International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 6, December 2025, pp. 5655~5666
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i6.pp5655-5666

Machine learning model for accurate prediction of coronary artery disease by incorporating error reduction methodologies

Santhosh Gupta Dogiparthi[1], Jayanthi K.[1], Ajith Ananthakrishna Pillai[2], K. Nakkeeran[3]
[1]Department of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India
[2]Department of Cardiology, Kauvery Hospital-Radial Road, Chennai, Tamilnadu, India
[3]School of Engineering, Fraser Noble Building, University of Aberdeen, Aberdeen, United Kingdom

Article Info

Article history:

Received Apr 9, 2025
Revised Jul 16, 2025
Accepted Sep 14, 2025

Keywords:

False negatives
Feature engineering
Misclassification cost
Threshold adjustment
Weighting

ABSTRACT

Coronary artery disease (CAD) remains a leading cause of mortality worldwide, with an especially high burden in developing countries such as India. In light of increasing patient loads and limited medical resources, there is an urgent need for accurate and reliable diagnostic support systems. This study introduces a machine learning (ML) framework that aims to enhance CAD prediction accuracy by specifically addressing the reduction of false negatives (FN), which are critical in medical diagnostics. Utilizing a stacked ensemble model comprising five base classifiers and a meta-classifier, the framework integrates cost-sensitive learning, classification threshold tuning, engineered features, and manual weighting strategies. The model was developed using a clinically acquired dataset from the Jawaharlal Institute of postgraduate medical education and research (JIPMER), consisting of 428 patient records with 36 original features. Evaluation metrics show that the proposed model achieved an accuracy of 92.19%, sensitivity of 98%, and an F1-score of 95.15%. These improvements are significant in a clinical context, potentially reducing missed diagnoses and improving patient outcomes. The model is intended for deployment in cardiology outpatient settings and demonstrates a scalable, adaptable approach to medical diagnostics.

Corresponding Author:

Santhosh Gupta Dogiparthi
Department of Electronics and Communication Engineering, Puducherry Technological University
Puducherry 605014, India
Email: santhoshgupta@ptuniv.edu.in

1. INTRODUCTION

Cardiovascular diseases, especially coronary artery disease (CAD), are a predominant cause of morbidity and mortality globally, contributing to approximately 28.1% of all deaths in India alone [1]. Given India's population surpassing 1.44 billion and an existing doctor-to-patient ratio significantly below the World Health Organization's recommended 1:1000 [2], the existing healthcare system faces a massive burden in timely and accurate diagnosis. This scenario highlights the need for intelligent decision-support systems capable of alleviating the diagnostic workload, particularly in resource-constrained environments.

The central problem in current machine learning (ML) based diagnostic models for CAD lies in their misclassification tendencies, especially false negatives (FN) cases where diseased patients are misclassified as healthy. This poses a severe risk in clinical settings, where early detection is crucial. Traditional models often optimize for accuracy or AUC, inadvertently neglecting sensitivity and error costs. A false negative in CAD diagnostics can delay life-saving interventions and mislead clinical decisions. Therefore, reducing FN must be a core objective for any ML model aimed at CAD prediction. Several ML

methodologies are available to reduce the FN depending on the specifics of different feature characteristics of the input dataset and the model algorithm implemented. In the broader landscape of predictive modeling, several methodologies such as adjusting the classification threshold [3], cost-sensitive learning [4], resampling techniques [3], feature engineering [5], algorithm selection [6], data augmentation [7], error analysis [3], ensemble methods [8], model calibration [9], active learning [10], manual and performance-based weight adjustment [3], [10], hyperparameter tuning [11], and custom stacking [12] are commonly adopted to improve classification outcomes, particularly by reducing false negatives. Out of these methodologies, depending upon the prediction model developed, different methodologies are to be incorporated to reduce the FN. Selected set of methodologies for heart disease prediction systems with better prediction accuracy and recall are vital in India to combat the growing health issues due to cardiovascular diseases. They facilitate early detection, optimize resource utilization, reduce healthcare costs, and support better healthcare outcomes, contributing to overall societal well-being.

Several works in literature have explored ML-based approaches for CAD detection. For instance, Khanna *et al.* [13] compared classification algorithms like SVM and neural networks for heart disease prediction, while Maini *et al.* [14] emphasized the utility of tailored prediction models for Indian populations. However, these studies primarily focused on accuracy enhancement without a detailed strategy for handling FN errors. Similarly, Zriqat *et al.* [15] and Babu *et al.* [16] discussed improved ML architectures but lacked targeted error-cost analysis or explicit threshold optimization methods. Thus, despite the progress in ML applications for CAD, a significant gap remains in handling error sensitivity, particularly reducing FNs.

This study proposes an ensemble ML model tailored for FN reduction in CAD prediction. The model combines five diverse base learners (linear and non-linear SVM, k-nearest neighbors, random forest, and AdaBoost) with a stacking meta-classifier. It incorporates several novel strategies: cost-sensitive learning [4] to penalize FN, threshold adjustment based on precision-recall trade-offs [3], domain-driven manual weight assignment [3], and derivation of meaningful composite features such as pulse pressure, MAP, and ECG abnormality scores. These modifications are aimed at fine-tuning the model to reduce misdiagnoses while maintaining generalization ability.

Our innovation lies in the integration of these FN-reduction methodologies into a clinically validated ensemble model. Unlike earlier studies, this approach emphasizes medical safety by lowering FN cases from six to two on a 428-patient dataset, improving recall from 94% to 98% a substantial leap in diagnostic reliability. This work not only addresses a critical clinical challenge but also sets a replicable precedent for deploying ML systems in real-world hospital workflows. The proposed model is currently being prepared for deployment at JIPMER and may be adapted for other disease domains with similar feature sets.

The remainder of this paper is organized as follows: section 2 presents the proposed model and its theoretical foundation. Section 3 details the methodology used in building and tuning the ml model. Section 4 reports the results and provides a comprehensive discussion comparing the model's performance with existing methods. Finally, section 5 concludes the paper and outlines directions for future work.

