Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Long-Term Prediction of Coronary Artery Disease via Ensemble Machine Learning Algorithms

Şehmus ASLAN[1*]

[1] Mardin Artuklu University, Business Administration Department, Mardin, sehmusaslan@artuklu.edu.tr, ORCID: 0000-0003-1886-3421

ARTICLE INFO

Article history:

Received 12 August 2024
Received in revised form 12 October 2024
Accepted 24 October 2024
Available online 23 December 2024

Keywords:

machine learning, coronary artery disease prediction, class imbalance, SMOTE, stackingC classifier.

Doi: 10.24012/dumf.1531523

* Corresponding author

ABSTRACT

Coronary artery disease (CAD) is the leading cause of death worldwide, necessitating early detection methods that are non-invasive, cost-effective, and reliable. In this study, the effectiveness of various machine learning (ML) models in predicting CAD was evaluated, with a focus on addressing class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). The Framingham CAD dataset was utilized, and SMOTE was applied with different k-values to balance the data, examining the impact on prediction performance. Eight significant features—age, diaBP, glucose, heart rate, sysBP, totChol, cigsPerDay, and BMI—were determined during preprocessing and used for further analysis. Among the models tested, the StackingC classifier demonstrated superior performance, achieving an accuracy of 95.81%, sensitivity of 95.9%, specificity of 95.7%, and an AUROC of 99.2% for k=1. These findings highlight the potential of the StackingC model as a robust tool for CAD prediction, offering a promising non-invasive method for early diagnosis.

Introduction

According to the WHO report [1], ischemic heart disease, also known as coronary heart disease, was the primary cause of death worldwide in 2019. It accounted for 16% of all deaths and experienced the most significant rise in fatalities since 2000; resulting in 8.9 million deaths in 2019 (see Fig. 1). Early detection of coronary artery disease (CAD) is essential to prevent an increased risk. Coronary angiography is the definitive diagnostic tool for CAD but is invasive and can result in complications like artery dissection, arrhythmia, and even death. Additionally, image-based detection methods are expensive and impractical for screening large populations, particularly in developing countries. Due to these limitations and the potentially dangerous nature of angiography, researchers have been seeking noninvasive, cost-effective, fast, and reliable methods for early CAD detection. Machine learning (ML) algorithms are among the techniques being explored for this purpose [2].

ML methodologies have shown significant potential in accurately diagnosing heart disease, improving the efficiency of medical professionals, and generating economic benefits. As datasets expand and ML algorithms advance, the impact of ML applications on automated heart disease prediction is expected to increase substantially. On the other hand, class imbalance issues are common in medical diagnostics, where majority class samples outnumber minority class samples. Traditional classification algorithms often misclassify minority observations as noise, leading to lower accuracy for the minority class. Addressing this imbalance is essential for improving the overall diagnostic performance of ML algorithms. In order to tackle this problem, the synthetic minority oversampling technique (SMOTE) algorithm is used in this study and efforts are made to improve the performance of SMOTE. This study has two primary contribution to the existing literature:

- The first is to determine the impact of the k hyper parameter in the K-Nearest Neighbors (KNN) classifier used for SMOTE on prediction performance.
- The second is to evaluate the effectiveness of various machine learning models in predicting CAD, aiming to identify the optimal ML model for coronary artery disease prediction.

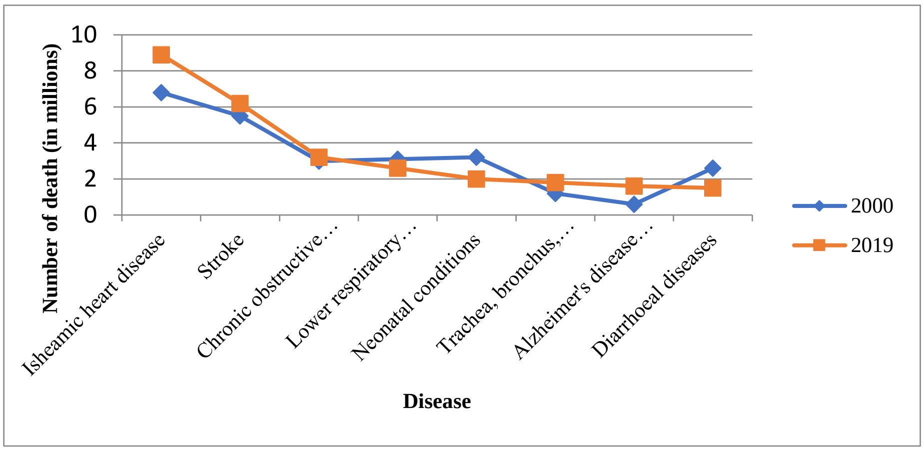

Figure 1. Leading causes of death globally

The remainder of the paper is structured as follows: Section 2 presents the literature review. Section 3 describes the dataset and outlines the proposed methodology. Section 4 discusses the experimental results and provides an analysis. Finally, Section 5 covers the conclusion and potential future work.

