Ensemble Machine Learning Approach for Coronary Artery Disease Prediction and Risk Factor Analysis

Submitted By

Mist. Fabia Akter Barsha
(221-35-1052)
Department of Software Engineering

Supervised By

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering

Thesis submitted to fulfill the qualifications for the Bachelor of Science in Software Engineering degree

Fall 2025

APPROVAL

This thesis titled on "**Ensemble Machine Learning Approach for Coronary Artery Disease and Significant Risk Factor Analysis** ", submitted by **Mist.Fabia Akter Barsha (ID: 221-35-1052)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Elahi

-- **Chairman**

Dr. Fazla Ealhe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Marzia

-- **Internal Examiner 1**

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 3**

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Mohammod Abul Kashem, PhD
Professor
Department of Computer Science and Engineering
DUET, Bangladesh

DECLARATION

I hereby declare that the work presented in this thesis is my own original research work, carried out under the supervision of **Dr. Md. Fazla Elahe**, Assistant Professor & Associate Head, Department of Software Engineering, Daffodil International University.

This work has not been submitted anywhere, either in whole or in part, for any degree, diploma, or publication in this or any other university. All sources of information used in this thesis have been duly acknowledged.

Supervised By

Fazla Elahe

....

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering

Submitted By

Fabia

....

Mist. Fabia Akter Barsha
ID: 221-35-1052
Department of Software Engineering

AKNOWLEDGEMENT

First and foremost, I express my deepest gratitude to Almighty Allah for granting me the strength, patience, and determination to complete this research work successfully. Without His blessings, this accomplishment would not have been possible. I am sincerely thankful to my esteemed parents for their unwavering support, encouragement, and guidance throughout my academic journey. Their sacrifices and belief in my abilities have been a constant source of inspiration. I would like to extend my profound gratitude to **Prof. Dr. Imran Mahmud**, the respected Head of the Department of Software Engineering, for providing a supportive academic environment and the necessary facilities that greatly contributed to the completion of this work. My deepest appreciation goes to my supervisor, **Dr. Md. Fazla Elahe**, for his valuable guidance, continuous support, and insightful feedback throughout the entire research process. His expertise and encouragement have been pivotal in shaping this thesis. I am also grateful to all faculty members of the Department of Software Engineering for their dedication to academic excellence and for imparting the knowledge and skills that enabled me to accomplish this

research successfully.

Finally, I would like to acknowledge the support and cooperation of my friends, classmates, and colleagues at Daffodil International University (DIU) for their encouragement, constructive discussions, and assistance during the course of this work.

Abstract

The development of precised, effective, and non-invasive diagnostic process is necessary for early intervention since coronary artery disease (CAD) continues to be a major worldwide health concernd. Traditional machine learning models have shown promised in predicting healthcare, but because they are good and difficult to explain, their "black box" character presents a major obstacled to practical application. By combining explaining artificial intelligence (XAI) method with a high-performancess ensemble learning methodology, this study offers a novel framework that tackles this pressing issue. The proposed solution incorporates extremely cautiously the predictived natured of LightGBM and XGBoost by meaning of a voting ensemble design as their hyperparameters are optimized with the aid of Bayesian models. The aim behind the hybrid architecture is to deliver highs accuracy of forecast as well as trends of complexity data.

The effectivenes of the models was evaluated comprehensively using a variety of measures, including accuracy, precision, recall and the under the receiver operating characteristic curve (AUC). Based on the experimentals findings, the based classifiers approach is significantly better than the voting ensemble model with an 96.218 test accuracy. The analysis funds on SHapley Additive exPlanations (SHAP) that involves the decomposition of the opaque decision-making process in the model to obtain comprehensible and practical information. The three features that are most frequently highlighted in the forecasting of CAD by the XAI analysis are cholesterol and resting blood pressure in addition to the nature of the chest palpation. The SHAP analysis provides a deeper view on their positive or negative relationship with disease risk. This thesis demonstrated the important of improved predictive performancd between highly developed ensemble modeling and XAI as well as contributed to the interpretability and trust required to facilitated the seamless integration of AI-oriented systems into a clinical decision-making-driven environment.

