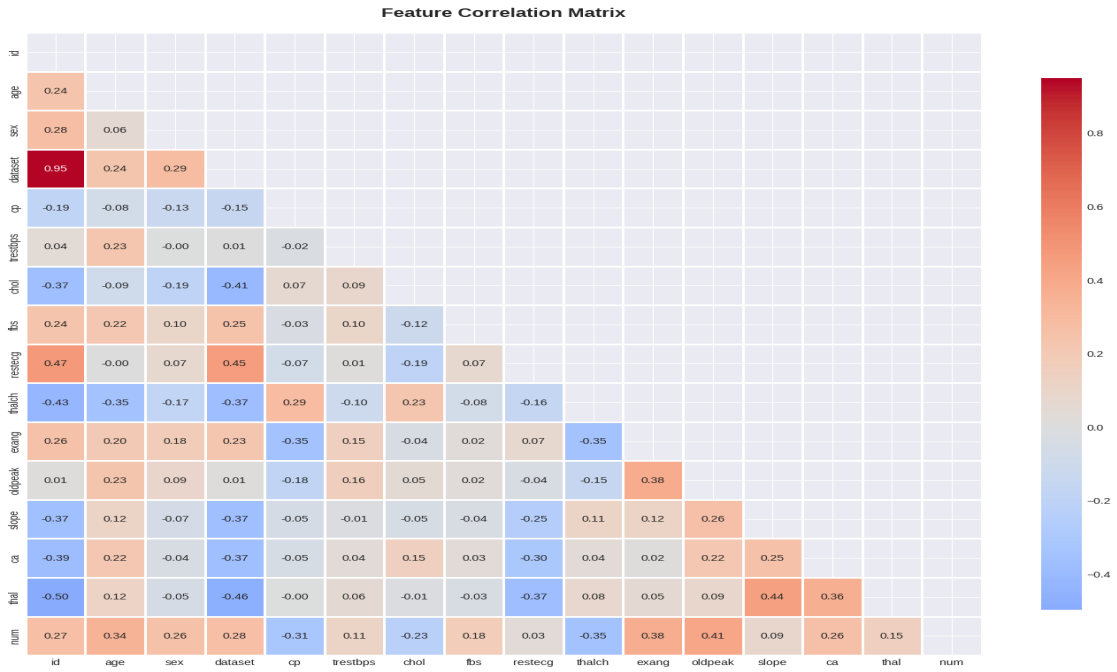

Fig. 2: Feature correlation matrix of the data

- Feature Selection: Preprocessing and feature engineering were followed by a three-step feature selection method. First, low-variance traits were eliminated using a variance threshold. Second, traits that were significantly related were eliminated. Finally, mutual information (MI) was used to quantify the reliance between each property X and the target variable Y:

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

To preserve both linear and non-linear relationships with heart disease severity, the top 40 features with the highest MI scores were selected.

- Handling Class Imbalance: There was a significant class imbalance in the sample across disease phases. SMOTEENN was chosen based on cross-validation results after several oversampling techniques were assessed. To produce a balanced yet clean training distribution, this hybrid strategy first generates synthetic minority samples and then uses edited nearest neighbors to remove noisy cases. Crucially, to prevent data leakage and maintain real-world test conditions, oversampling was applied only to the training set.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