2. PROPOSED MODEL FOR CAD PREDICTION

This section presents the design and theoretical foundation of the proposed machine learning framework for predicting coronary artery disease. It further elaborates on the incorporation of multiple false negative (FN) reduction strategies to the ensemble learning structure to improve diagnostic reliability by minimizing FN errors, which are critical in medical decision-making.

2.1. Stacked ensemble architecture

The proposed machine learning model is designed as a stacked ensemble framework that integrates multiple base classifiers to leverage the individual strengths of each. Specifically, it combines five classifiers viz., support vector machine (SVM) with both linear and nonlinear kernels [17], [18], k-nearest neighbors (KNN) [19], random forest (RF) [20], and AdaBoost [21] followed by a meta-classifier that synthesizes their outputs for final decision-making. Ensemble learning methods, particularly stacking, are known to enhance model generalization and performance across diverse datasets. In this context, the choice of classifiers was informed by their complementary nature: SVM for margin-based separation, KNN for local decision boundaries, RF for feature importance and bagging, and AdaBoost for handling difficult-to-classify instances. The meta-classifier in the final layer captures and balances these behaviors to minimize overall classification error, particularly false negatives.

2.2. Integration of false negative reduction methodologies

A distinguishing innovation in our model is the incorporation of multiple strategies explicitly targeting the reduction of false negatives (FN), a critical concern in medical diagnostics. First, a cost-sensitive

learning approach is employed where the misclassification cost of FN is set significantly higher than that of false positives, thereby influencing the model to be more conservative when predicting negative outcomes. Second, manual weighting is applied to base classifiers based on their historical performance and clinical insight this emphasizes more reliable models in the final decision. Third, the decision threshold is optimized using a precision-recall trade-off, selecting a value that maximizes the F1-score, which is particularly suited for imbalanced datasets. Finally, domain-specific engineered features such as pulse pressure, mean arterial pressure, ECG abnormality score, and comorbidity counts are introduced to enhance model interpretability and prediction strength. This multi-pronged strategy collectively reduces FN instances, making the model more trustworthy and clinically viable. Similar FN-reduction strategies have been successfully applied in recent works using cost-sensitive ensemble methods and threshold-moving techniques [22]–[24].

3. METHODS

This section outlines the end-to-end methodology adopted for building the proposed ML model for coronary artery disease prediction. It covers the dataset characteristics, feature engineering strategies, model training process, and evaluation metrics. The methodological pipeline ensures high reproducibility and transparency in data processing, model development, and performance analysis.

3.1. Dataset description

A dataset was prepared by collecting the demographical, clinical assessment, ECG, lab and ECHO features of 428 patients from Department of Cardiology, JIPMER, Puducherry. The dataset has 36 different features with the last feature indicating in binary values about the presence '1' or absence '0' of coronary artery disease (CAD). Full details about the remaining 35 features are provided in Table 1.

Table 1. Details of 35 features from the JIPMER CAD dataset

Feature Name	Data Range	Type	Feature Category
Age (Years)	20–83	Numerical	Demographical
Weight (kg)	50–90	Numerical	Demographical
Length (cm)	130–180	Numerical	Demographical
Gender	1–M,0–F	Binary	Demographical
BMI (kg/m^2)	19.37–35.5	Numerical	Demographical
Diabetes Mellitus	1–Y, 0–N	Binary	Demographical
Hypertension	1–Y, 0–N	Binary	Demographical
Current Smoker	1–Y, 0–N	Binary	Demographical
Ex-smoker	1–Y, 0–N	Binary	Demographical
Dyslipidemia	1–Y, 0–N	Binary	Demographical
Systolic BP (mmHg)	90–189	Numerical	Clinical Assessment
Diastolic BP (mmHg)	52–112	Numerical	Clinical Assessment
Pulse rate (/min)	44–140	Numerical	Clinical Assessment
Angina	1–Y, 0–N	Binary	Clinical Assessment
Dyspnea	1–Y, 0–N	Binary	Clinical Assessment
Rhythm	0–Sinus, 1–Fibrillation	Binary	ECG
Q-Wave	1–Y, 0–N	Binary	ECG
QS Wave	1–Y, 0–N	Binary	ECG
QRS Complex	1–Y, 0–N	Binary	ECG
Axis Deviation	1–Y, 0–N	Binary	ECG
ST-T changes	1–Y, 0–N	Binary	ECG
T-Wave	1–Y, 0–N	Binary	ECG
ST elevation	1–Y, 0–N	Binary	ECG
ST depression	1–Y, 0–N	Binary	ECG
T inversion	1–Y, 0–N	Binary	ECG
Left ventricular hypertrophy	1–Y, 0–N	Binary	ECG
Poor R Wave Progression	1–Y, 0–N	Binary	ECG
Bundle Branch Block	0–Absence, 1–LBBB/ RBBB	Nominal	ECG
RBS (mg/dl)	60–733	Numerical	Lab and ECHO
Creatinine (mg/dl)	0.3–2.7	Numerical	Lab and ECHO
Blood Urea (mg/dl)	9.3–70	Numerical	Lab and ECHO
Haemoglobin (gm/dl)	7.1–22	Numerical	Lab and ECHO
Platelet Count (1000/ml)	90–586	Numerical	Lab and ECHO
Ejection Fraction (%)	15–66	Numerical	Lab and ECHO
Regional wall motion abnormality	0–Normal, 1–Abnormal	Binary	Lab and ECHO

3.2. Data preprocessing and feature engineering

Prior to model training, the raw clinical dataset obtained from JIPMER underwent systematic preprocessing. This included handling missing values, outlier treatment, and normalization of continuous

variables such as age, systolic/diastolic blood pressure, and ejection fraction. Categorical variables such as gender, smoking status, and ECG indicators were encoded into numerical representations using one-hot or ordinal encoding schemes, as appropriate. Feature engineering was conducted to enhance the model's discriminative power. Ten derived features were added, including clinically significant constructs like pulse pressure (systolic minus diastolic BP), mean arterial pressure (MAP), comorbidity count, ECG abnormality score, and ratios such as RBS to BMI. Additionally, categorical recoding for variables such as heart rate and ejection fraction was introduced to map these into clinically interpretable categories. The combined use of raw and engineered features yielded a total of 46 predictors, providing a richer and more robust input space for training the ensemble model.