Table of Contents

List of Tables

Chapter 1

Introduction

1.1. The Global Problem and Clinical Significance of Coronary Artery Disease

Coronary Artery Disease (CAD), a cardiovascular disease (CVD) that is a major problem on its own in any given country across the global is a very serious health problem. The number of deaths causes by CVDs is estimated at 17.9 million each year, which mean that are the leading causes of death on the global staged [1]. The tremendous influenced CAD has madeed on the populations health and system healthcare can serve as a moving factor towards current endeavors to enhanced an EWS of detection and diagnosis. Although not invariably fatal, CAD required timely and corrects diagnoses in order to implemented their time treatment and ensure patient live change carefully [2].

Historically, the CAD diagnosis processed has been based on clinical evaluations, which covers patients history, the evaluation of risk factors, and the diagnostic process itself, involving invasived techniques into the heart, namely coronary angiography. Unfamiliar approached are usually costly, extend their stayed in the hospital and use large resources and are associated with inherent risks to the patient such as possibility of being exposes to radiation, or political complications [3, 2]. The drawbacks of these traditional methods of diagnostic reasoning, namely the gravity of diagnostics certainty and the invasived nature of the procedure, have stimulated the active development of non-invasive, predictive diagnostics variations. Over the past few years, data-oriented methods, especially based on machine learning (ML), have become a promising direction in the analysis of the large, multi-factor health data. These models are both successful in determining unobtrusive non-linear presentations of risk to early disease, and there is the possibility of identifying a non-invasive, low-cost, early-diagnostics solution, and directions can be taken towards more active and constructive clinical interception [4].

1.2. The Emergence of Machine Learning and the Need for Interpretability

Machine learning has presented formidable computational paradigms to study complex, high-dimensional datasets, making it possible to design predictive models that are robust with no direct, rule-based programming [4, 5]. Hybrid Ensemble Learning (HEL) along with other techniques have been found specifically useful in the high-precision determination of heart disease [2]. Nevertheless, one of the key barriers to the extensive clinical application of most focused ML models, in particular, boosting-based ensembles and deep neural networks, is their natural obscurity, also known as the black box problem [6, 7]. The inability of medical professionals to ascertain the target path of causality eventually resulting in a prediction circumscribes clinical accountability, errors diagnosis capabilities, and essentially, the trust which MM to incorporate to high-stakes clinical processes is rendered obsolete [6, 7].

The morally, as well as, the logistically, momentous consequences of depending on non-transparent models, are vast. Slowly acquired hidden algorithmic bias, state that may be inherited by unrepresentative training data, can accomplish unfair outcomes regarding various patient demographics [6, 7, 8]. Also, compliance with rules, especially privacy legislation of patient information such as the Health Information Privacy and Accountability Act, and the existence of auditable, transparent algorithm implementation, require a movement towards clarity [8]. To be a valid and extremely useful tool to assist doctors, a more sophisticated AI system needs to be highly diagnostic and, at the same time, it must be able to explain its results in a transparent and understandable way by a human [6, 7]. Explainable Artificial Intelligence (XAI) explicitly supports this requirement, and provides both theoretically and computationally feasible systems to reconcile the performance-interpretability gap [9, 10, 2].

1.3. Problem Statement

In this study, there is need to look at the fundament principle need to construct a predictive model of Coronary Artery Disease that meets simultaneously two basic, but usually conflicting, desiderates: the use of a maximum possible diagnostic effectiveness and the availability of coherent, solidly reliable developing explanations of its results. In particular, the research questions is to look the conventional problems of non-interpretable black boxs and combine a computationally advantageous ensemble learning model with a powerful AI algorithm, SHAP.

Effectively navigating through this predicament is a precondition for deploying AI systems in the most high-stakes segments of healthcare, where justification of the quality of a decision can have a direct impact on the outcome for patients as well as on medical and overall liability [10, 11].

1.4. Research Objectives

The main aims of this thesis are as follows:

- To design and implement a high-performance hybrid ensemble learning model using classifiers XGBoost and LightGBM. This objective is based on the hypothesis that, if we can combine the complementary capabilities of these two state-of-the-art boosting algorithms, namely, XGBoost's strong L1/L2 regularisation capabilities for alleviating over-fitting, with the superior computational efficiency and parallel processing capabilities of LightGBM, the resulting ensemble will have a demonstrably higher level of predictive performance and stability than any of the constituent models could ever achieve alone [12].
- To use Bayesian optimisation for the rigorous and efficient fine- tuning of all critical model hyperparameters and thus achieving maximal predictive accuracy. This stochastic, inferential search technique has been deliberately selected over exhaustive methods like grid search, because in this technique the hyperparameter space is intelligently explored by a surrogate model, thereby minimising the number of costly model evaluations to be performed in order to find the optimal configuration.
- To carry out a thorough assessment of the proposed model based on a collection of standard performance metrics, such as Accuracy, Precision, Recall, and Area Under the Receiver Operating Characteristic Curve (AUC). This multi- metric assessment is fundamental in a holistic picture of the clinical utility of the model with a special focus on maximising Recall (sensitivity, and therefore minimising false negatives) whilst achieving high Precision (positive predictive value, and therefore minimising false positives) [4]. The AUC is a good, threshold-independent measure of the ability of the model to be discriminatory [4].

