Diagnosing Coronary Artery Disease with Perfect Machine Learning Methods and SHAP Explainability

Hardeep Kaur
Student, Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India

Sandeep Kaur Dhanda
Assistant Professor, Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India

Simarjot Kaur
Assistant Professor, Department of Computer Science Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India

ABSTRACT
Diagnosis of Coronary Artery Disease (CAD)-one of the highest non-communicable diseases that are affecting millions of people in every corner of the world-requires suitable diagnostic instruments that are effective, precise, and easy to understand. The study proposes an early detection machine learning pipeline for CAD based on the Z-Alizadeh Sani dataset. This pipeline consists of domain-specific preprocessing, SMOTE-based class balancing, hybrid feature selection using RFECV and RFE trimming, and evaluation using several classifiers. XGBoost outperformed all models that were employed with a sensitivity of 0.9637 and ROC AUC of 0.9503. To increase the safety of the categorization, a clinically tuned threshold was used. SHAP analysis revealed the key variables to model openness, such as common chest pain, EFTTE, DM, and BMI. The suggested method stands superior with its diagnostic sensitivity and interpretability when compared with existing norms for clinical applicability.

Keywords
Coronary Artery Disease, Machine Learning, SHAP, XGBoost, Feature Selection, SMOTE, Medical Diagnosis.

1. INTRODUCTION
Coronary Artery Disease (CAD) is the prominent cardiovascular illness responsible for killing millions of people every year. It happens when coronary arteries are narrowed or completely blocked by atheromatous plaques, preventing blood flow to the heart muscle. According to the WHO, cardiovascular diseases, including CAD, accounted for 17.9 million deaths each year, the leading cause of global death [1]. That said, it tends to impact low-and middle-income countries the most because they usually do not have easy access to invasive diagnostic technologies. That's what is going to be further discussed in this topic [2].

Early Getting CAD diagnosis subscribes to the logic of decreasing hospital admissions, preventing heart attacks, and improving the overall quality of life. Traditional methods of diagnosis, like coronary angiography, are rather accurate but are invasive, expensive, and unsuitable for screening large populations [3]. Even non-invasive methods like stress testing and echocardiography require medical expertise and available facilities that render them impractical in mass-scale applications [4].

In this field, machine learning (ML) is proving to be an efficient tool to build non-invasive, cheap, and scalable diagnostic solutions. With high sensitivity and specificity, ML algorithms discover hidden patterns from complex clinical data and utilize them to accurately forecast early disease risk [5]. Over the previous decade, many machine learning techniques like neural networks, support vector machines, decision trees, and ensemble methods were used in healthcare, with good results in disease classification and forecasting [6]. However, many of the current computer-aided diagnosis systems have deficiencies in interpretability or do not exceed the sensitivity benchmarks.

To address these constraints, the study presents a comprehensive ML pipeline for CAD diagnosis, which includes aspects like model optimization, engineering of clinical features, and post hoc interpretability through SHAP. The purpose is to develop a system that not only augments the prediction power but at the same time fosters confidence by explaining how every clinical factor has contributed to the final judgment. The sensitivity of the proposed model has exceeded that of some state-of-the-art works, like the benchmark proposed by Naji et al. (2024), and is shown by virtue of a threshold optimization step and a global feature importance analysis. [7]. Through this integrated approach, the enhancement of early CAD detection in typical real-world practices can be envisaged via a synergy of accurate predictions and clinical interpretability.

2. RELATED WORK
Currently, early detection and risk evaluation for coronary artery disease (CAD) have become very dependent on machine learning (ML), which is one of the major health concerns worldwide. The most recent research studies have used a combination of modern and traditional ML techniques to improve accuracy in diagnosis and provide support for the clinical decision, and explainability.

CAD holds great promise for diagnosis. Abdar et al. [6] have designed an outstanding hybrid machine learning model, which combines probabilistic reasoning and soft computing. But the clinical applicability was ruled out because the processes involved were not comprehensible. Naji et al. [7] developed an artificial neural network model with a good score on the Z-Alizadeh Sani dataset, using feature selection via LASSO.

Garavand et al. [8] implemented a variety of methodologies in undertaking a comparative study of ML models, including Random Forest, Decision Trees and ANN. In their results, they suggested that whereas interpretability was limited in ensemble methods, these methods showed better performance in terms of accuracy. Similarly, Bilal et al. [9] underlined the importance of feature selection and that ensemble learning methods (i.e., XGBoost) yielded better sensitivity and F1-score for CAD detection.

Performance evaluation of several ML classifiers, including logistic regression and deep learning, was done by Muhammad et al. [10] Their results indicated that deep learning models work well with larger datasets, while conventional modeling has a better track record with smaller clinical datasets, such as Z-Alizadeh Sani. In another study, Kataria and Kumar [11] proposed a gradient boosting approach supported by SHAP explainability to elucidate the clinical significance of important predictors.

A CAD prediction model has been proposed by Wang et al. [12] utilizing AutoGluon and AutoML framework. The model incorporated SHAP values that provided clinical transparency and a very high accuracy (91.67%) showing potential for diagnostic use.