Literature Review

Up to now, a substantial number of studies have been conducted to predict CAD, extensively applying machine learning. Notably, Alizadehsani et al. [2] and Kutiame et al. [3] have provided comprehensive reviews on CAD prediction. Kurt et al. [4] compared various classification techniques for CAD prediction in 1245 subjects, finding the multi-layer perceptron (MLP) to be the best predictor with an and area under the receiver operating characteristic curve (AUROC) of 0.783, outperforming logistic regression (LR), classification and regression trees (CART), radial basis function (RBF), and self-organizing feature maps (SOFM). Babaoğlu et al. [5] explored principal component analysis (PCA) with support vector machines (SVM) for assessing exercise stress tests. The results showed that reducing features with PCA improved SVM model accuracy and reduced errors. Alizadehsani et al. [6] introduced the Z-Alizadeh Sani dataset and achieved a 94.08% accuracy in diagnosing CAD using effective features and data mining methods.

Furthermore, Akila and Chandramathi [7] proposed a hybrid method involving decision trees (C4.5) and MLP, achieving classification accuracies of 98.66% for C4.5 and 96.66% for MLP. Lo et al. [8] combined four heart disease datasets, using seven machine learning methods to predict CAD, and developed a new classifier with optimal performance using the TOPSIS algorithm. Miao et al. [9] developed an ensemble machine learning model using adaptive boosting, achieving accuracies between 77.78% and 96.72% across multiple datasets. Alizadehsani et al. [10] also aimed for high accuracy in diagnosing major coronary arteries stenosis, achieving rates of 86.14% for the LAD artery, 83.17% for the LCX artery, and 83.50% for the RCA, which are the highest reported in the literature.

Forssen et al. [11] systematically evaluated machine learning methods like L1 regression and random forest (RF) classifiers for CAD prediction using metabolomic data. Beunza et al. [12] compared machine learning algorithms on the Framingham Heart Study data, finding artificial neural networks (ANN) achieved the highest AUROC in R-Studio (0.71) and SVMs performed best in RapidMiner (0.75). Abdar et al. [13] tested ten traditional algorithms, ultimately selecting three types of SVMs optimized using genetic algorithms and particle swarm optimization, achieving a 93.08% accuracy and 91.51% F1-score with the optimized N2Genetic-nuSVM model on the Z-Alizadeh Sani dataset. Likewise, Dahal and Gautam [14] implemented five supervised classification methods for CAD prediction and determined the best model through performance comparison.

Dipto et al. [15] developed a prototype system using various machine learning algorithms to identify the most suitable model for a clinical dataset, finding the ANN to achieve the highest accuracy. Additionally, Dutta et al. [16] introduced a two-layer convolutional neural network (CNN) for classifying imbalanced clinical data, achieving high accuracy and specificity. Joloudari et al. [17] enhanced CAD diagnosis accuracy through an integrated machine learning approach, utilizing random trees (RT), C5.0 decision trees, SVM, and chi-squared automatic interaction detection (CHAID) decision trees. Muhammad et al. [18] developed machine learning models for CAD prediction using data from hospitals in Nigeria, with the random forest model achieving the highest accuracy and AUROC. Wang et al. [19] introduced a cloud-random forest (C-RF) model for assessing CAD risk. Using the Framingham dataset from Kaggle, the model achieved 85% accuracy and demonstrated superior performance in classification accuracy, error rates, and AUROC value compared to CART, SVM, CNN, and RF models. Wang et al. [20] enrolled 3,112 CAD patients and 3,182 controls

Table 1. Summary of literature review on CAD prediction using ML techniques

Authors	**Method Used**	**Dataset**	**Class Imbalance Handling**
Kurt et al. [4]	Logistic Regression, CART, Neural Networks	Trakya University dataset (1245 samples)	No specific focus on class imbalance
Alizadehsani et al. [6]	SVM with Feature Selection and Information Gain	Z-Alizadeh Sani (303 samples)	No specific focus on class imbalance
Alizadehsani et al. [10]	SVM with Kernel Fusion, Information Gain, Feature Selection	Z-Alizadeh Sani (303 samples)	No specific focus on class imbalance
Lo et al. [8]	Ensemble Learning, Co-Expressed Observations, TOPSIS	UCI Machine Learning Repository (822 samples)	No specific focus on class imbalance
Miao et al. [9]	Adaptive Boosting Algorithm, Ensemble Learning	UCI Machine Learning Repository (920 samples)	No specific focus on class imbalance
Forssen et al. [11]	L1 regression, Random Forest	4C Metabolomics dataset (3409 samples)	No specific focus on class imbalance
Beunza et al. [12]	SVM, Neural Networks, Decision Trees, Random Forest	Framingham Heart Study (4240 samples)	Applied oversampling to balance dataset
Dahal and Gautam [14]	SVM, LR, Bagging CART, RF, KNN	Z-Alizadeh Sani (303 samples)	Applied SMOTE to balance dataset
Dipto et al. [15]	Logistic Regression, SVM, ANN	Z-Alizadeh Sani dataset (303 samples)	Applied SMOTE to balance dataset
Dutta et al. [16]	CNN with LASSO	NHANES dataset (37,079 samples)	Addressed class imbalance with CNN structure
Masih et al. [21]	Multilayer Perceptron-based DNN	Framingham Heart Study (4583 samples)	Applied SMOTE to balance dataset
Wang et al. [20]	Cloud-Random Forest	Framingham Heart Study (4238 samples)	Applied SMOTE to balance dataset
Trigka et al. [22]	Stacking Ensemble with SMOTE	Framingham Heart Study (4238 samples)	Applied SMOTE to balance dataset
Mishra et al. [23]	SVM, Logistic Regression, Random Forest, Decision Tree	Various hospitals in Odisha (3254 samples)	No specific focus on class imbalance
Saeedbakhsh et al. [23]	SVM, ANN, Random Forest	Isfahan Cohort Study dataset (11495 samples)	No specific focus on class imbalance
This Study (2024)	**StackingC**	**Framingham CAD dataset (4238 samples)**	**SMOTE with varying k-values**