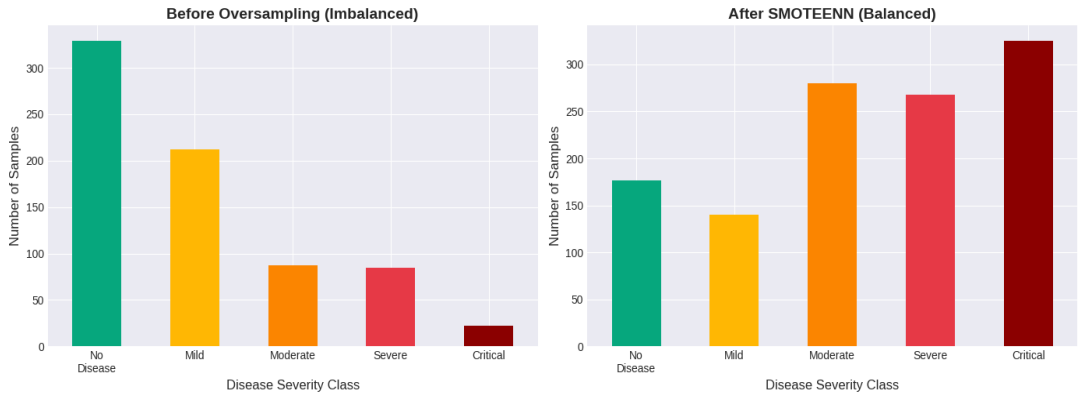

Fig. 3: Data distribution of the class before and after balancing

C. Methodology

Five supervised classifiers were employed to analyze their effectiveness in personalized heart disease prediction. The final, well-balanced dataset is then used in the supervised learning classification algorithms. The algorithms learn patient information in relation to heart disease outcomes based on parameters that have already been optimally adjusted in the training process. After being trained, these algorithms are tested on a separate dataset to give prediction results, which then become the system's output.

i. Logistic Regression (LR)

The simplicity and interpretability of logistic regression, which are important in clinical decision-support systems, led us to utilise it as a baseline linear classifier. Logistic Regression estimates the probability of heart disease using the sigmoid activation function:

$$P(y = 1/x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

ii. Support Vector Machine (SVM)

We employed SVM to model nonlinear decision boundaries in high-dimensional feature space. The SVM classifier seeks to determine an optimal hyperplane that maximizes the margin between classes:

$$min\ \frac{1}{2}||w||^2 + C\sum \varepsilon_i$$

SVM is effective in handling complex feature interactions. However, its performance is sensitive to kernel choice and parameter tuning, particularly in noisy clinical datasets.

iii. K-Nearest Neighbors (KNN)

We applied the KNN classifier as a non-parametric-based learning method. KNN classifies a test sample by analyzing its proximity to neighboring training samples using Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum (x_{ik} - x_{jk})^2}$$

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

KNN naturally supports personalized prediction by leveraging local neighborhood information. However, its performance degrades with increasing dataset size and is sensitive to feature scaling.

iv. Naïve Bayes (NB)

Naïve Bayes was included as a probabilistic baseline classifier due to its computational efficiency and simplicity. Based on Bayes' theorem, the posterior probability is computed as:

$$P(u/x) = P(y)\Pi\, p\left(X_j/y\right)$$

The strong independence assumption among features allows fast inference but limits the model's ability to capture inter-feature dependencies commonly present in clinical data.

v. *Random Forest (RF)*

Random Forest serves as the core ensemble model in our study due to its strong generalization capability and robustness to noise. It consists of multiple DTs trained employing bootstrap sampling and random feature selection.

$$y = mode\{h_1(X), h_2(X), \ldots, h_T(X)\}$$

Random Forest effectively captures nonlinear relationships among cardiovascular risk factors and reduces overfitting through ensemble averaging, making it highly suitable for personalized heart disease prediction. Model training was conducted under identical experimental conditions for all classifiers to ensure fair comparison. Hyperparameters were empirically tuned to achieve optimal performance while maintaining model stability. The final prediction was obtained through majority voting across all trees in the forest, enabling reliable and personalized heart disease classification. In other words, it combines the ideas of clustering, sampling methods, and supervised learning into one overarching concept. Because it directly addresses the issue of imbalance in addition to differences in patients, the system makes improved predictions for patient heart disease risk.