Amini et al. [13] Radiomics data for myocardial perfusion from SPECT imaging were analyzed using machine learning techniques. This could describe how the interpretation of imaging biomarkers and SHAP-based explanation could be coupled to improve diagnosis with interpretable machine learning models categorizing CAD risk.

All these potentially necessitate the development of predictive interpretable models that can be used in hospitals for meaningful application in diagnosing CAD.

3. DATASET AND PREPROCESSING

The present study was based on the Z-Alizadeh Sani dataset, which is a clinically validated dataset popularly used in CAD prediction studies. The database contains 303 patient records and 54 attributes [7]. Therefore, 'Cath'(CAD: 1, Normal: 0) has been used as the target variable for supervised learning. Categorical variables such as hypertension and sex are encoded as binary values (e.g., Male = 1, Female = 0; Yes = 1, No = 0), whereas ordinal ones have symptoms of valvular heart disease encoded as an ordered list of integers according to the degree [10]. Model-derived clinically relevant properties were composed similarly to the LDL/HDL ratio, BMI, Age interaction, and BUN/Creatinine ratio based on diagnostic indicators recommended in the literature [2][12]. SimpleImputer was used for handling missing values brought about by median imputation. In addition, StandardScaler was utilized to normalize the feature range. The final preprocessed dataset comprised 57 features, balancing both raw clinical indicators and derived insights to enhance model robustness.

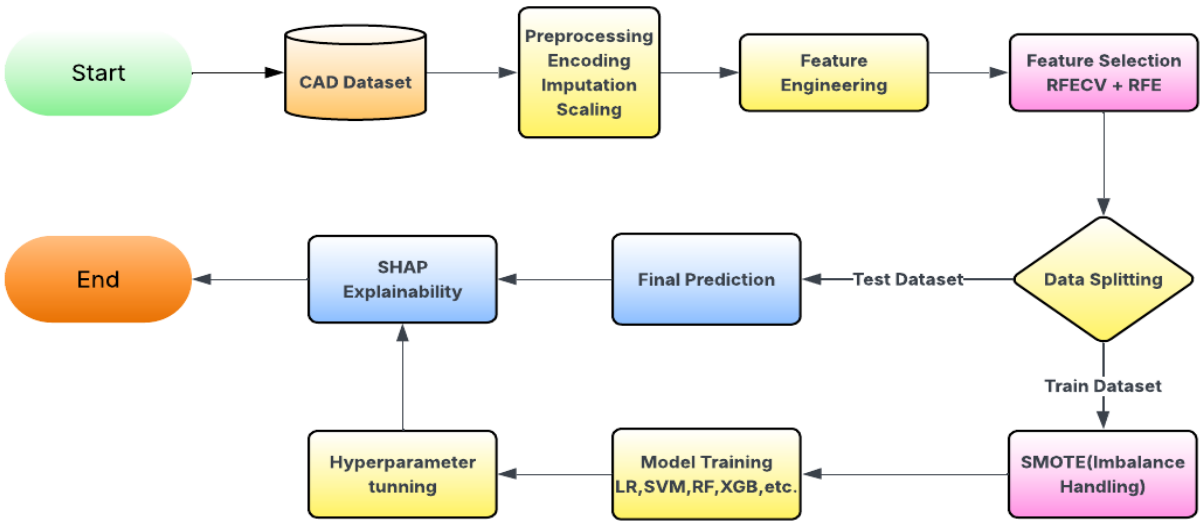

Figure 1. Overall Workflow for Coronary Artery Disease Detection using Optimized Machine Learning Pipeline.

4. FEATURE SELECTION

To identify the most informative subset of features for coronary artery disease (CAD) prediction, Recursive Feature Elimination with Cross-Validation (RFECV) was applied using a Logistic Regression estimator with class balancing and ROC AUC as the scoring metric. Stratified 5-fold cross-validation ensured robust performance estimation across imbalanced classes. A minimum of 15 features was enforced, and RFECV initially selected 28 features. To refine the selection further, RFE was used to prune the set down to exactly 23 features. This final set balanced statistical importance and clinical interpretability, including features such as Age, BMI, DM, EF-TTE, HTN, PR, Q Wave, Function Class, Typical Chest Pain, and engineered metrics like BMI × Age interaction. The selected features aligned well with established clinical indicators of CAD, enhancing the interpretability and medical relevance of the final model [2][6][7]. Figure 2 presents the correlation heatmap of the selected features, where minimal multicollinearity was observed, supporting their complementary nature in predictive modeling.