from three centers in China, developing a RF model. This model achieved an AUROC of 0.948 in the development cohort and 0.944 and 0.940 in two validation cohorts. An easy-to-use tool combining 15 indexes was also created for clinical application, enhancing CAD management and prevention. Trigka and Dritsas [22] evaluated various machine learning models with and without SMOTE, finding the stacking ensemble model combined with SMOTE and 10-fold cross-validation to outperform others. Huang and Huang [24] used the NHANES dataset to identify CAD risk factors with machine learning, achieving an AUROC of 0.89 with the XGBoost model. Özbilgin et al. [25] proposed a non-invasive CAD diagnosis method using iris images and SVM, achieving a 93% accuracy rate. Saeedbakhsh et al. [23] used SVM, ANN, and RF to predict CAD with high accuracy, identifying key predictors and extracting eleven high-confidence rules. Li et al. [26] proposed a deep learning hybrid model for CAD prediction, significantly improving accuracy and addressing overfitting issues. Table 1 presents a summary of literature review on CAD prediction using ML techniques

After analyzing the related literature, it is evident that various studies have explored the use of machine learning models for coronary artery disease (CAD) prediction, employing methods such as support vector machines (SVM), random forests, and convolutional neural networks (CNN), demonstrating moderate to high accuracy and AUROC scores. However, a significant gap exists in addressing the issue of class imbalance, a common problem in medical datasets where CAD-positive cases are underrepresented. This imbalance often leads to reduced model performance, particularly in predicting minority class instances, which are critical in healthcare applications. Additionally, while previous research has shown the potential of ensemble models, like stacking classifiers, to enhance performance by leveraging the strengths of multiple models, there has been limited focus on optimizing these models for CAD prediction. Another unexplored area in the literature is the effect of different SMOTE (Synthetic Minority Oversampling Technique) k-values on model performance, which can have a substantial impact on how well machine learning models manage imbalanced datasets.

This study addresses these gaps by applying SMOTE to balance the dataset and experimenting with various k-values to optimize model performance. Moreover, it evaluates a stacking ensemble model (StackingC), which combines multiple classifiers and demonstrates superior performance over traditional machine learning models. The findings reveal that StackingC consistently outperforms other models, achieving high accuracy, sensitivity, specificity, and AUROC, making it a robust and reliable tool for CAD prediction, especially in the context of imbalanced datasets. This research not only advances the application of ensemble models for CAD prediction but also provides valuable insights into how tuning the k-parameter of SMOTE can enhance model performance, effectively addressing key gaps identified in the literature.

Material and Method

This section outlines the approach used in the study, detailing the materials, datasets, and methodologies employed to predict coronary artery disease (CAD) risk using machine learning techniques.

Dataset description

This research paper utilizes the Framingham CAD dataset, available on Kaggle [27], to conduct an empirical examination of CAD risk assessment methods. The dataset originates from a continuous cardiovascular study involving residents of Framingham, Massachusetts. The objective is to predict whether a patient will experience coronary heart disease within the following ten years.

The Framingham CAD dataset encompasses 15 risk indicators associated with CAD, with a collective sample size of 4238 participants, of whom 644 (15.2%) were diagnosed with CAD. Detailed statistical information regarding the features concerning the target class labels is provided in Table 2 and 3. The Framingham CAD dataset was chosen for several reasons. First, it includes well-established risk factors for CAD, such as age, systolic and diastolic blood pressure, cholesterol levels, glucose, and smoking habits, all of which are critical in predicting the disease. These features have been widely used in cardiovascular research and are known to provide reliable insight into coronary artery disease. The dataset's breadth and depth allow for a comprehensive analysis of the disease risk, making it a suitable foundation for machine learning-based predictive modeling.

Second, the dataset's public availability on platforms like Kaggle ensures transparency and reproducibility, which are essential in scientific research. Its accessibility allows researchers to validate findings and develop comparable models, increasing the robustness of machine learning approaches in CAD prediction. Additionally, despite its initial class imbalance (with 15.2% of participants diagnosed with CAD), the dataset offers a sufficient sample size to apply techniques like Synthetic Minority Oversampling Technique (SMOTE) [28]. SMOTE was used to balance the dataset, creating equal representation of CAD and non-CAD cases, which enhances the accuracy and reliability of the predictive models.

Overall, the Framingham CAD dataset provides an ideal combination of relevance, data quality, and availability. Its use of established risk factors and its large sample size, together with the applied class-balancing techniques, make it a robust choice for developing accurate and reliable machine learning models for early CAD detection.

Data Preprocessing

The number of participants who have been diagnosed with CAD is 644 (15.2%), indicating an imbalanced dataset. In accordance with the principle of balancing the quantity of CAD and non-CAD samples, SMOTE method [28] was employed to transform the Framingham dataset into a balanced dataset. SMOTE relies on the K-Nearest Neighbors (KNN) [29] classifier generating synthetic data for the minority class in order to achieve a uniform distribution across both classes. One of the objectives of this study is to observe the effects of the k hyper parameter of the KNN classifier used for SMOTE on the prediction performance. Hence, SMOTE was applied with 3 different k values which are 1, 3, and 5, resulting in the creation of three distinct datasets. Following the implementation of SMOTE, the datasets achieved balance, consisting of 7188 participants with an equal distribution of 3594 instances each for both CAD and non-CAD classes. Furthermore, in order to prevent a reduction in the observation count, any missing values within the continuous variables were filled by averaging the remaining non-missing values within that specific variable. Lastly, all attributes with numeric values were normalized within the range of [0, 1].