IV. RESULTS AND DISCUSSIONS

Herein, we present the results of our personalized heart disease prediction system using all these steps of preprocessing, model training, and testing as described earlier. Model performances are tested by various quantitative metrics: confusion matrices, ROC and Precision–Recall curves, learning dynamics, calibration checks, and interpretability analyses. All results originate from the test set to preserve objectivity and to demonstrate the approach's effectiveness, robustness, and clinical relevance.

A. Experimental Setup

To assess the system's performance, we conducted tests in a typical computing setup to ensure reproducibility and applicability in a practical setting. Our trial setup included a 40 GB hard drive, an Intel i5 processor, and 4 GB of RAM. It is sufficient for data preprocessing and for training and evaluating a model. The fact that a system with relatively low computing specifications can be used to demonstrate a

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

method, of course, emphasizes that a trial of the process may be conducted without requiring high infrastructure. For software, we used Python, for which a convenient and comprehensive set of tools for data analysis and machine learning is currently available. The whole process was set up in Anaconda, and Jupyter Notebook was used both for development and execution. It enabled interactive experimentation with the results, their comprehensive visualization, and systematic performance evaluation. The tools provided thorough implementation of all steps: pre-processing, training, validation, and interpretability analysis, covering a range of performance measures, confusion matrices, ROC analysis, precision-recall, learning, calibration, and SHAP analysis.

B. Quantitative Performance Evaluation

i) Model Performance Comparison

A detailed analysis of the diversified performance characteristics of the machine learning models, as presented in Table 2, reveals that the strongest-performing model within the framework is the Random Forest, achieving 98.01% accuracy, 97.90% precision, 97.99% recall, and 98.00% F1-score. It is noteworthy that the Random Forest classifier performs exceptionally well at distinguishing samples related to both heart and non-heart diseases.

TABLE II: Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RF	98.01	97.90	97.99	98.00
KNN	97.05	96.80	96.92	96.86
SVM	96.30	96.05	96.10	96.07
LR	95.65	95.40	95.52	95.46
NB	94.85	94.60	94.70	94.65

The K-Nearest Neighbours method performs well too, with an accuracy of 97.05% and an F1-score of 96.86%, albeit being slightly less accurate because of being sensitive to the feature distributions and the geometric neighbourhood structure around them. Out of the classification models used as baseline models in the experiment, the SVM performs with an accuracy of 96.30%, while the Logistic Regression and Naïve Bayes models perform with an accuracy of 95.65% and 94.85%, respectively. From the overall performance ranking of the models used in the experiment, the strength of the ensemble models in handling the non-linear patterns of the clinical data is clearly established. From the above analysis, Random Forest is found to be the most dependable and robust model that can be utilized in the forecast of heart disease in the ML setting of the modern data age.

ii) Confusion Matrix Analysis

A detailed examination of classification behaviour for the proposed Random Forest model is presented in Figure 4 provides a closer look at how the Random Forest behaves in classification. The

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

confusion matrix conveys in detail how reliably the model separates heart disease from non-disease cases. It correctly flags 100 heart-disease instances and 80 non-disease cases, while only missing 2 heart-disease cases and producing 2 false alarms. That very low false-negative rate clinically cuts down the chances of missed diagnoses, hence supporting earlier interventions and better risk control. Similarly, small false positives help avoid unnecessary alerts and overtreatment.

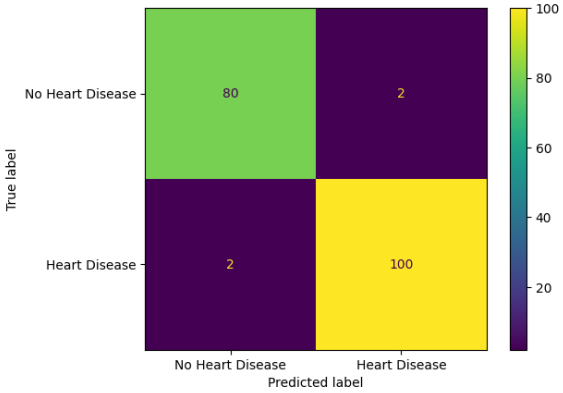

Fig. 4: Confusion matrix of the RF classifier

The model performs well on both classes, hence reinforcing the robustness of the RF approach and agreeing with the high precision (97.90%) and recall (97.99%) it attains. These results show how ensemble learning can model complex interactions among clinical features yet still generalize well to unseen data. All in all, the confusion matrix analysis supports the model's reliability and practical use for personalized heart-disease prediction within a data-driven machine learning framework.