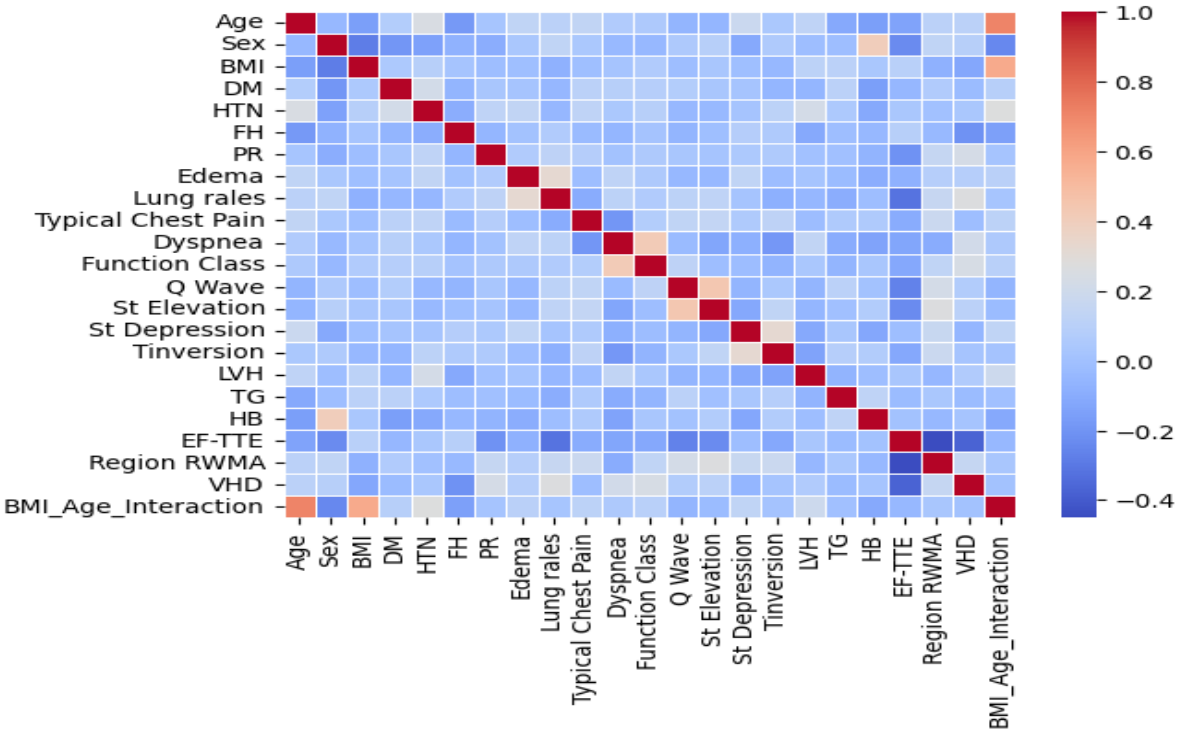

Figure 2: Correlation Heatmap of Selected Features

Correlation heatmap of the 23 final features selected after RFECV and RFE pruning. Blue tones indicate negative correlation, red tones indicate positive correlation, and white indicates low or no correlation. The matrix elaborates that most characteristics have very little correlation, thereby ensuring a diversity of predictive information.

5. MODEL DEVELOPMENT AND EVALUATION

In addition to the selection of proper algorithms, the medical context and statistical soundness must be carefully considered in the development of a clinically meaningful model for Coronary Artery Disease (CAD). This chapter describes the key components of the modeling process, from data validation to model comparison, along with justification for the rationale behind each step.

5.1 Cross-Validation Strategy

Clinical dataset issues are rife with overfitting owing to extreme class imbalance and very small sample sizes. Therefore, cross-validation is applied stratified by fold sizes, wherein PCG was divided into 5 equal folds with an equal share of CAD-positive and CAD-negative cases in every fold. Whereas in every round, the training is carried out on 4 folds and testing on one.

By using some stratification, every layer will represent the actual distribution of the data, which is a key part of making machine learning predictions to carry over to real-world scenarios in healthcare [14].

5.2 Dealing with Class Imbalance

The dataset has more instances of CAD compared to normal instances. A model trained on such an imbalanced dataset would learn to be biased towards the majority class. To counter this effect during training, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Unlike traditional oversampling, which simply duplicates minority class examples, SMOTE generates new samples by interpolating between nearby points in the feature space [15].

This approach enriches the minority class without introducing overfitting risks. Importantly, SMOTE was applied only to the training folds, leaving the validation data untouched there by preserving the integrity of the evaluation process [16].

5.3 Choice of Models: Principles and Medical Relevance

The diverse set of machine learning algorithms used to detect Coronary Artery Disease (CAD), balancing interpretability and predictive strength. balancing interpretability and predictive strength. This included linear models for transparency, tree-based models for decision logic, and neural networks for capturing complex nonlinear interactions.

5.3.1 Logistic Regression (LR)

Logistic Regression maps clinical variables to a probability score using the sigmoid activation:

$$p(y = 1|X) = \frac{1}{1+e^{-(\beta_0+\sum\beta_i x_i)}}$$

Its linear structure in the log-odds makes it easy to trace predictions back to specific clinical features, thus maintaining clinical interpretability [17].

5.3.2 Support Vector Machine (SVM)

SVM constructs a maximum-margin hyperplane to distinguish CAD and non-CAD cases. For nonlinear cases, it uses kernel functions like:

$$k(x_i, x_j) = \emptyset(x_i)^T \emptyset(x_j)$$

This allows SVM to operate in higher-dimensional spaces, making it robust for complex or small clinical datasets [18].