Feature selection

In this study, Random Forest (RF) [30] was utilized alongside 10-fold cross-validation to eliminate irrelevant features, which can degrade model performance. The combination of robustness, scalability, interpretability, and automatic feature selection makes RF a popular choice for selecting the best features in machine learning tasks. The main steps of feature selection using RF in this study are described as follows:

Step 1: RF was trained using the entire dataset with 10-fold cross-validation.

Step 2: Then, for each attribute in the dataset, we temporarily removed this attribute and trained the classifier again using the remaining attributes.

Step 3: After training the RF without each attribute, we evaluated the performance of the RF using, accuracy metric through cross-validation.

Step 4: The importance of each attribute was determined based on how much the performance of the RF degrades when that attribute is removed. If removing an attribute causes a significant drop in performance, it suggests that the attribute is important for classification.

Step 5: Finally, the attributes were ranked based on their importance scores. Attributes with higher scores were considered more important for classification, while attributes with lower scores might be deemed less important or even irrelevant.

After obtaining the rank of significance for the features (see Fig. 2), a decision was made regarding which features to use for the model. Starting from the feature with the lowest score, each feature was individually removed from the model, and RF classifier was run. The accuracy metric reached its maximum level with 8 features as it is shown Table 4. The 8 features that maximize the accuracy of the RF model are as follows: age, diaBP, glucose, heartRate sysBP, totChol, cigsPerDay, BMI. Hence, only these features were utilized for the remainder of the study

Machine learning models

In particular, the focus was on evaluating Bayesian Network (BN) [31] and the K-Star algorithm [32]. Ensemble ML algorithms were also explored, including Random Forest (RF) [30], Stacking [33], and StackingC [34]. After numerous preliminary experiments, the optimal hyper parameters for the algorithms were determined and are detailed in Table 5:

Table 2. Numerical features of Framingham dataset

Attribute	Description	Min	Max	Mean ± stdDev
Age	Age of the patient in years	32	70	49.5 ± 8.56
Cigs/day	Number of cigarettes smoked per day	0	70	9 ± 11.92
totChol	Total cholesterol level	113	464	236.8 ± 43.69
SysBP	Systolic blood pressure	83.5	295	132.3 ± 22.1
DiaBP	Diastolic blood pressure	48	142.5	82.9 ± 11.97
BMI	Body Mass Index	15.54	56.8	25.8 ± 4.07
Heart rate	Heart rate (beats per minute)	44	143	75.7 ± 11.99
Glucose	Glucose level	40	394	81.8 ± 23.89

Table 3. Nominal features of Framingham dataset

Attribute	Description	Categories	n
Gender	Gender of the patient	Male Female	1819 2419
Education	Level of education	Some high school High school graduate Some college or vocational school College graduate	1825 1253 687 473
Current smoker	Indicator of whether the participant is smoking	Yes No	2094 2144
BPMeds	Indicator of whether the participant is on blood pressure medication	Yes No	124 4114
prevStroke	Indicator of whether the participant has a history of stroke	Yes No	25 4213
prevHyp	Indicator of whether the participant has a history of hypertension	Yes No	1316 2922
Diabetes	Indicator of diabetes presence	Yes No	109 4129

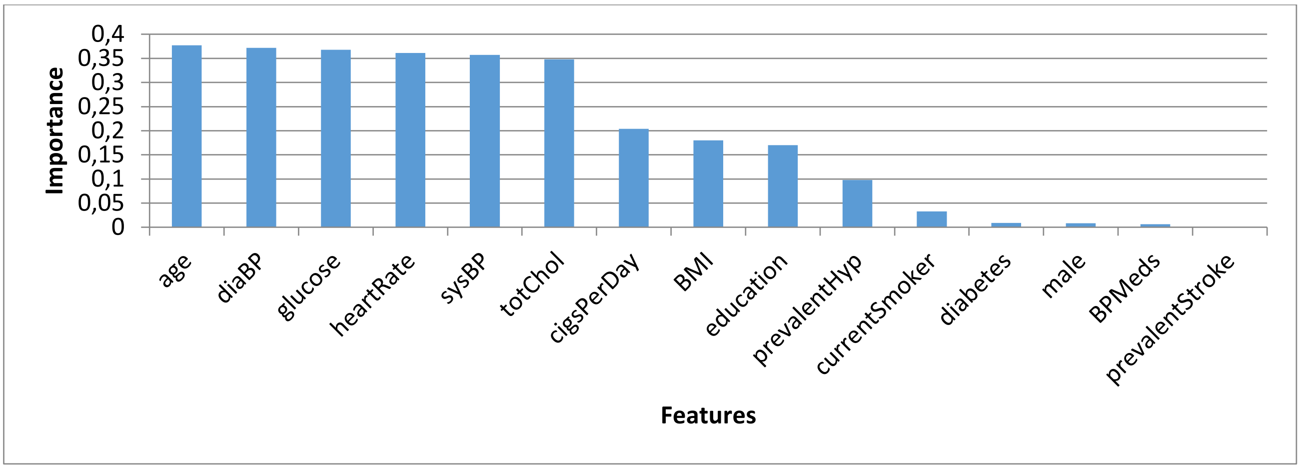

Figure 2. Selection of important influencing factors of CAD using RF

Table 4. Performance evaluation related to the number of features

Feature Number	6	7	**8**	9	10	11	12	13	14	15
Accuracy (%)	90.16	90.33	**90.35**	90.16	89.89	89.83	89.84	89.79	89.72	89.72