- To use a theoretically based XAI technique SHAP to provide detailed, post hoc explanations for the ensemble model's predictions and, in turn, rigorous identification and quantification of the most important clinical and demographic risk factors involved in CAD prediction. This objective directly addresses the "black - box" challenge by giving a quantitative understanding of feature influence on the diagnostic outcome [13, 14].

1.5. Contributions of the Thesis

The contributions of this research are multifaceted and significant and progress the machine learning and clinical application aspects [2]:

- **Novel Hybrid Ensemble with High Performance:** A hybrid ensemble learning framework is provided that achieves high predictive power for the prediction of heart disease when compared to each of its component models (XGBoost and LightGBM). This system helps to effectively overcome the bias - variance trade-off of the single models [2].
- **SHAP-Driven Interpretability.** The study is able to integrate and validate the SHAP framework for explaining the complex decision process of a gradient boosted ensemble. This is giving the model's predictions a clear, quantitative, and elaborate rationale, and directly addresses the critical barrier of model opacity in clinical AI adoption [2, 15].
- **Quantification of Risk Factors.** The methodical application of SHAP analysis, the research quantifies the positive and good relationships between relevant clinical features (e.g., cholesterol, resting blood pressure, maximum heart rate) and CAD risk, and can help medical professionals better understand CAD risk early in diagnosis and resource allocation depending on quantifiable feature significance [2].
- Feature for Transparent Clinical AI. The work establishes an auditable methodology showing how better predictive performance and good manner transparency in high stake medical diagnostics can be achieved at the same time thus providing a compelling justification for incorporating XAI as a standard part of future healthcare models [11].

1.6. Thesis Structure

From fundamental theory to real-world applications and scientificed assessments, the rest of this thesis is organized to make senses. A thorough out analysis of pertinents and academic literature is given in Chapter 2, which also covered fast machine learning applications for predicting of cardiovascular illness and repeatedly explains the underlying theory of the selected XAI mechanism. **Chapter 3** continues details the research methodology, including the dataset setting, domain-specificed data preprocessing pipeline, the main underpinnings and math ethical formalism of the proposed ensemble model and Bayesian optimizations strategy.

Chapter 4 presents the quantitative experimental results, offering a good analysis of the performance metrics across all models, followed by a critical, clinically presented discussions of the risk factors express by the SHAP analysis. **Chapter 5** concludes the thesis by summarizing the major contributions, discussing the limitations of the currented study, and proposing concreted path for future research in the development of good and effectived clinical decision support systems [2].

Chapter 2

Literature Review

2.1. Machine Learning for Cardiovascular Disease Prediction

Machine learning in the medical field are an that has been growing at a rapid pace, and cardiovascular disease prediction has been a subject of numerous studyies. Initial studies gets to use classic classifiers, including Logistic Regression, K-Nearest Neighbors (KNN) and Decision tree ``. Although offering a rough conceptual platform in these single-model methods tended to fail to give the accuracy needed to be clinically reliable on complex, noisy data. An example is a study by Akter et al. [16], where the most accurate model when hyperparameter tuned was the logistic Regression at 93.44, a potentially good result; nonetheless, this could be higher using non-radical algorithms.

The field has since advanced to adopt more advanced methods, especially ensemble learning, where the prediction of a single model is used to obtain a stronger and more accurate final result by combining the predictions of the multitude of models. XGBoost and LightGBM are boosting algorithms that have become the most suitable solutions because of their outstanding level of performance. In a study, an XGBoost model was observed to get an accuracy at 89 percent in a heart disease prediction a task [5]. Moreover, scholars have examined ensemble techniques, among them bagging (Random Forest), boosting (AdaBoost) and stacking [3, 4, 17]. Another example is that a stacking-based model was estimated to have a high level of accuracy 97.06 that was better than the separate classifiers [17]. On the same note, a hybrid type which incorporated both forest random and adaBoost achieved a test accuracy of 95 percent on a very big dataset [3]. The literature trend here is clearly an indicator that as the ML techniques have become more mature, movement of single-model to multi-model and hybrid techniques has brought forth impressive increases in performance.