5.3.3 Random Forest (RF)

RF builds multiple decision trees using bootstrap samples and aggregates their predictions through majority voting:

$$\hat{y} = mode(T_1(x), T_2(x), \dots, T_n(x))$$

This ensemble method reduces overfitting and naturally highlights feature importance, a useful trait in clinical diagnostics [19].

5.3.4 Gradient Boosting Machine (GBM)

GBM improves accuracy by sequentially correcting errors of previous models. Each iteration minimizes the loss:

$$F_m(x) = F_{m-1}(x) + \eta \, . \, h_m(x)$$

Its additive nature enables the model to learn subtle interactions between clinical factors [20].

5.3.5 XGBoost

XGBoost extends GBM with regularization:

$$\text{Obj} = \sum l(\hat{y}_i, y_i) + \sum \Omega(f_k), \Omega(f_k) = \gamma^T + \frac{1}{2}\lambda ||w||^2$$

This prevents overfitting and offers fast, accurate predictions in structured medical datasets [21].

5.3.6 LightGBM

LightGBM grows trees leaf-wise instead of level-wise and uses histogram-based binning, accelerating training while preserving accuracy. It is ideal for datasets with high cardinality features [22].

5.3.7 Multi-Layer Perceptron (MLP)

MLP is a fully connected neural network where each neuron performs:

$$z = \sigma\left(\sum w_i x_i + b\right)$$

MLPs approximate nonlinear patterns in data, useful for capturing subtle and non-obvious CAD risk factors [23].

5.4 Hyperparameter Optimization

Each algorithm's performance depends on its internal configuration. RandomizedSearch CV [24] is used , which samples a subset of possible combinations to efficiently tune parameters like tree depth, learning rate, or regularization strength. This search was embedded within 5-fold cross-validation**,** ensuring that parameter selection generalized across patient subgroups. Sensitivity was prioritized throughout the tuning process to reduce the risk of missing true CAD cases.

5.5 Evaluation Metrics and Clinical Interpretation

In CAD diagnosis, it is critical not only to achieve high predictive accuracy but also to minimize life-threatening errors like false negatives. Therefore, a comprehensive set of clinical and statistical metrics was employed, each supported by established literature [25][27].

5.5.1 Sensitivity (Recall)

Measures the model's ability to detect actual CAD patients. A higher value reduces the risk of undiagnosed cases*[25].*

$$Sensitivity = \frac{TP}{TP + FN}$$

5.5.2 Specificity

Assesses the ability to correctly identify non-CAD cases, minimizing unnecessary anxiety or invasive tests[25].

$$Specificity = \frac{TN}{TN + FP}$$

5.5.3 Precision

Indicates the proportion of correctly identified CAD cases among all predicted positives, important for reducing overtreatment[25].

$$Precision = \frac{TP}{TP + FP}$$

5.5.4 F1-Score

The harmonic mean of precision and sensitivity; useful when dealing with imbalanced datasets [26].

$$F1 - Score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

5.5.5 Accuracy

Captures the overall prediction correctness but may be misleading if classes are imbalanced[25].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

5.5.6 ROC AUC

Reflects the model's discrimination ability over all thresholds. AUC values closer to 1 indicate excellent separability between CAD and non-CAD cases [25].

5.5.7 Negative Predictive Value (NPV)

Measures how often a negative prediction truly corresponds to the absence of disease, which is crucial when ruling out CAD [27].

6. MODEL RESULTS AND COMPARISON

After training and tuning all models under a unified evaluation framework, their performance was assessed using cross-validated predictions on the entire dataset. Table 1 presents a comprehensive summary of seven candidate models based on a variety of diagnostic metrics.

Each model was evaluated using the same 5-fold cross-validation strategy and SMOTE-adjusted training folds to ensure a fair and unbiased comparison. The metrics include Accuracy, Sensitivity (Recall for CAD class), Specificity, Precision, F1-Score, Area Under the ROC Curve (ROC AUC), Negative Predictive Value (NPV), and raw confusion matrix values (TP, TN, FP, FN). Training time was also recorded for completeness.

Table 1. Simplified Model Performance Summary (Cross-Validation Pipeline Results)

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	ROC AUC	NPV	TP	TN	FP	FN	Time (s)
Logistic Regression	0.8912	0.8989	0.8721	0.9512	0.9243	0.9575	0.7676	192	75	12	24	1.59
ANN (MLP)	0.8912	0.8850	0.9066	0.9645	0.9230	0.9571	0.7529	189	78	9	27	23.06
SVM	0.9044	0.9267	0.8491	0.9440	0.9352	0.9548	0.8122	198	73	14	18	20.03
XGBoost	**0.9209**	**0.9637**	**0.8146**	**0.9338**	**0.9485**	**0.9503**	**0.8850**	**206**	**70**	**17**	**10**	**23.10**
Gradient Boosting	0.8945	0.9359	0.7916	0.9232	0.9295	0.9444	0.8195	200	68	19	16	104.34
Random Forest	0.8747	0.9452	0.6997	0.8921	0.9179	0.9418	0.8208	202	60	27	14	94.65
LightGBM	0.9077	0.9452	0.8146	0.9324	0.9387	0.9267	0.8433	202	70	17	14	15.22

Among all models, XGBoost emerged as the top performer with a sensitivity of 0.9637, the highest across all models. This metric is particularly critical in CAD diagnosis, where false negatives can lead to missed interventions. Furthermore, its F1-Score (0.9485) and ROC AUC (0.9503) indicate a strong balance between recall and precision. Importantly, XGBoost also maintained high NPV (0.8850), suggesting confidence in ruling out CAD when predictions are negative.