3.3. FN reduction methodologies

In developing a medical diagnostic tool to perform binary classification as, patient with disease (positive class) or without disease (negative class), there are two possible misclassification errors namely FP and FN are available. In this scenario:

- A FN is when the ML tool incorrectly classifies a diseased patient as healthy, which can be very detrimental in terms of patient health.
- A FP is when the ML tool incorrectly classifies a healthy patient as diseased, which can lead to unnecessary stress and further expensive tests.

In the following sub sections, detailed information about modifications to reduce the FN and pipeline architecture of the tuned ML model are explained.

3.3.1. Cost-sensitive learning

Most of the ML methodologies assumes the misclassification errors are justified as they are inherently provided by the model itself. Rather than such a justification, the incorrect predictions of FP or FN should be seen as a question on the reliability of the ML model prediction. Incorrect predictions of FN and FP compared to the TP and TN leads to reduction in the recall and hence lower accuracy level for the final predicted values of the different features by the ML model. To avoid such errors in prediction by the ensemble model, a misclassification cost matrix is introduced to train all base classifiers to be followed by a meta classifier to reduce the wrong predictions. To reflect the severity of errors, a cost matrix is defined such that cost of a FN is higher than a FP. Thus, calculated elements for the cost matrix are not to be simply used as a multiplication factor, rather to be used as an influencing factor on how the classifier evaluates its decision. In the following we provide the details on how the cost matrix can be used to adjust the predicted values by the ML model.

Very common methodology is to modify the decision threshold of the classifier using the elements calculated for the construction of the cost matrix. Let us denote the predicted probabilities of the positive class as p (output of the classifier), and assume the default threshold value for classifying a positive instance as 0.5. If the cost of a FN (C_{FN}) is higher than the cost of a FP (C_{FP}), there is a requirement to decrease the threshold value to reduce the FN. On the other hand, if the C_{FP} is greater than the C_{FN}, there is a requirement to increase the threshold value to reduce the FP. The adjusted threshold ($T_{adjusted}$) can be calculated as:

$$T_{adjusted} = \frac{C_{FP}}{C_{FP} + C_{FN}},$$

and the corresponding prediction is adjusted as follows:
If $p \geq T_{adjusted}$, classify as positive,
If $p < T_{adjusted}$, classify as negative.

These adjustments balance the costs associated with the FP and FN as per the specific cost matrix. Also, it ensures that the model's predictions align with the specific cost considerations, potentially leading to better outcomes in real-world scenarios where different types of errors, namely, type-I and type-II errors [14] have different consequences. In this research study, the misclassification cost matrix used is $[0, 17, 20, 0]$ for $[TP, FP, FN, TN]$ where all the base classifiers and meta classifier are trained and tested with their respective datasets and cost matrix.

3.3.2. Manual weight adjustment

Manual weight adjustment involves the modification of the prediction weightings in the individual base model's ensemble to improve the overall performance. This methodology leverages domain expertise to provide proper significance to models that are expected to perform better in certain aspects.

Consider an ensemble of N = 5 base models. Let matrix X (428×35) be the input feature matrix, and y (428×1) be the target vector. The prediction from the i^{th} model for a given input x (elements of matrix X) is denoted as $\hat{y}_i(x)$. The ensemble prediction is typically a weighted combination of these individual predictions:

$$\hat{y}_{ensemble}(x) = \sum_{i=1}^{N} w_i \, \hat{y}_i(x),$$

where w_i are the weightings assigned to the predictions from each base model. In manual weight adjustment, these weightings w_i are set based on the feature's domain expertise.

For our JIPMER dataset, the manual weightings used are [0.5, 0.8, 0.7, 0.9, 0.85] and the final model scores are obtained by multiplying with weightings and prediction scores of respective train and test datasets.

3.3.3. Threshold adjustment

Adjusting the decision threshold of a classifier can bring a trade-off between the precision and recall. The optimal threshold can be determined by maximizing the F1-Score and the necessary steps to be followed are:

- Prediction scores: for each test instance, obtain the predicted probability scores $\hat{p}(x)$ from the model.
- Thresholds: define a range for possible thresholds θ to evaluate. For each threshold θ, classify the data points as:

$$\hat{y}_\theta(x) = \begin{cases} 1 & if \;\; \hat{p}(x) \geq \theta \\ 0 & if \;\; \hat{p}(x) < \theta \end{cases}$$

- Compute metrics: for each threshold θ, compute precision and recall.
- F1-Score calculation: calculate the F1-Score for each threshold.
- Optimal threshold: select the threshold θ^* that maximizes the F1-Score:

$$\theta^* = arg \; \max_{\theta} F1\text{-}Score(\theta).$$

3.3.4. Engineered/derived features

Derived features, also known as engineered features, are new variables created by transforming or combining existing features. These features can capture more entangled relationships interconnecting the data that raw features might not reveal. In the context of medical data, derived features can be particularly useful in enhancing the predictive capacity and capability of the model by incorporating domain knowledge and specific medical insights. Its main purpose is:

- By capturing additional information that raw features alone might not provide, derived features can help to improve the accuracy, precision, recall, and overall performance of the ML models.
- Derived features can often make the models more interpretable by bringing out the important relationships and patterns in the data that are meaningful to the nature of the dataset (medical in our case study).
- Properly engineered features can help to reduce the influence of the noisy data in the prediction process by focusing more on the relevant aspects of the data.

Derived features can be of different kinds such as, statistical, temporal, transformation, etc. Depending on the prediction requirements any particular feature kind can be adopted. Nevertheless, inclusion of more feature(s) to the existing dataset features will introduce different constraints to the ML model like the overfitting, relevance and complexity.