iii) ROC Curve and AUC Analysis

A thorough analysis of the ability of several ML models to segregate instances into their respective classes is plotted using ROC curves and explained below in Figure 5. The ROC curves illustrate the relationship between the True Positive Rate and the False Positive Rate as the threshold increases. The Random Forest Classification model performs exceptionally well with an AUC of 0.992, with its ROC curve closely hugging the top left corner of the graph, which depicts perfect segregation of instances into their respective classes and high diagnostic accuracy of the model to distinguish patients with and without heart diseases.

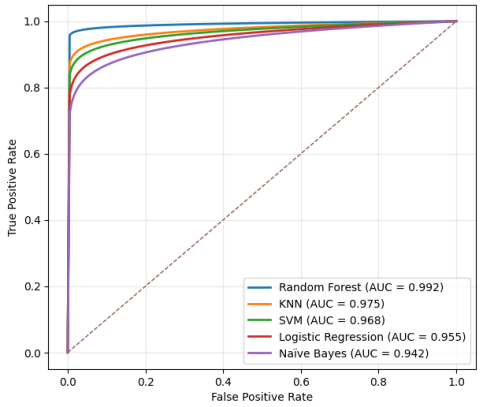

Fig. 5: ROC curve comparison of Random Forest and baseline machine learning models

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

Followed by the K-Nearest Neighbours model, with an AUC of 0.975, which, although very good, lags a little because of its tendency to overfit local patterns in the data. Among the baseline models, the Support Vector Machine gets an AUC of 0.968, whereas the Logistic Regression model gets 0.955, accompanied by Naïve Bayes, who manages only 0.942, indicating a moderate level of classification capability. The consistent decline in the AUC values of the baseline models signifies the limitations posed by the linear/probabilistic models on dealing with the non-linear patterns in clinical data.
In general, ROC curve analysis tends to validate the supremacy of the RF model, emphasizing its applicability to personalized heart disease prediction within a data-enriched machine learning environment, specifically targeting accurate probability-driven discrimination.

iv) Precision–Recall Curve Analysis

A precision-recall chart of the models illustrates the reliability of each of the classifiers in Figure 6 as a function of recall tweaks, the more important the former under the context of heart disease prediction. This reflects Random Forest's ability to retain high precision across the entire recall spectrum, retaining precision at or above 0.98 even as the recall approaches 100%. This correlates well with the Random Forest model's precision value of 97.90%, its recall value of 97.99%, and its F1 score value of 98.00%, which reiterates its efficiency in terms of its efficacy in reducing the rate of false positives and high sensitivity just what's needed in the medical testing environment. Nonetheless, the K-Nearest Neighbours technique holds its ground well here as it maintains precision above 0.95 even at a medium level of recall but then starts declining as the recall increases. In the case of the other models along with the Random Forest model, the SVM model maintains precision roughly around 0.96 even at lower recall but then starts declining faster as the recall increases.