Although models like Random Forest and LightGBM also achieved high sensitivity, their trade-off in specificity and overall balance made them less suitable as the final choice. Logistic Regression and ANN maintained strong precision, but fell short in terms of sensitivity compared to XGBoost.

6.1 ROC Curve Analysis for All Models

The Receiver Operating Characteristic (ROC) curve provides a threshold-independent assessment of classifier performance. Figure 3 displays the ROC curves of all tuned models, with XGBoost achieving a strong area under the curve (AUC) of 0.9503. Although Logistic Regression (0.9575) and MLP (0.9571) showed slightly higher AUC values, their performance in terms of recall and F1-score was lower than that of XGBoost.

In medical diagnostics, the ROC AUC is especially valuable as it measures a model's ability to distinguish between classes at various threshold levels, offering a global evaluation of sensitivity-specificity trade-offs [25].

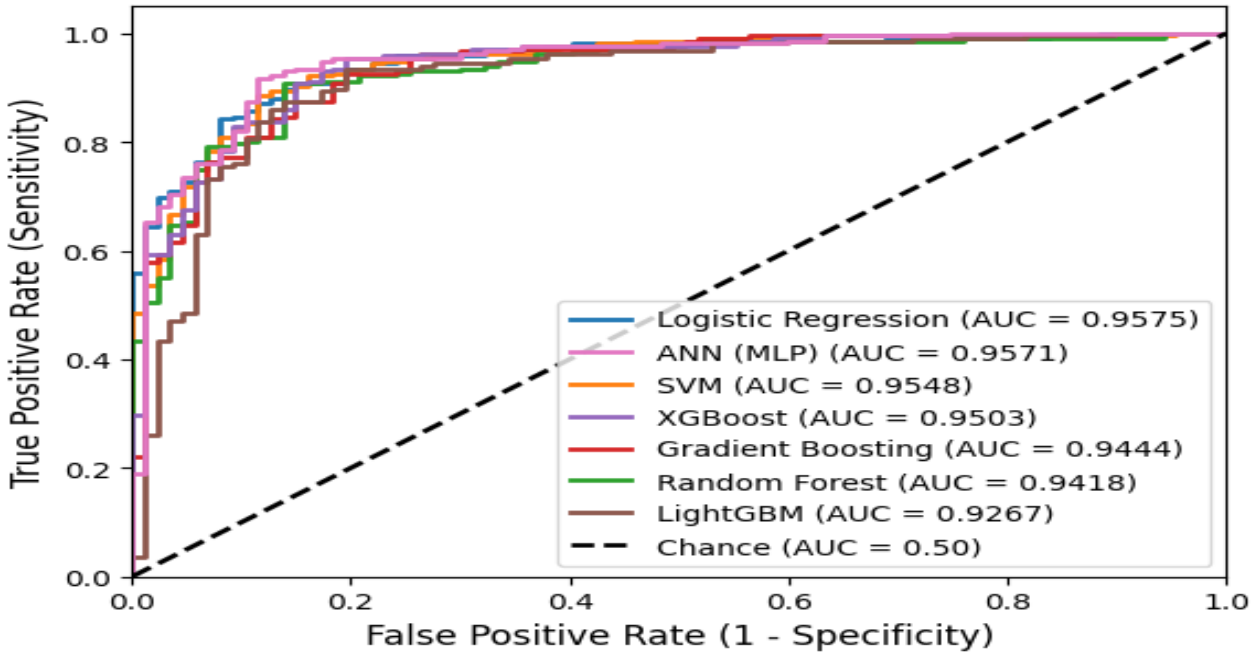

Figure 3. ROC Curves for All Tuned Models

6.2 Benchmark Comparison with Naji et al. (2024)

The performance of the best-performing model (XGBoost) was compared against the benchmark metrics reported by Naji et al. (2024) [7], who used an artificial neural network trained on LASSO-selected features. Table 2 summarizes this head-to-head comparison.

Table 2. Comparison of the Best Model (XGBoost) with Naji et al. (2024)

Metric	The Best Model (XGBoost)	Naji et al. (ANN)	Comparison
Accuracy	0.9209	0.9038	Better (+0.0171)
Sensitivity	0.9637	0.9443	Better (+0.0194)
Specificity	0.8146	0.8027	Better (+0.0119)
Precision	0.9338	0.9251	Better (+0.0087)
F1-Score	0.9485	0.9336	Better (+0.0149)
ROC AUC	0.9503	0.9246	Better (+0.0257)
NPV	0.8850	0.8595	Better (+0.0255)

6.3 Confusion Matrix Analysis for Final Model

To visualize the XGBoost model's diagnostic behavior, its confusion matrix is presented in Figure 4. This visual summarizes how well the model distinguishes between CAD-positive and CAD-negative patients using a fixed decision threshold (optimized for F1-score).