A total of 10 derived features are obtained for this work, where their names, relations with other features and their significance are listed below.

- Pulse pressure: It is defined as:

$$Pulse\ Pressure = Systolic\ BP - Diastolic\ BP.$$

  This derived feature provides insights into cardiovascular health and arterial stiffness.
- Mean arterial pressure (MAP*):* It represents the average pressure in a patient's arteries during a cardiac cycle. It is a useful indicator of perfusion pressure of the organs. It is defined as:

$$MAP = (Systolic\ BP + 2 \times Diastolic\ BP)/3.$$

- Comorbidity count: This derived feature counts the patient's number of comorbid conditions. The presence of multiple comorbidities [such as diabetes mellitus (DM), hypertension (HTN), and dyslipidemia (DLP)] can significantly impact the patient's overall health and risk profile. It is useful to assess the burden of chronic diseases on a patient and helps in stratifying risk and tailoring treatment plans. It is defined as:

$$Comorbidity\ Count = DM + HTN + DLP.$$

- Heart rate category: It categorizes the patient's heart rate into three levels: Low, Normal, and High. Heart rate is a critical vital sign that can indicate underlying conditions such as bradycardia (low heart rate), tachycardia (high heart rate), and normal heart functioning. This derived feature helps to swiftly identify any abnormalities in the heart rate that require immediate medical attention. It is computed as:

$$Heart\_Rate\_Category = discretize\ (data.PR,[0, 60, 100, Inf],\ 'categorical',\ 'Low',\ 'Normal',\ 'High').$$

- ECG abnormalities count: This derived feature adds up various abnormalities found in an ECG reading. Each of these components (QW, QS, QRSC) [22], [23] represents different types of ECG abnormalities that can indicate various heart conditions. Sum of these components becomes yet other derived feature that provides a consolidated measure of the overall ECG abnormality burden, which can be useful to predict cardiac events and the severity of the heart disease. It is calculated as:

$$ECG\_Abnormalities\_Count = QW+QS+QRSC+STT+TW+STE+STD+TI+LVH+PRW+BBB.$$

- Random blood sugar to BMI ratio: It measures the relationship between random blood sugar (RBS) levels and body mass index (BMI). It helps in understanding how blood glucose levels are affected by body weight. This ratio can be particularly useful in managing diabetes and obesity-related conditions, providing insights into the metabolic status of the patient. This ratio is calculated as:

$$RBS\_BMI\_Ratio = RBS/BMI.$$

- Left ventricular ejection fraction (LVEF) Category: This derived feature quantifies the amount of blood pumped by the left ventricle in each heart contraction. Categorizing LVEF into Low, Normal, and High helps in assessing the functional status of the heart. A low LVEF indicates heart failure or cardiomyopathy, while normal and high categories are indications of good heart functioning. This categorization is crucial for diagnosing and monitoring heart conditions. LVEF category is computed as:

$$LVEF\_Category = discretize(data.LVEF,[0, 40, 55, Inf],\ 'categorical',\ 'Low',\ 'Normal',\ 'High').$$

- Angina relative risk: Relative risk values are based on the gender and age for patients with typical angina (ANG) [25]. These values are manually assigned based on predefined risk categories as specified by the cardiologists. The probability values assigned for CAD as per the category depicted in Table 2.

Table 2. Probability of CAD based on age and gender

Age	Typical Angina	
	Men	Women
30-39	0.76	0.26
40-49	0.87	0.55
50-59	0.93	0.73
60-69	0.94	0.86

- Diabetes mellitus relative risk: relative risk for patients with diabetes mellitus (DM) have two-to four-fold possibility of developing coronary disease [26]. So, presence of DM is manually set to 0.7 and absence as 0.3.
- Smoking status: smoking status is another critical determinant of various health risks. Current smokers and ex-smokers have different risk profiles compared to individuals who have never smoked. This feature

combines the information from *Current_Smoker* and *Ex_Smoker* features into a single categorical variable. It uses the categorical values of the two smoking feature to map into a single value that corresponds to one of the following three categories: i) 0 for '*Never Smoked*', ii) 0.5 for '*Ex_Smoker*', and iii) 1 for '*Current_Smoker*'.

By combining into a single categorical variable, the complexity of handling multiple related features is reduced. This helps in better interpretation and analysis.

3.4. Model training and evaluation

The pipeline architecture of the tuned ML model is depicted in Figure 1. This workflow captures all stages of the modeling process, including data pre-processing, derived feature creation, model training with cost-sensitive learning, manual weightings, and classification threshold adjustment to optimize predictive performance. The proposed pipeline offers a structured and effective methodology for handling complex medical datasets.

Initially, the dataset was partitioned into 70% for training and 30% for testing. After training, the model predictions were validated across the full dataset using cross-validated predictions to ensure robustness. The modeling process starts by loading the training and test sets, followed by preprocessing and feature selection. The five base classifiers are trained, their outputs fed into a meta-classifier, and the ensemble model is evaluated using a test dataset.

For the base configuration (without any FN-reduction methods), the training procedure includes the following steps: (i) Load the dataset and perform a 70:30 train-test split; (ii) Apply data cleaning and feature selection; (iii) Train the ensemble model using five base classifiers and a meta-learner; (iv) Test the full dataset on the trained model; (v) Compare predicted values with actual labels to generate the confusion matrix; (vi) Compute evaluation metrics such as accuracy (ACC), precision (P), recall/sensitivity (S), specificity (SP), F1-score, Matthew's Correlation Coefficient (MCC), and area under the curve (AUC).

The pipeline ensures interpretability and reproducibility while enabling the use of ensemble methods with integrated FN reduction. A detailed comparison of the performance metrics for the baseline model (without FN reduction) is presented later in section 4 (Results and Discussion), where Table 3 reports the confusion matrix and associated evaluation outcomes.