Table 5. Optimal hyper parameters for ML models for the Framingham dataset

Random Forest (RF)	numIterations numFeatures	300 5
K-Star (K*)	globalBlend	30
Bayesian Network (BN)	estimator searchAlgorithm	SimpleEstimator K2
Stacking	classifiers metaClassifier	RF, K*, BN K*
StackingC	classifiers metaClassifier	RF, K*, BN K*

Experimental Results

This section presents the experimental results of the study, detailing the evaluation metrics and test outcomes of the machine learning models used for coronary artery disease (CAD) prediction.

Evaluation Metrics

To assess the performance of machine learning models, a confusion matrix was utilized. This matrix showcases the disparities between actual and predicted classes. Each row of the confusion matrix corresponds to instances in the predicted class, while each column represents instances in the real class, and vice versa. Typically, the confusion matrix comprises four distinct terms: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). True Positive (TP) denotes the number of predicted values correctly identifying the presence of disease, while True Negative (TN) represents the count of predicted values correctly identifying the absence of disease. False Positive (FP) indicates the count of predicted values incorrectly classified as positive (when they were actually negative), and False Negative (FN) refers to the count of predicted values incorrectly classified as negative (when they were actually positive). Using these terms, performance metrics such as accuracy, sensitivity, specificity, and Area Under the Receiver Operating Characteristic Curve (AUROC) were calculated. A description of each performance metric is provided below.

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (3)$$

In this study the performance of the ML models' was assessed using accuracy, specificity, sensitivity, and AUROC values.

Experimental setup

The algorithms were implemented in WEKA 3.9.6 and run on a PC with a 2.3 GHz AMD Ryzen 5 processor and 8 GB of RAM with Windows 11. A 10-fold cross-validation was applied to measure the efficiency of the machine learning models on the balanced dataset of 7188 instances after SMOTE, for each of the three values of k: 1, 3, and 5.

Test results

In this paper, five machine learning techniques were compared: Bayesian Network (BN) [31], K-Star [32], Random Forest (RF) [30], Stacking [33], and StackingC [34]. The comparison was performed on Framingham dataset balanced by SMOTE [28] using three different values of the k hyper parameter: 1, 3, and 5.

Table 6. Performance evaluation of ML models for k=1

	Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**AUROC (%)**
BN	90.35	83.4	97.2	93.5
K-Star	84.24	98.3	70.2	97.9
RF	90.43	90.8	90.1	96.5
Stacking	95.92	95.4	96.5	99.1
StackingC	95.81	95.9	95.7	99.2

Table 7. Performance evaluation of ML models for k=3

	Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**AUROC (%)**
BN	90.58	82.8	98.4	93.6
K-Star	81.15	94.3	68.0	93.9
RF	87.54	89.1	86	94.7
Stacking	95.08	93.9	96.3	98.6
StackingC	95.23	94.3	96.1	98.7

Table 8. Performance evaluation of ML models for k=5

	Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**AUROC (%)**
BN	88.97	82.2	95.7	93.1
K-Star	78.85	91.5	66.2	90.9
RF	86.14	87.7	84.6	93.7
Stacking	94.23	92.7	95.8	98.1
StackingC	93.93	93.2	94.6	98.2

According to Table 6, 7, 8 the k hyper parameter in SMOTE influences the performance of the ML models. Generally, lower values of k yield higher performance metrics, while higher values of k show a decrease in these metrics. Considering the overall performance across different values of the k hyper parameter in SMOTE, the Stacking and StackingC classifiers stand out as the best options for Framingham dataset due to their consistently high accuracy, sensitivity, specificity, and AUROC. However, StackingC has shown slightly better performance compared to Stacking. StackingC demonstrates high accuracy across all values of k (95.81% for k=1, 95.23% for k=3, and 95.23% for k=5). Furthermore, both sensitivity (95.9%) and specificity (95.7%) are high for k=1, suggesting that the model is equally good at identifying CAD cases and non-CAD cases. This balance is crucial for medical diagnosis to avoid both false alarms and missed diagnoses.

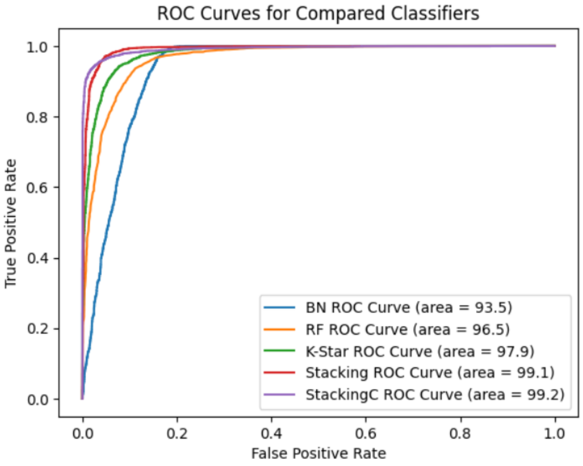

Figure 3. ROC curves comparing machine learning models at k=1 (SMOTE)