The model correctly identified 206 out of 216 CAD-positive patients and 70 out of 87 CAD-negative patients. With only 10 false negatives and 17 false positives, the confusion matrix supports the high sensitivity and balanced precision–recall tradeoff that XGBoost achieved during training.

Confusion matrices offer a threshold-specific view that is particularly useful for clinical decision-making, where false negatives can have serious implications [26].

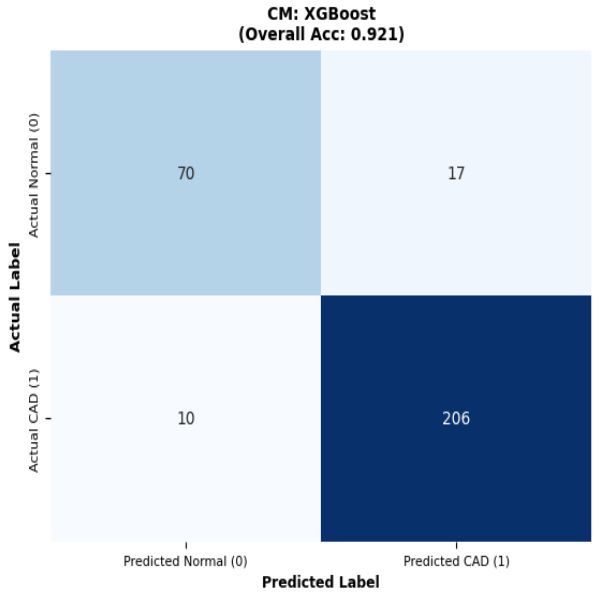

Figure 4. Confusion Matrix of Final XGBoost Model

7. SHAP-BASED EXPLAINABILITY

7.1 Global feature importance in CAD prediction

Interpretability is a critical factor in the adoption of machine learning models in clinical environments, especially for life-threatening conditions like Coronary Artery Disease (CAD). To provide transparency and support clinical trust, the study employed SHAP (SHapley Additive exPlanations), a model-agnostic interpretability method grounded in cooperative game theory [27]. SHAP values reveal how much each feature contributes to a prediction, both at the global and individual level.

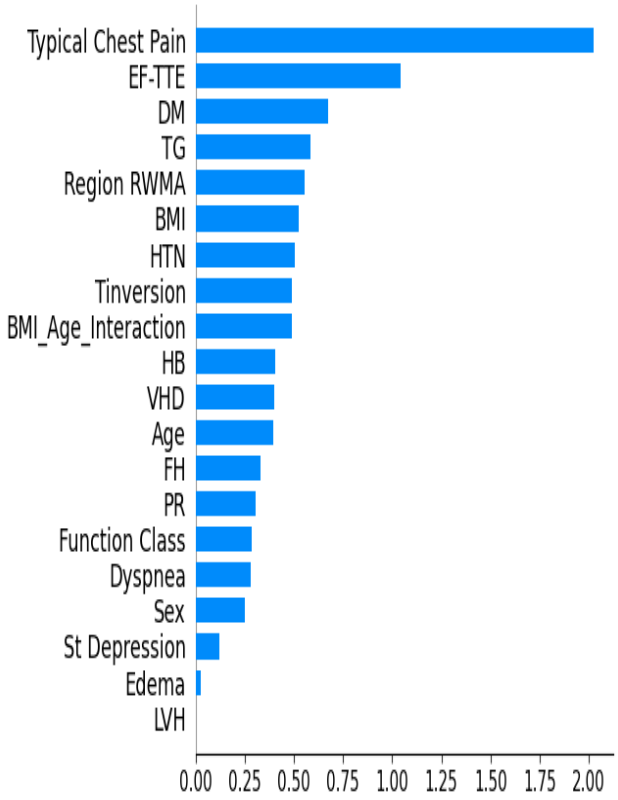

Figure 5. SHAP summary bar plot showing average importance of top features.

As seen in Figure 5, the most influential feature was Typical Chest Pain. Its dominance in the model is aligned with clinical reality, where chest pain remains the most common presenting symptom of CAD. Following this, EF-TTE (Ejection Fraction from echocardiography) was the second most important feature, underscoring the relevance of left ventricular function in determining cardiac risk. Other high-ranking features included Diabetes Mellitus (DM), Triglyceride (TG) levels, and Regional Wall Motion Abnormality (RWMA) all medically validated cardiovascular risk indicators [11].

7.2 Direction and strength of individual predictions

While the bar plot provides an overview of feature importance, it does not reveal how each feature influences predictions for individual patients. This is illustrated by the SHAP beeswarm plot, where each point represents a patient. The x-axis shows whether the feature increases or decreases CAD prediction probability, and color denotes the actual feature value (red = high, blue = low).