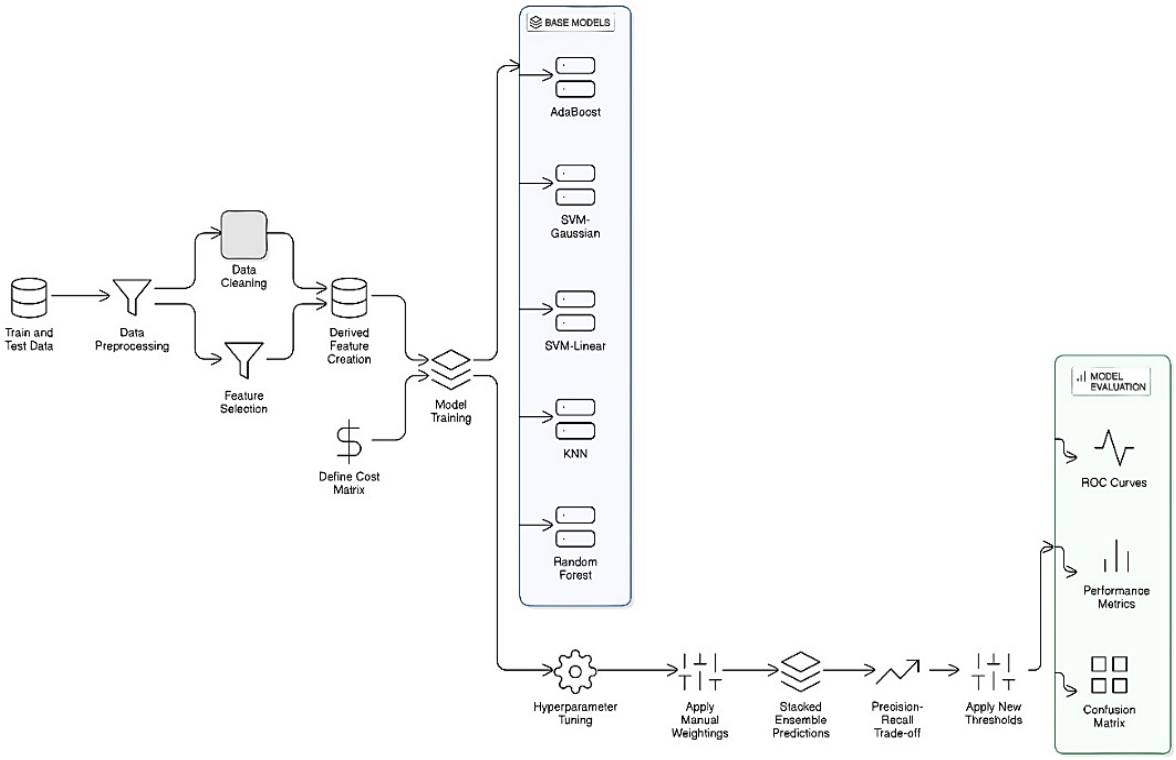

Figure 1. Pipeline architecture of the proposed ML model, including preprocessing, feature engineering, FN-reduction, training, and evaluation

4. RESULTS AND DISCUSSION

This section presents and analyzes the performance of the proposed ML model for CAD prediction, focusing first on the baseline configuration without any error-reduction methodologies, followed by improvements achieved through the integration of individual and combined false negative (FN) reduction strategies.

4.1. Baseline model performance

Table 3 summarizes the confusion matrix and performance metrics for the baseline ML model trained and evaluated without incorporating any FN-reduction methodologies. This configuration uses the full ensemble pipeline with original and derived features but applies no cost-sensitive learning, manual weighting, threshold tuning, or engineered strategies. As shown, while the model demonstrates respectable overall accuracy and AUC, it still suffers from a relatively higher false negative (FN) count, which is a critical concern in clinical applications where undiagnosed CAD cases can have severe consequences.

The results from Table 3 highlight the need to address the FN issue directly. To evaluate the impact of FN-reduction strategies, additional experiments were conducted, applying each method individually and in combination.

Table 3. Performance metrics of the baseline model without FN reducing methodologies

S.No.	Dataset	Confusion Matrix				Performance Metrics						
		TN	FP	FN	TP	ACC	P	S	SP	F1	MCC	AUC
1.	JIPMER	20	8	6	94	89.06%	92.16%	94.00%	71.43%	93.07%	67.23%	92.04%

4.2. Individual FN-reduction methodologies

Here we report the effectiveness of the ML model in accurate prediction due to the incorporation of FN reduction methodologies detailed earlier. First, we detail the performance of the ML model with only one of the FN reduction methodologies is incorporated. In Table 4, all available features from the JIPMER dataset are applied and provided all the performance metrics of the model with individual methodologies utilized in the study i.e., engineered features, cost-sensitive learning, manual weightings provision and calculation of new thresholds using precision-recall trade-off. First model i.e., baseline model (S.No.1 in Table 4) that used only the original features resulted in maximum percentages for various metrics as: accuracy 89.06%, precision 92.16%, recall 94%, F1-score 93.07% and AUC 92.04%. These performance metrics are not the best we wished for as the original dataset features has only captured the essential characteristics required for accurate classification. Second model (S.No.2 in Table 4) is addition of engineered features along with the original features showed slight improvement in the precision 93% and recall 93%, but with a marginal decrease in AUC 91%. Marginal decrease in the AUC metrics of this case can be attributed to the noise and/or redundancy introduced by the inclusion of the derived features for the training and prediction. Third model (S.No.3 in Table 4) is the incorporation of the cost matrix that works only on the original features to reduce the FN, provided significant performance improvement in accuracy 91.41%, recall 97% and F1-score 94.63%, along with a slight compromise in AUC 90.39%. This improvement of the accuracy is a direct reflection of the high reduction of FN metric in this model. Either fourth (S.No.4 in Table 4) model that incorporated manual weightings or the fifth (S.No.5 in Table 4) model that incorporated threshold adjustments to the original features dataset did not show any significant change in the performance compared to the first model (S.No. 1 in Table 4). This indicates that there are no added advantages in using both manual weightings and threshold adjustments methodologies together for accurate prediction of JIPMER CAD medical kind of datasets. The manual weightings did not alter the model's decision boundary enough to cause any appreciable change in performance. As the original model already included the optimal threshold in the prediction process, the fifth model that incorporated the threshold adjustments has no effect on reducing the FN. The model's performance metrics suggest that the default threshold was appropriate, and any further adjustments did not enhance or degrade the model's performance.