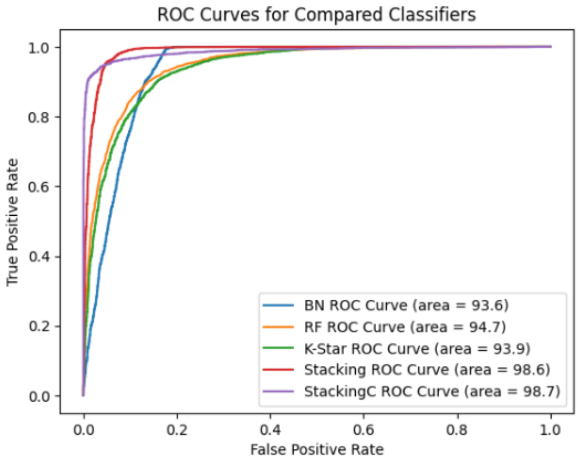

Figure 4. ROC curves comparing machine learning models at k=3 (SMOTE)

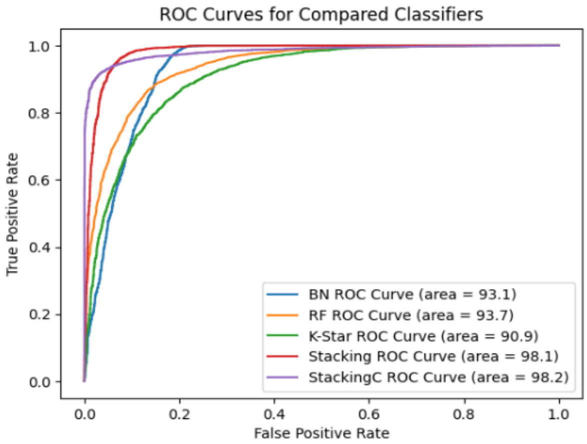

Figure 5. ROC curves comparing machine learning models at k=5 (SMOTE)

To provide a deeper understanding of the classification performance of the ML models, ROC curves are displayed in Figures 3, 4 and 5. The ROC curve is crucial for evaluating classifier performance because it provides a comprehensive view of the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) across different threshold settings. The AUROC values are consistently high for StackingC, indicating strong overall performance and reliability in distinguishing between positive and negative cases of CAD (99.2 for k=1, 98.7 for k=3, and 98.2 for k=5). StackingC gave the best performance for k=1 in SMOTE with 10-fold cross-validation, and the confusion matrix of this model is given in Table 9.

Table 9. Confusion matrix for stackingC using k=1 in SMOTE for Framingham dataset

	Predicted CAD	**Predicted non-CAD**	**Total**
Actual CAD	3446	148	3594
Actual non-CAD	153	3441	3594
Total	3599	3589	

Additionally, a comparative analysis between the current study and previous studies was performed using accuracy, specificity, sensitivity, and AUROC evaluation criteria, with the results presented in Table 10. Beunza et al. [12] utilized support vector machines (SVM) on the same Framingham dataset, achieving an accuracy of 69% and an AUROC of 75%, which is substantially lower than the results achieved by the StackingC model in this study. Similarly, Wang et al. [20] employed a cloud-random forest (C-RF) model on the Framingham dataset and achieved an accuracy of 85% and an AUROC of 85%, again demonstrating the superior performance of the StackingC model used in this paper.

In comparison, Trigka and Dritsas [22] applied a stacking ensemble model with SMOTE to predict CAD and achieved an accuracy of 90.9% and an AUROC of 96.1%. While this shows strong performance, the StackingC model in the current study outperformed it with an accuracy of 95.81% and AUROC of 99.2%, underscoring the effectiveness of the ensemble learning method used in this study.

Moreover, Masih et al. [21] applied a convolutional neural network (CNN) model, achieving an accuracy of 96.5% on the Framingham dataset. Although the CNN model achieved slightly higher accuracy, the StackingC model's AUROC of 99.2%, alongside its balanced sensitivity and specificity, makes it a more reliable and robust model for distinguishing between CAD and non-CAD cases.

These comparisons emphasize that while machine learning models such as SVM, random forests, stacking models, and CNNs have been widely used for CAD prediction, the StackingC model in this study demonstrates superior overall performance, particularly in terms of balancing accuracy, sensitivity, specificity, and AUROC. This balance is crucial for medical diagnostics, where both false positives and false negatives can have significant consequences. The StackingC model's exceptional performance across multiple evaluation metrics suggests it is a robust tool for CAD prediction, and its applicability could extend to clinical settings for early diagnosis.

In summary, although there have been numerous studies using machine learning techniques for CAD prediction, this study's StackingC model offers improved performance when compared to many previous approaches, making it a valuable contribution to the field.

Table 10. Comparison of the ML models from previous studies for Framingham dataset

Authors	Dataset	Method	Accuracy(%)	Sensitivity(%)	Specificity(%)	AUROC(%)
Beunza [12]	Framingham	SVM	69	42	84	75
Wang [20]	Framingham	C-RF	85	84	86	85
Trigka [22]	Framingham	Stacking	90.9	87.6	NaN	96.1
Masih [21]	Framingham	CNN	96.5	91.9	98.28	NaN
This study	Framingham	StackingC	95.81	95.9	95.7	99.2

Conclusion

In this study, it is aimed to enhance the prediction of coronary artery disease (CAD) using various machine learning models, focusing on the Synthetic Minority Oversampling Technique (SMOTE) and its impact on model performance. The results demonstrated that the application of SMOTE, particularly with lower k-values, significantly improved the performance metrics of the machine learning models. Among the models tested, the StackingC classifier consistently outperformed others, achieving high accuracy, sensitivity, specificity, and AUROC values across different k-values. This suggests that StackingC is highly effective in distinguishing between CAD and non-CAD cases, making it a reliable tool for medical diagnostics.