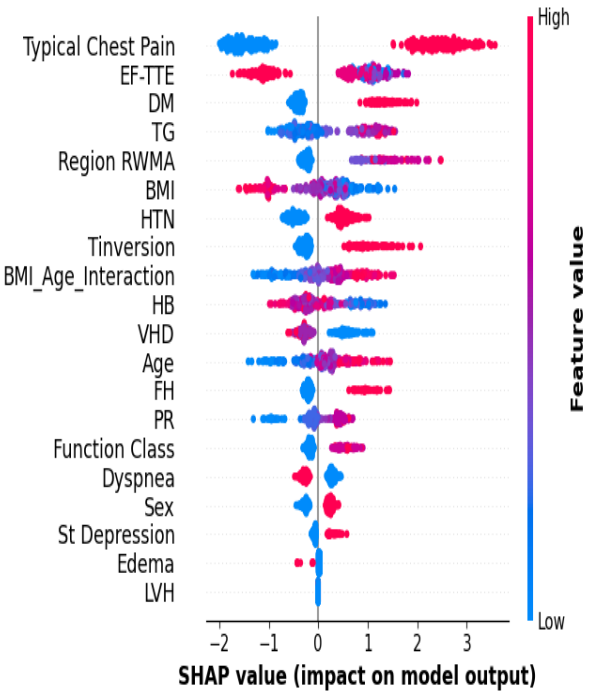

Figure 6. SHAP beeswarm plot showing direction and density of feature contributions.

In Figure 6, patients with high Typical Chest Pain (red) values overwhelmingly push the model toward predicting CAD. Similarly, low values of EF-TTE (blue), indicating reduced cardiac function, also increase the likelihood of a positive prediction. These patterns reinforce the clinical validity of the model's logic and enhance trust in its decision-making process.

7.3 Typical Chest Pain: A dominant clinical indicator

To further explore the role of top features, SHAP dependence plots were examined. Figure 7 focuses on Typical Chest Pain, a binary symptom recorded as 1 for presence and 0 for absence. Patients with a value of 1 showed significantly higher SHAP scores, confirming the model's strong reliance on this symptom for CAD prediction. The clear distinction in SHAP values across the binary states highlights chest pain as the most decisive input factor.

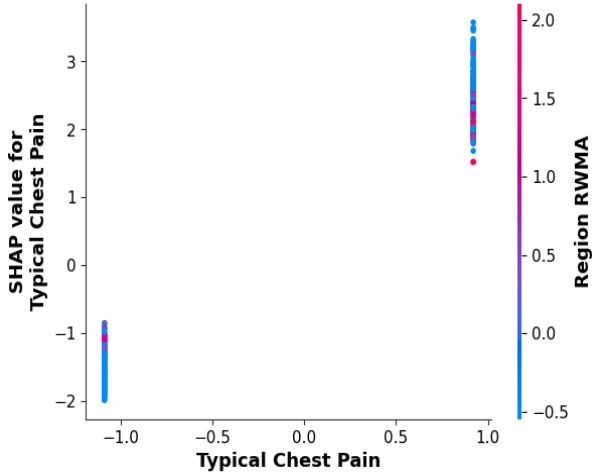

Figure 7. SHAP dependence plot showing the effect of Typical Chest Pain on CAD prediction.

7.4 EF-TTE: Interpreting cardiac efficiency in predictions

Figure 8 illustrates the relationship between EF-TTE values and CAD risk. As EF-TTE decreases, the SHAP value sharply rises suggesting that patients with reduced ejection fraction are at a significantly higher risk of CAD as per the model. This relationship reflects real-world clinical practices where left ventricular dysfunction is a red flag in cardiac diagnostics. The smooth nonlinear pattern also The smooth, nonlinear pattern of the model shows its capacity to broadly apply risk beyond strict requirements.

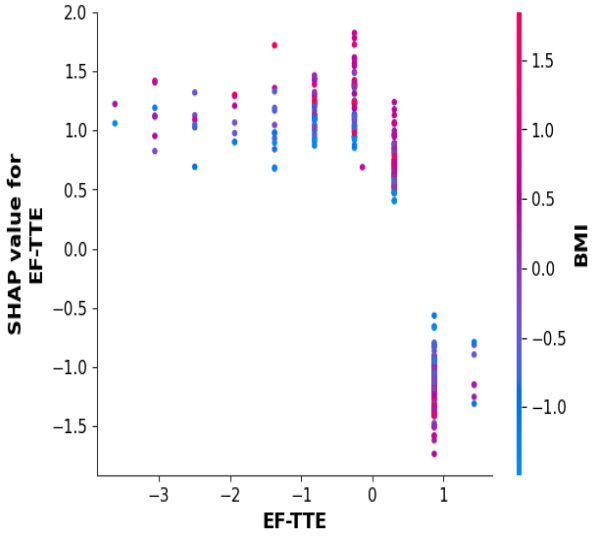

Figure 8 Shows the SHAP dependency plot and the effect of EFTTE on CAD prediction.

8. CONCLUSION AND FUTURE WORK

This research presents a machine learning pipeline integrating clinical knowledge with algorithmic accuracy to facilitate earlier identification of Coronary Artery Disease (CAD).