4.3. Combination of FN-reduction methodologies

In Table 5, we report the effectiveness of the ML model in predicting the JIPMER dataset when combination of more than one FN reducing methodologies are incorporated. First combination model (S.No.1 in Table 5) that used original and derived features along with manual weightings provided the strongest performance with accuracy 90.63%, recall 96.00% and F1-score 94.12%, while AUC got decreased to 91.00%. Second combination model (S.No.2 in Table 5) used original and derived features along with threshold adjustments, mirrored the first combination model's performance metrics indicating that there is no

additional benefit compared to first combination model that used original and derived features along with manual weightings. The third combination model (S.No.3 in Table 5) uses original and derived features with a cost matrix, demonstrates balanced performance metrics with an accuracy 89.06%, precision 93.00%, and an AUC 90.63% but not provide any significant improvement compared to first (or third) combination model. By observing the first three combination model cases, the addition of derived features with any methodology did not show substantial improvement due to potential noise or redundancy. The fourth combination model (S.No.4 in Table 5) used original features along with manual weightings and a cost matrix, resulted in percentages of performance metrics as: accuracy 91.41%, precision 92.38%, F1 score 94.63% and recall 97.00%. The fifth combination model (S.No.5 in Table 5) uses original features along with manual weightings and threshold adjustment maintains with the original model's (S.No.1 in Table 4) performance. The sixth combination model (S.No.6 in Table 5) utilizing a cost matrix and threshold adjustment with original features, emerges as the best ML predicting algorithm model, resulted in maximum percentages for various metrics as: accuracy 92.19%, recall 98.00%, F1 score 95.15%, and MCC 76.08%, though its AUC is slightly lower at 90.39%. By comparing the last three combination model cases, the effectiveness of the cost matrix in models highlights its importance in improving recall by reducing the FN by very good amount.

Table 4. Performance metrics of the baseline model with individual FN methodologies

S. No.	Category	No. of Features	Confusion Matrix				Performance Metrics						
			TN	FP	FN	TP	ACC	P	S	SP	F1	MCC	AUC
1	Baseline model	36	20	8	6	94	89.06%	92.16%	94.00%	71.43%	93.07%	67.23%	92.04%
2	Engineered features	46	21	7	7	93	89.06%	93.00%	93.00%	75.00%	93.00%	68.00%	91.00%
3	Cost-sensitive learning	36	20	8	3	97	91.41%	92.38%	97.00%	71.43%	94.63%	73.68%	90.39%
4	Manual weight adjustment	36	20	8	6	94	89.06%	92.16%	94.00%	71.43%	93.07%	67.23%	92.04%
5	Threshold adjustment	36	20	8	6	94	89.06%	92.16%	94.00%	71.43%	93.07%	67.23%	92.04%

Table 5. Performance metrics of the baseline model with combination of different FN methodologies

S. No.	Category	No. of Features	Confusion Matrix				Performance Metrics						
			TN	FP	FN	TP	ACC	P	S	SP	F1	MCC	AUC
1	Engineered features + Manual weight adjustment	46	20	8	4	96	90.63%	92.31%	96.00%	71.43%	94.12%	71.42%	91.00%
2	Engineered features + Threshold adjustment	46	20	8	4	96	90.63%	92.31%	96.00%	71.43%	94.12%	71.42%	91.00%
3	Cost-sensitive + Engineered features	46	21	7	7	93	89.06%	93.00%	93.00%	75.00%	93.00%	68.00%	90.63%
4	Manual weight + Cost-sensitive	36	20	8	3	97	91.41%	92.38%	97.00%	71.43%	94.63%	73.68%	90.39%
5	Threshold + Manual weight	36	20	8	6	94	89.06%	92.16%	94.00%	71.43%	93.07%	67.23%	92.04%
6	Cost-sensitive + Threshold adjustment	36	20	8	**2**	98	**92.19**%	92.45%	**98.00**%	71.43%	**95.15**%	**76.08**%	90.39%

In the techniques reported in the literature [14], [16], [27] good prediction accuracy for the respective dataset utilized were achieved either by building a new sophisticated ML algorithm or developing a deeper network. In particular, very few of them are concentrating on reduction of type–I and type–II errors [14]. In this work, we incorporated various methodologies to focus mainly on reduction of FN as the model almost picked out those patients who are having heart disease. From Table 4 and 5, it is evident that the model is well suited for selecting methodologies like cost-sensitive and threshold adjustment as FN are reduced to a count of only 2 patients which resulted in an accuracy improvement from 89.06% to 92.19%. This proves that the proposed idea of different methodologies can work as a better ML framework to provide efficient prediction model with reduction of type-I and type-II errors. This result demonstrates the maximum clinical value by successfully minimizing life-threatening misclassifications while preserving high diagnostic precision, highlighting the optimal trade-off achieved by integrating multiple error reduction strategies. It is important to note that the reported false negative (FN) count of 2 was obtained from the 30% test set, corresponding to approximately 128 patients. This count reflects the model's evaluation performance on held-out data and not the entire dataset.

5. CONCLUSION

In this research study, a stacked ensemble-based ML classifier model is developed for coronary artery disease prediction of JIPMER out-patient dataset with high accuracy. The research work reported in this paper is focused on reducing the FN to improve the obtained prediction accuracy. This is achieved by

tuning or modifying the stacked ensemble model with FN reduction methodologies. Experimental results showed a clear reduction in FN count and improved performance across precision, recall, F1-score, and AUC metrics. Notably, combining cost-sensitive learning and threshold adjustment reduced FN to just two cases in the 30% test set, with a corresponding F1-score of 95.15% and recall of 98% highlighting the clinical impact of the proposed methodology. Any dataset having similar features like the JIPMER medical record dataset considered in this research can utilize our proposed ML model with incorporation of suitable FN reducing methodologies for accurate prediction.