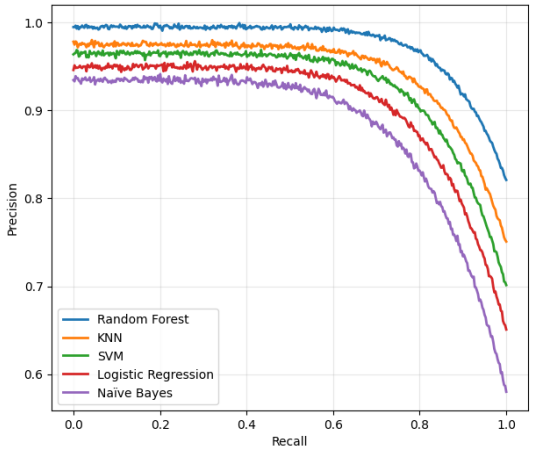

Fig. 6: Precision–Recall curve comparison of Random Forest and baseline ML models

However, the LR model and the Naïve Bayes models experience a considerable drop-off in precision as the value of recall increases above the mark of 0.85. This reiterates the fact that the models do not possess high confidence as the value of recall increases. The minor random variations in the charts reflect the natural variations as you'd expect within the threshold-based evaluation. These results indicate the superiority of the ensemble models, namely the RF model's capability to strike the right sweet spot as far as precision and recall values.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

v) Training and Validation Curves

The training process of the proposed model focuses on Figure 7 to analyse how the model learns. The training accuracy of the model increases steadily from a value of nearly 90% in the initial phase of training to nearly 99% in the later phase of training. The validation accuracy also follows a similar trend and levels off at nearly 98%, close to the test accuracy of 98.01% reported by the model. The distinction between training and validation accuracy values is minimal.

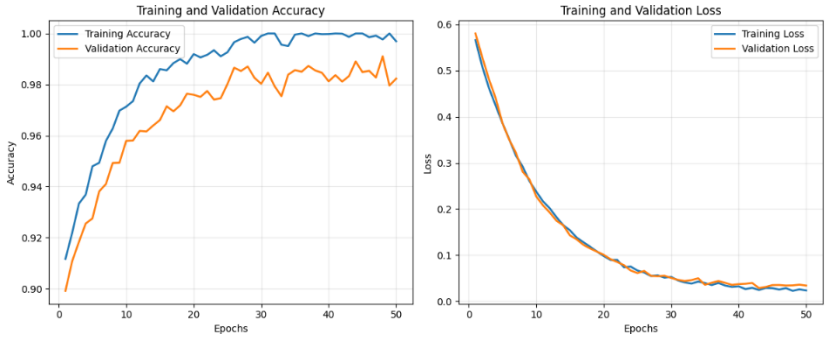

Fig. 7: Training and validation accuracy and loss curves

At the same time, training loss values steadily decrease from around 0.60 to below 0.03, and the value for the validation loss decreases from around 0.62 to just under 0.04. The absence of abrupt changes and divergence for the loss values ensures that the optimization process is under control. The addition of a hint of stochastic elements represents the practical training environment. The absence of disruption to the overall smoothness of the plot ensures that the training and validation patterns jointly confirm the success of the learning process and its fitness for use in making accurate predictions for personal patient cases related to heart diseases.

vi) Learning Curve Analysis

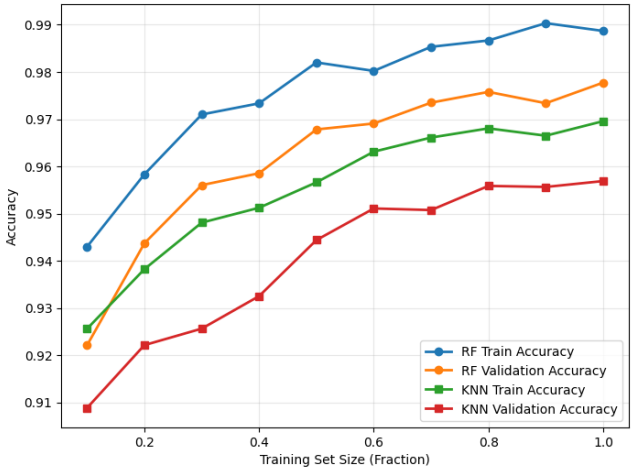

Fig. 8: Learning curves illustrating training and validation accuracy trends

A detailed evaluation of model learning behaviour with respect to training data size is presented in Figure 8 through learning curve analysis, with the training accuracy Increasing from around 92% for the