To optimize the predictive accuracy of the ML models, it was identified the eight most significant features using the Random Forest (RF) model. The features that maximize the accuracy of the RF model are age, diaBP, glucose, heart rate, sysBP, totChol, cigsPerDay, and BMI. Hence, only these features were utilized for the remainder of the study. This feature selection process was critical in enhancing the model's performance, ensuring that the most relevant and impactful variables were included in the prediction models.

This study’s findings highlight the importance of addressing class imbalance in medical datasets to enhance the accuracy and reliability of predictive models. The use of SMOTE in balancing the dataset proved crucial, particularly for models like StackingC, which showed superior performance in all evaluation metrics. The StackingC model achieved an accuracy of 95.81%, sensitivity of 95.9%, specificity of 95.7%, and an AUROC of 99.2% for k=1, underscoring its robustness and potential for clinical application. The StackingC model holds significant promise for real-world applications, particularly in healthcare settings where early and accurate detection of CAD is critical. This model could be integrated into clinical decision support systems to assist healthcare professionals in identifying high-risk patients. With its high sensitivity and specificity, the model can reliably distinguish between CAD-positive and CAD-negative cases, reducing the chances of both false positives and false negatives. This would be particularly valuable in resource-constrained environments where access to invasive diagnostic methods, such as coronary angiography, is limited. By offering a non-invasive, cost-effective, and rapid diagnostic tool, the StackingC model could improve early detection and intervention, potentially reducing mortality rates from CAD.

Despite its strong performance, the study has some limitations that should be acknowledged. First, the dataset used is the Framingham CAD dataset, which, while comprehensive, may not fully represent the diverse global population. The model’s performance might vary when applied to other populations with different risk factors and demographic characteristics. Additionally, the study primarily focused on balancing the dataset using SMOTE, but there could be other methods for handling class imbalance that might yield different results. The impact of hyper-parameter tuning and alternative resampling techniques, such as ADASYN or NearMiss, was not explored in this study.

Future research could address these limitations by validating the StackingC model on larger, more diverse datasets, potentially from different geographic regions and healthcare settings. Expanding the feature set to include other relevant risk factors, such as genetic data or lifestyle factors, could improve the model's generalizability. Additionally, exploring other machine learning techniques, such as deep learning models, or hybrid approaches that combine ensemble learning with neural networks, could further enhance performance.

Another promising direction for future research is the development of interpretable models that allow healthcare professionals to understand the underlying reasons for a model’s predictions. Explainability in AI models is crucial, especially in high-stakes fields like healthcare. Research could focus on incorporating explainability techniques like SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) to make the predictions more transparent to clinicians.

Ethics Committee Approval

There is no need to obtain permission from the ethics committee for the article prepared. There is no conflict of interest with any person / institution in the article prepared.

References

[1] World Health Organization, “The top 10 causes of death,” Jun. 09, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[2] R. Alizadehsani *et al.*, “Machine learning-based coronary artery disease diagnosis: A comprehensive review,” *Comput. Biol. Med.*, vol. 111, p. 103346, Aug. 2019, doi: 10.1016/j.compbiomed.2019.103346.

[3] S. Kutiame, R. Millham, A. F. Adekoya, M. Tettey, B. A. Weyori, and P. Appiahene, “Application of Machine Learning Algorithms in Coronary Heart Disease: A Systematic Literature Review and Meta-Analysis,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.0130620.

[4] I. Kurt, M. Ture, and A. T. Kurum, “Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease,” *Expert Syst. Appl.*, vol. 34, no. 1, pp. 366–374, Jan. 2008, doi: 10.1016/j.eswa.2006.09.004.

[5] I. Babaoğlu, O. Fındık, and M. Bayrak, “Effects of principle component analysis on assessment of coronary artery diseases using support vector machine,” *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2182–2185, Mar. 2010, doi: 10.1016/j.eswa.2009.07.055.

[6] R. Alizadehsani *et al.*, “A data mining approach for diagnosis of coronary artery disease,” *Comput. Methods Programs Biomed.*, vol. 111, no. 1, pp. 52–61, Jul. 2013, doi: 10.1016/j.cmpb.2013.03.004.

[7] S. Akila and S. Chandramathi, “A Hybrid Method for Coronary Heart Disease Risk Prediction using Decision Tree and Multi Layer Perceptron,” *Indian J. Sci. Technol.*, vol. 8, no. 34, Dec. 2015, doi: 10.17485/ijst/2015/v8i34/85947.

[8] Y.-T. Lo, H. Fujita, and T.-W. Pai, “PREDICTION OF CORONARY ARTERY DISEASE BASED ON ENSEMBLE LEARNING APPROACHES AND CO-EXPRESSED OBSERVATIONS,” *J. Mech. Med. Biol.*, vol. 16, no. 01, p. 1640010, Feb. 2016, doi: 10.1142/S0219519416400108.

[9] K. H. Miao, J. H., and G. J., “Diagnosing Coronary Heart Disease using Ensemble Machine Learning,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 10, 2016, doi: 10.14569/IJACSA.2016.071004.