This work has substantial implications for real-time decision support in cardiology, especially in resource-limited settings. The developed model is now being prepared for deployment within the clinical workflow at JIPMER, with plans for real-world validation and expansion to other cardiovascular conditions. Future research will focus on integrating this predictive framework with deep learning-based segmentation models using angiography images, thereby advancing toward a unified and automated diagnostic pipeline. Additionally, the proposed model can be further enhanced by incorporating recent developments in hybrid and cost-based ensemble learning techniques to improve predictive accuracy and clinical adaptability.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Akinchan Bhardwaj, Former Senior Residents, Department of Cardiology, JIPMER, Puducherry for extending his support towards creation of the database used in this study. We also like to thank the JIMPER Institute Ethics Committee for permitting us to use their patients' data for this research study.

FUNDING INFORMATION

No funds, grants, or other support was received.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Santhosh Gupta Dogiparthi	✓	✓	✓					✓	✓	✓				
Jayanthi K	✓					✓		✓		✓		✓	✓	
Ajith Ananthakrishna Pillai				✓			✓			✓			✓	
K Nakkeeran		✓								✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors reveals that they have no conflicts of interest to disclose.

INFORMED CONSENT

Being a retrospective study, the informed consent for this research study is waived off, as anonymized patient data has been provided. Henceforth ethics approval is sufficient for this research study.

ETHICAL APPROVAL

Ethical approval for this study was obtained from the Institute Ethics Committee of JIPMER (IEC Approval No. 1087/2019/OBS, dated 12th August 2022). The committee granted permission for the use of patient data in this research.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, DSG. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] S. Hay *et al.*, "India: Health of the Nation's States — The India state-level disease burden initiative," 2017. doi: 10.13140/RG.2.2.26119.88489.

[2] Worldometer, "India population (live)," *Worldometer*. 2024. [Online]. Available: https://www.worldometers.info/world-population/india-population/

[3] R. Deo, S. Desai, S. Behra, C. Pulate, A. Yadav, and N. Powar, "Threshold selection for classification models in prognostics," *PHM Society European Conference*, vol. 8, no. 1, Jun. 2024, doi: 10.36001/phme.2024.v8i1.4139.

[4] I. Araf, A. Idri, and I. Chairi, "Cost-sensitive learning for imbalanced medical data: a review," *Artificial Intelligence Review*, vol. 57, no. 4, Mar. 2024, doi: 10.1007/s10462-023-10652-8.

[5] H. Lee and P. Tsoi, "Feature-enhanced machine learning for all-cause mortality prediction in healthcare data," *arXiv:2503.21241*, Mar. 2025.

[6] E. Zvuloni, J. Read, A. H. Ribeiro, A. L. P. Ribeiro, and J. A. Behar, "On merging feature engineering and deep learning for diagnosis, risk prediction and age estimation based on the 12-lead ECG," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 7, pp. 2227–2236, Jul. 2023, doi: 10.1109/TBME.2023.3239527.

[7] P. Theerthagiri and J. Vidya, "Cardiovascular disease prediction using recursive feature elimination and gradient boosting classification techniques," *Expert Systems*, vol. 39, no. 9, Jun. 2022, doi: 10.1111/exsy.13064.

[8] B. T. Pham *et al.*, "Ensemble machine learning models based on reduced error pruning tree for prediction of rainfall-induced landslides," *International Journal of Digital Earth*, vol. 14, no. 5, pp. 575–596, May 2021, doi: 10.1080/17538947.2020.1860145.

[9] S. Sharma, R. Nayak, and A. Bhaskar, "Multi-view feature engineering for day-to-day joint clustering of multiple traffic datasets," *Transportation Research Part C: Emerging Technologies*, vol. 162, p. 104607, May 2024, doi: 10.1016/j.trc.2024.104607.

[10] M. Abbasi, S. Lopez Florez, A. Shahraki, A. Taherkordi, J. Prieto, and J. M. Corchado, "Class imbalance in network traffic classification: An adaptive weight ensemble-of-ensemble learning method," *IEEE Access*, vol. 13, pp. 26171–26192, 2025, doi: 10.1109/ACCESS.2025.3538170.

[11] M. E. Almandouh, M. F. Alrahmawy, M. Eisa, M. Elhoseny, and A. S. Tolba, "Ensemble based high performance deep learning models for fake news detection," *Scientific Reports*, vol. 14, no. 1, p. 26591, Nov. 2024, doi: 10.1038/s41598-024-76286-0.

[12] M. Vasconcelos and L. Cavique, "Mitigating false negatives in imbalanced datasets: An ensemble approach," *Expert Systems with Applications*, vol. 262, 2025, doi: 10.1016/j.eswa.2024.125674.

[13] D. Khanna, R. Sahu, V. Baths, and B. Deshpande, "Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease," *International Journal of Machine Learning and Computing*, vol. 5, no. 5, pp. 414–419, Oct. 2015, doi: 10.7763/ijmlc.2015.v5.544.

[14] E. Maini, B. Venkateswarlu, B. Maini, and D. Marwaha, "Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India," *Medical Journal Armed Forces India*, vol. 77, no. 3, pp. 302–311, Jul. 2021, doi: 10.1016/j.mjafi.2020.10.013.

[15] I. A. Zriqat, A. M. Altamimi, and M. Azzeh, "A comparative study for predicting heart diseases using data mining classification methods," *Preprint, arXiv:1704.02799*, 2017.

[16] S. V. Babu, P. Ramya, and J. Gracewell, "Revolutionizing heart disease prediction with quantum-enhanced machine learning," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, doi: 10.1038/s41598-024-55991-w.