Aggressive preprocessing was undertaken, and relevant clinical features created, whereby the key variables were selected using RFECV and RFE on the Z-Alizadeh Sani data. Seven machine-learning models were trained using a 5-fold cross-validation framework employing SMOTE resampling to counteract class imbalance. The performance of those models was evaluated upon several metrics, including sensitivity, specificity, ROC AUC, F1-score et cetera.

XGBoost scored a sensitivity of 0.9637 and an ROC AUC of 0.9503, confidently outperforming the best similar contenders, and set a bar that was raised even further above that established by Naji et al. (2024). There must be very high sensitivity in medical diagnosis because false negatives can be detrimental to a patient's prognosis. The high NPV of the model will contribute to its reliability in a clinical setting to diagnose disease.

Finally, an SHAP analysis was performed for internal relationship exploration of final optimized model. The SHAP visualization shows that the most impactful parameters in the model-predicted predictions are diabetes mellitus, triglyceride levels, EF-TTE, and normal chest discomfort. These findings correlate with clinical studies that studied the relationship between machine learning outputs and functional cardiology practice.

Findings indicate that, when correctly applied for example, through optimal feature selection, balancing classes, and provision for explainability machine learning stands to provide not just excellent predictive accuracy, but also clarity of understanding and clinical relevance, in the diagnosis of coronary artery disease (CAD). It would thus create a wonderful intersection between data science and healthcare, where decision support systems provide credible yet comprehensible systems.

Future studies should develop new approaches for the implementation of attention-based deep learning models to incorporate sequential EGG or imaging data. Plus, it should cross-link to the population-based electronic health records and test the model in prospective clinical databases. Lastly, integrating the models' outcomes with current clinical decision support systems might add significant boosts to their impact in the hospital environments.

9. REFERENCES

[1] World Health Organization. 2021. Cardiovascular diseases (CVDs). Available at: https://www.who.int

[2] Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., et al. 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries. The Lancet, 364(9438), 937–952.

[3] Tavakol, M., Ashraf, S., and Brener, S. J. 2012. Coronary angiography: Background, complications and limitations. Heart, 98(12), 915–921.

[4] Heo, R., Nakazato, R., Kalra, D., and Min, J. K. 2014. Noninvasive imaging in coronary artery disease. Seminars in Nuclear Medicine, 44(5), 398–409.

[5] Rajpurkar, P., Irvin, J., Chen, E., Zhu, K., and Ng, A. 2023. AI in healthcare: The path forward. NEJM AI, 1(1), 3–13.

[6] Abdar, M., Zomorodi-Moghadam, M., Das, R., and Ting, I. H. 2019. A new machine learning-based classification for diagnosis of coronary artery disease. Computer Methods and Programs in Biomedicine, 179, 104992.

[7] Naji, M. S., Hashim, S. Z. M., and Salih, M. M. 2024. Artificial neural network model for predicting coronary artery disease using LASSO-selected features. International Journal of Computing and Digital Systems, 13(1), 33–41.

[8] Garavand, A., Behmanesh, A., and Samadbeik, M. 2022. Efficient model for coronary artery disease diagnosis: A comparative study. Journal of Healthcare Engineering, Article ID 5597723.

[9] Bilal, M., Khan, M. J., Raza, B., and Alzahrani, M. 2023. Comparison of machine learning algorithms for coronary artery disease detection using feature selection. Indonesian Journal of Electrical Engineering and Computer Science, 31(1), 187–193.

[10] Muhammad, L. J., Algehyne, E. A., Usman, S. S., and Dada, E. G. 2021. Supervised machine learning models for coronary artery disease prediction: Performance analysis and comparison. Applied Sciences, 11(6), 2874.

[11] Kataria, V., and Kumar, N. 2025. Predicting coronary artery disease using explainable machine learning. International Journal of Scientific Research in Engineering and Management, 9(1), 1–6.

[12] Wang, J., Xue, Q., Zhang, C. W. J., and Tang, X. 2024. Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework. Frontiers in Cardiovascular Medicine, 11, 1360548.

[13] Amini, M., Pursamimi, M., and Hajianfar, G. 2023. Machine learning-based diagnosis and risk classification of coronary artery disease using myocardial perfusion imaging SPECT: A radiomics study. Scientific Reports, 13, 14920.

[14] Refaeilzadeh, P., Tang, L., and Liu, H. 2009. Cross-validation. In Encyclopedia of Database Systems. Springer.

[15] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.

[16] Douzas, G., and Bacao, F. 2018. Effective data generation for imbalanced learning. Expert Systems with Applications, 91, 464–471.

[17] Hosmer, D. W., and Lemeshow, S. 2000. Applied Logistic Regression. 2nd ed., Wiley.

[18] Cortes, C., and Vapnik, V. 1995. Support-vector networks. Machine Learning, 20, 273–297.

[19] Breiman, L. 2001. Random forests. Machine Learning, 45(1), 5–32.

[20] Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

[21] Chen, T., and Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD, 785–794.

[22] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. 2017. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

[23] Hornik, K., Stinchcombe, M., and White, H. 1989. Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366.

[24] Bergstra, J., and Bengio, Y. 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281–305.

[25] Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.

[26] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. 2019. A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

[27] Lundberg, S. M., and Lee, S.-I. 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.