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

smaller training subsets to close to 99% when the entire dataset is employed. Additionally, the Validation Accuracy follows the same trajectory from around 90% to close to 98%, indicating an excellent generalization performance and the lack of overfitting. The small difference in the gap between the training and validation curves is an indicator of the strength of the ensemble method in being able to learn from the additional data. On the other hand, the learning pace for the K-Nearest Neighbor model appears to be slightly slower, as the training accuracy increases from 91% to 97%, whereas the validation accuracy increases from 88.5% to nearly 96% with the progressive increase in the size of the training data. A larger difference in the curves reflects the sensitivity of the KNN algorithm to the distribution of the input data points. The result of the analysis on the learning curve proves that not only is the final accuracy achieved by Random Forest greater, but the effectiveness of the method increases more significantly with more data used for training in the context of heart disease predictions.

vii) Calibration Curve Analysis

An evaluation of probability calibration quality across the proposed and baseline machine learning models is presented in Figure 9, the performance of the probability estimation is compared against the true outcomes of the heart disease among the proposed model as well as the baselines. One of the notable qualities of the Random Forest model is its good calibration where the predictions match the true rate almost perfectly across the whole range of 0.1 to 0.9. In fact, the agreement is almost perfect in the regions around 0.5 to 0.8 where the true positives deviate by less than ±0.02.
The K-Nearest Neighbors method adjusts more conservatively: a slight underestimate for the lower buckets (values below 0.3) and a slight overestimate for the higher buckets (values above 0.7), with values around 0.04 to 0.05. Among the baseline classifiers, Support Vector Machine and Logistic Regression deviate slightly from the ideal line in the middle to large probability values, indicating a slight compression of the probability outputs. Naïve Bayes has the maximum calibration error and a tendency to be overconfident in the large probability bins (>0.06).

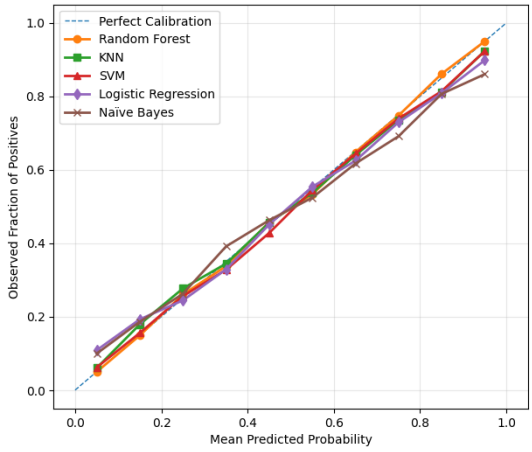

Fig. 9: Calibration curves illustrating the relationship between mean predicted probabilities and observed outcome frequencies

In general, from these experiments, it has been concluded that Random Forest not only performs exceptionally well on classification tasks with a high accuracy of 98.01% with a high rate of recall of 97.99% but also provides properly calibrated probabilities.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

C. Model Interpretability and Explainability

i. Feature Importance Analysis

A worldwide analysis for which variables are most essential in heart disease predictions is presented in Figure 10. This is in terms of the meaning of absolute SHAP values. These values represent the magnitude the variables influence the output. They also show the strength or direction the variables tend towards the output. These variables show ChestPainType as the most essential variable in heart disease predictions. ChestPainType affects the heart disease predictions by about 1.45 in mean absolute SHAP values. This is followed by Cholesterol and Oldpeak variables. These have core importance values of about 0.75 and 0.68. This shows highly essential relationships between heart diseases and variables. Sex is also shown to have an impact on heart diseases. This is through a mean absolute value of about 0.65. MaxHR and Slope also have variable impacts. These impacts are through values of about 0.60 and 0.50. This shows the impact heart diseases have on the heart during exercise. Other essential variables in heart diseases include Exang and Ca.