[17] A. Mehta, P. Sengupta, D. Garg, H. Singh, and Y. S. Diamand, "Benchmarking the effectiveness of classification algorithms and SVM kernels for dry beans," *arXiv:2307.07863*, 2023.

[18] S. Niazmardi, "Cluster-based random radial basis kernel function for hyperspectral data classification," *arXiv:2409.05013*, Sep. 2024.

[19] M. T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, "Heart disease risk prediction using deep learning techniques with feature augmentation," *Multimedia Tools and Applications*, vol. 82, no. 20, pp. 31759–31773, Mar. 2023, doi: 10.1007/s11042-023-14817-z.

[20] X. Zhu, B. Xie, Y. Chen, H. Zeng, and J. Hu, "Machine learning in the prediction of in-hospital mortality in patients with first acute myocardial infarction," *Clinica Chimica Acta*, vol. 554, p. 117776, Feb. 2024, doi: 10.1016/j.cca.2024.117776.

[21] S. Gamil, F. Zeng, M. Alrifaey, M. Asim, and N. Ahmad, "An efficient AdaBoost algorithm for enhancing skin cancer detection and classification," *Algorithms*, vol. 17, no. 8, p. 353, Aug. 2024, doi: 10.3390/a17080353.

[22] R. H. Laftah and K. H. K. Al-Saedi, "Explainable ensemble learning models for early detection of heart disease," *Journal of Robotics and Control (JRC)5*, vol. 5, no. 5, pp. 1412–1421, 2024, doi: 10.18196/jrc.v5i5.22448.

[23] Q. Zhenya and Z. Zhang, "A hybrid cost-sensitive ensemble for heart disease prediction," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, Jan. 2021, doi: 10.1186/s12911-021-01436-7.

[24] C. Esposito, G. A. Landrum, N. Schneider, N. Stiefl, and S. Riniker, "GHOST: Adjusting the decision threshold to handle imbalanced data in machine learning," *Journal of Chemical Information and Modeling*, vol. 61, no. 6, pp. 2623–2640, Jun. 2021, doi: 10.1021/acs.jcim.1c00160.

[25] G. A. Diamond and J. S. Forrester, "Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease," *New England Journal of Medicine*, vol. 300, no. 24, pp. 1350–1358, Jun. 1979, doi: 10.1056/NEJM197906143002402.

[26] M. K. Ali, K. V. M. V. Narayan, and N. Tandon, "Diabetes & coronary heart disease: current perspectives," *Indian Journal of Medical Research*, vol. 132, no. 5, pp. 584–597, Nov. 2010, doi: 10.4103/ijmr.2010_132_05_584.

[27] C. Gupta, A. Saha, N. V Subba Reddy, U. Dinesh Acharya, N. V. S. Reddy, and U. D. Acharya, "Cardiac disease prediction using supervised machine learning techniques," *Journal of Physics: Conference Series*, vol. 2161, no. 1, p. 012013, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012013.

BIOGRAPHIES OF AUTHORS

Santhosh Gupta Dogiparthi (B.Tech, M.Tech, SMIEEE, LMISTE) received his under graduation from Amara Institute in 2012 and master's degree from Narasaraopeta Engineering College in 2014 from the stream of electronics and communication engineering. He is a part-time research scholar in the department of electronics and communication engineering at Puducherry Technological University, Puducherry. Currently, he is working as an assistant professor in the Department of Electronics and Communication Engineering at JNTUK University College of Engineering, Narasaraopet, Andhra Pradesh. His research interests include biomedical signal processing, image processing and prediction methodologies. He can be contacted at email: santhoshgupta@ptuniv.edu.in.

Jayanthi K. (B.E, M.Tech, Ph.D, MIEEE, MISTE) completed her under graduation from Madras University in 1997 and obtained her master's and doctorate degrees from Pondicherry Central University in 1999 and 2007, respectively. She has held various administrative positions as associate dean (research), associate dean (examination) and associate director (academic courses) since 2014 and is currently serving as a professor in the department of electronics and communication engineering at Puducherry Technological University. She has been actively involved in interdisciplinary research in the contemporary areas like AI in medical imaging and medical data analytics, IoT in water management and smart transportation system, artificial intelligence in 5G wireless communication systems and networks. She can be contacted at email: jayanthi@ptuniv.edu.in.

Ajith Ananthakrishna Pillai (MD, DM, FIC (Germany) FRCP (Lon) FRCP (Edin) FRCP (Glasgow) FACC, FSCAI) is a very popular and leading interventional cardiologist in India with vast experience in complex structural and coronary procedures. Trained in interventional cardiology in multiple international reputed centers like CVC in Frankfurt, UCLA in Los Angeles, Emory Heart in Atlanta, Karolinska in Sweden, ASAN in Seoul and Toyohashi in Japan. He is an accomplished teacher in the field of coronary and structural field. He is currently a professor and Head of Cardiology Department at Kauvery Hospital, Chennai. He can be contacted at email: ajithanantha@gmail.com.

K. Nakkeeran (Ph.D., FInstP, FIET, CEng, SMIEEE, SMOptica) received the B.Eng. degree from the Coimbatore Institute of Technology, Coimbatore, India, in 1993, and the M.Tech. and Ph.D. degrees from Anna University, Chennai, India, in 1995 and 1998, respectively. In 1999, he joined the Institute of Mathematical Sciences, Chennai, where he was a post-doctoral fellow for ten months. In 1999, he became a research associate with the department of physics, University of Burgundy, Dijon, France. In 2002, he became a post-doctoral fellow with the department of electronic and information engineering, The Hong Kong Polytechnic University. In 2005, he joined the School of Engineering, University of Aberdeen, Aberdeen, U.K., where he has been a senior lecturer since 2011. His research interests include solitons, fibre lasers, fibre sensors, modelling and simulations of optical devices, long-haul optical fibre communications, and nonlinear science. He can be contacted at email: k.nakkeeran@abdn.ac.uk.