[10] R. Alizadehsani *et al.*, “Coronary artery disease detection using computational intelligence methods,” *Knowl.-Based Syst.*, vol. 109, pp. 187–197, Oct. 2016, doi: 10.1016/j.knosys.2016.07.004.

[11] H. Forssen, R. Patel, N. Fitzpatrick, A. Timmis, H. Hemingway, and S. Denaxas, “Evaluation of Machine Learning Methods to Predict Coronary Artery Disease Using Metabolomic Data”.

[12] J.-J. Beunza *et al.*, “Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease),” *J. Biomed. Inform.*, vol. 97, p. 103257, Sep. 2019, doi: 10.1016/j.jbi.2019.103257.

[13] M. Abdar, W. Książek, U. R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak, “A new machine learning technique for an accurate diagnosis of coronary artery disease,” *Comput. Methods Programs Biomed.*, vol. 179, p. 104992, Oct. 2019, doi: 10.1016/j.cmpb.2019.104992.

[14] K. R. Dahal and Y. Gautam, “Argumentative Comparative Analysis of Machine Learning on Coronary Artery Disease,” *Open J. Stat.*, vol. 10, no. 04, pp. 694–705, 2020, doi: 10.4236/ojs.2020.104043.

[15] I. C. Dipto, T. Islam, H. M. M. Rahman, and M. A. Rahman, “Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease,” *J. Data Anal. Inf. Process.*, vol. 08, no. 02, pp. 41–68, 2020, doi: 10.4236/jdaip.2020.82003.

[16] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, “An efficient convolutional neural network for coronary heart disease prediction,” *Expert Syst. Appl.*, vol. 159, p. 113408, Nov. 2020, doi: 10.1016/j.eswa.2020.113408.

[17] J. H. Joloudari *et al.*, “Coronary Artery Disease Diagnosis; Ranking the Significant Features Using a Random Trees Model,” *Int. J. Environ. Res. Public. Health*, vol. 17, no. 3, p. 731, Jan. 2020, doi: 10.3390/ijerph17030731.

[18] L. J. Muhammad, I. Al-Shourbaji, A. A. Haruna, I. A. Mohammed, A. Ahmad, and M. B. Jibrin, “Machine Learning Predictive Models for Coronary Artery Disease,” *SN Comput. Sci.*, vol. 2, no. 5, p. 350, Sep. 2021, doi: 10.1007/s42979-021-00731-4.

[19] C. Wang *et al.*, “Development and Validation of a Predictive Model for Coronary Artery Disease Using Machine Learning,” *Front. Cardiovasc. Med.*, vol. 8, p. 614204, Feb. 2021, doi: 10.3389/fcvm.2021.614204.

[20] J. Wang, C. Rao, M. Goh, and X. Xiao, “Risk assessment of coronary heart disease based on cloud-random forest,” *Artif. Intell. Rev.*, vol. 56, no. 1, pp. 203–232, Jan. 2023, doi: 10.1007/s10462-022-10170-z.

[21] N. Masih, H. Naz, and S. Ahuja, “Multilayer perceptron based deep neural network for early detection of coronary heart disease,” *Health Technol.*, vol. 11, no. 1, pp. 127–138, Jan. 2021, doi: 10.1007/s12553-020-00509-3.

[22] M. Trigka and E. Dritsas, “Long-Term Coronary Artery Disease Risk Prediction with Machine Learning Models,” *Sensors*, vol. 23, no. 3, p. 1193, Jan. 2023, doi: 10.3390/s23031193.

[23] S. Saeedbakhsh, M. Sattari, M. Mohammadi, J. Najafian, and F. Mohammadi, “Diagnosis of

coronary artery disease based on machine learning algorithms support vector machine, artificial neural network, and random forest,” *Adv. Biomed. Res.*, vol. 12, no. 1, p. 51, 2023, doi: 10.4103/abr.abr_383_21.

[24] A. A. Huang and S. Y. Huang, “Use of machine learning to identify risk factors for coronary artery disease,” *PLOS ONE*, vol. 18, no. 4, p. e0284103, Apr. 2023, doi: 10.1371/journal.pone.0284103.

[25] F. Özbilgin, Ç. Kurnaz, and E. Aydın, “Prediction of Coronary Artery Disease Using Machine Learning Techniques with Iris Analysis,” *Diagnostics*, vol. 13, no. 6, p. 1081, Mar. 2023, doi: 10.3390/diagnostics13061081.

[26] F. Li, Y. Chen, and H. Xu, “Coronary heart disease prediction based on hybrid deep learning,” *Rev. Sci. Instrum.*, vol. 95, no. 1, p. 015115, Jan. 2024, doi: 10.1063/5.0172368.

[27] “The Framingham CAD dataset.” Accessed: Jun. 10, 2024. [Online]. Available: https://www.kaggle.com/datasets/navink25/framingham

[28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[29] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/BF00153759.

[30] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[31] M. Friedman, “The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance,” *J. Am. Stat. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937, doi: 10.1080/01621459.1937.10503522.

[32] J. G. Cleary and L. E. Trigg, “K*: An Instance-based Learner Using an Entropic Distance Measure,” in *Machine Learning Proceedings 1995*, Elsevier, 1995, pp. 108–114. doi: 10.1016/B978-1-55860-377-6.50022-0.

[33] D. H. Wolpert, “Stacked generalization,” *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[34] A. Seewald, *How to Make Stacking Better and Faster While Also Taking Care of an Unknown Weakness.* 2002, p. 561.