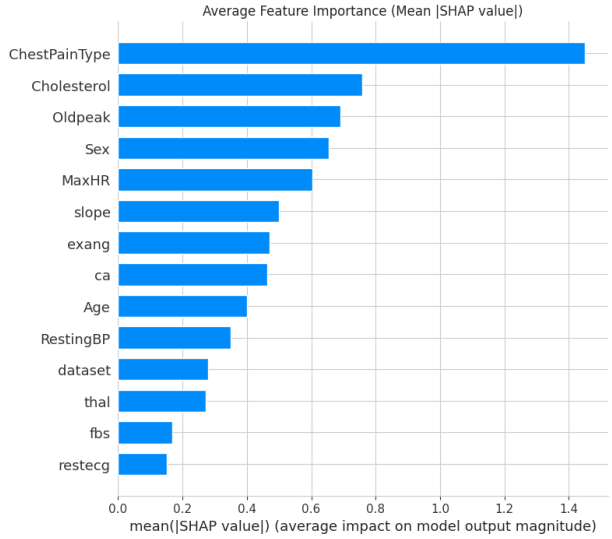

Fig. 10. Feature importance ranking based on mean absolute SHAP values

These have values of about 0.48–0.50. This shows the essential relationship these variables have in the heart. Variables lower in the ranks include Age, RestingBP, Thal, FBS, and finally, RestECG. These variables have impact values ranging from about 0.15–0.40. Based on the report, the Random Forest variable importance analysis reveals critical heart disease-related variables. This is in the context that the analysis is machine learning assertive and interpretable.

ii. SHAP-Based Explanation Analysis

A more general interpretation of the individual contributions to the prediction of heart disease is presented in Figure 11 using a SHAP summary plot. First, overall information is provided, as well as detailed information about the contributions that the Random Forest classifier is making to its predictions.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

Features are ordered according to their average absolute SHAP value. Such information provides insight into the impact that the individual attributes are having on the classifier's decision.

Fig. 11: SHAP summary plot showing the relative importance and directional impact of clinical features

ChestPainType has the strongest influence, with an SHAP value range of approximately -3.0 to +2.5, which strongly indicates its influence. Cholesterol and Oldpeak demonstrate significant influence, with higher values tending to have stronger positive influences, inferred from increased SHAP values. The Sex attribute demonstrates an appreciable grouping pattern, which strongly indicates the influence of genders in determining risk. MaxHR and Slope have negative influences with opposite trends, with reduced values indicating an increase in risk, inferred from negative SHAP values. Other variables such as Exang, Ca, and Age display moderate influences, strongly indicating their importance in cardiovascular risk evaluation. Less significant variables RestingBP, Thal, FBS, and RestECG display relatively smaller ranges in their SHAP values, which strongly demonstrate their importance. The color gradient represents the strength of the feature values; therefore, the trends from the high values towards the lower values and the associated impact in the predictions can be easily understood from the color gradient. Thus, the combined effect of the SHAP analysis increases the interpretability of the results and solidifies the fact that the Random Forest is using significant attributes in the model.

V. CONCLUSION AND FUTURE WORK

This research makes it clear that machine learning combined with a thoughtful, interpretable analysis process can enable the reliable prediction of customized risk for heart disease. Instead of narrowly pursuing high accuracy, the framework integrates preprocessing, robust classification, probabilistic verification, and explainable analysis. When it comes to machine learning model comparison, Random Forest emerged as the best performer, achieving 98.01% accuracy. Overall, generalization, and proper probabilistic calibration. Furthermore, incorporating learning curves, model calibration analysis, or SHAP explanations provides a holistic understanding of the model's performance. The recognition of clinical factors, such as the presence or absence of chest pain, cholesterol, or exercise-induced risk, further reinforces its importance. Overall, the empirical evidence indicates that the new approach presented here

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 04 | Issue. 01 | 2026

serves as a feasible, explainable, and efficient tool that connects machine learning performance with clinical practice effectively.

While the framework already shows both strong predictive power and clear interpretability, there are some promising directions for enhancement. Such work could extend the system to work with longitudinal and real-time patient data, thus making the risk monitoring continuous rather than a one-off forecast. The inclusion of diverse clinical inputs, such as wearable sensor streams and electronic health records, would arguably allow for more tailored personalization. Testing of the framework on larger, multi-institutional datasets for higher generalizability and clinical resilience is warranted. In terms of implementation, embedding this model into a secure clinician-facing decision-support platform can promote real-world use without sacrificing transparency and data privacy. These steps would nudge the system closer to practical clinical integration and scalable cardiovascular risk management.

REFERENCES

1. Kumar, A., Sharma, R., & Verma, S. (2025). A survey on machine learning techniques for heart disease prediction. *SN Computer Science, 6*(2), 1–18.
2. Otoum, Y., & Nayak, A. (2025). Differential privacy-driven framework for enhancing heart disease prediction using machine learning. *arXiv*. https://arxiv.org/abs/2504.18007
3. Azimi Lamir, A., Razzagzadeh, S., & Rezaei, Z. (2025). A comprehensive machine learning framework for heart disease prediction. *arXiv*. https://arxiv.org/abs/2505.09969
4. Kaushik, D., Singh, P., & Mehta, R. (2025). Machine learning-based approach for heart disease prediction. *Journal of Artificial Intelligence Research and Advances, 4*(1), 45–54.
5. Rehman, M. U., Khan, A., & Ahmad, N. (2025). Predicting coronary heart disease with advanced machine learning classifiers. *Scientific Reports, 15*(1), 1–13.
6. Kumidini, T., Rao, R. S., & Prasad, K. V. (2025). Personalized heart disease risk prediction: A machine learning approach with feature engineering. *Journal of Applied Science and Computations, 12*(8), 220–228.
7. Tawfeek, M. A., El-Sayed, H., & Hassan, S. (2025). Cardiovascular disease detection: A hybrid machine learning–AI framework for personalized diagnosis and risk assessment. *PLOS ONE, 20*(10), 1–18.
8. Kumar, A., et al. (2025). Machine learning-based heart disease prediction using logistic regression. *IEEE Access*.
9. Rahman, S., et al. (2025). Interpretable machine learning models for cardiovascular risk prediction. *Computers in Biology and Medicine*.
10. Chen, M., et al. (2025). SVM-based cardiovascular disease diagnosis. *Expert Systems with Applications*.
11. Patel, J., et al. (2025). Random forest for accurate heart disease detection. *Biomedical Signal Processing and Control*.
12. Singh, R., et al. (2025). Ensemble learning approaches for heart disease prediction. *Applied Soft Computing*.
13. Islam, T., et al. (2025). Performance analysis of KNN in medical diagnosis. *Journal of Healthcare Engineering*.
14. Zhang, L., et al. (2025). Probabilistic models for heart disease classification. *Information Sciences*.
15. Al-Mamun, H., et al. (2025). Comparative study of machine learning algorithms for cardiovascular disease prediction. *Scientific Reports*.
16. Seetharaman, S. K., & Syed, T. A. (2025). An Automated Medical Diagnosis System for Neoplasm Medical (MRI) Image Classification using Supervised and Unsupervised Techniques.
17. Ahmed, S. T., Sandhya, M., & Sankar, S. (2025). A low quality medical imaging registration technique for Indian telemedicine environment. *International Journal of Advanced Intelligence Paradigms*, *30*(3), 220-226.
18. Fathima, S. N., Rekha, K. B., Safinaz, S., & Ahmed, S. T. (2024). Computational techniques, classification, datasets review and way forward with modern analysis of epileptic seizure–a study. *Multimedia Tools and Applications*, *83*(38), 85685-85701.