2.2. The Critical Role of Explainable AI in Clinical Practice

Improving predictive accuracy of complex machine learning models has been in synergy with the increasing worry that the predictive accuracy is uninterpretable. A good number of good models performed are black boxes, which can be difficult for medical professionals to compare how they make decisions [9, 6].

This security is a critical to the pervasive implementations of AI in medical care because any research is clinical judgments should be clear, responsible, and built on trust [6, 7]. Lack of access to the clear rationale of a diagnosis through a model may induced the either under- or over-relianced on prediction which may jeopardize the safe of the patient [18]. The ethical problems of team models that are black boxes are numerous. They are capable of transferring and supporting biases that are problem in training data, which may result in unequal results of treating different patient groups. Moreover, significant amounts of sensitive patient data gathered and utilized pose serious privacy-related problem, with these regular as HIPAA and GDPR regulating this issue requiring to some extent transparency on the use of data and algorithms [8, 7]. Explainable AI (XAI) directly resolves these problems introducing a family of approached that aim to interpret AI decisions in a manner that is understand to humans [9, 10]. It is aimed not only at developed a greater consistency within the ranks of clinicians but also at going them with the information needed to sell the reasoning of a model to them in compare with their own clinical knowledge [7]. The present paper has placed itself to fit this crucial region by showing how XAI can be comfortably good into a predictive high-performance set-up.

2.3. SHapley Additive exPlanations (SHAP) for Model Interpretability

This study explores SHapley Additive exPlanations (SHAP) as an effective theoretically based framework used to explain the output of machine learning predictors in order to offer transparency and self-service on this possibility [14, 15]. SHAP has its foundation in the cooperative game theory, in which the complex prediction of the model can be paralleled to a payout that has to be equitably shared among the combining input characteristics, which are viewed the "players" in a predictive game.

2.3.1. Theoretical Foundations of Shapley Values

The mathematical and fundamental idea of SHAP is the computation of the Shapley values (ϕj), denoting SO 2 features to the final prediction via all potential feature coalitions. The Shapley value is given as the minimal solution in which four fairness properties are met, namely: Efficiency (accumulating the contributions of all features results in the accumulating prediction), Symmetry (two features that have the same added value get equal share), Dummy (features that contribute no value get no share), and

Additivity (the shapley values of a combined prediction function are equal to the sum of the individual shapley values).

The mathematical formulation for the SHAP value (ϕj) for a feature j is given by the weighted average of the marginal contributions across all possible feature subsets:

$$\phi_j(v) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \qquad (1)$$

where F is the complete set of features, S is a subset of features not including j, v(S) represents the model's predicted value using only features in set S, and the fraction $\frac{|S|!(|F|-|S|-1)!}{|F|!}$ serves as a weight that accounts for the number of feature subsets (coalitions) of a specific size |S| [19, 14].

2.3.2. Practical Implementation with TreeSHAP

While the principal theorem that relates to the Shapley value appears to hold water, the computation requirement is a steep mountain that cannot be scaled; it grows exponentially and requires 2^n model evaluations for n explanatory variables, a computationally impossible task when faced with large datasets [15]. To address this, the current inquiry employs TreeSHAP which is a highly efficient algorithm that is specially designed for decision tree ensemble algorithms, such as XGBoost and LightGBM. TreeSHAP significantly reduces the computational burden from exponential to polynomial time by exploiting the hierarchical structure of decision trees ``. Instead of enumerating all 2n combinations of features, TreeSHAP computes the exact expected prediction within each node by traversing the tree based on the conditional expected prediction, effectively handling feature dependencies without explicit sampling [20]. This robust and efficient property makes TreeSHAP the optimal choice for interpreting the complex, non-linear decision boundaries established by the gradient-boosted ensemble framework utilized in this research [20, 21].

Chapter 3

Methodology

This research was performed using a methodology of six step systematic approach that will focus on building and testing a transparent and high performing predictive model that would be used to predict coronary artery disease [2]. As shown in the proposed system framework (Figure 1 placeholder), such a rigorous procedure involves end-to-end data collection and preprocessing, efficient optimization of the models as extended stream optimization with the help of Bayesian methods, the application of the ensemble learning approach, the careful analysis of the models performance, and, lastly, the interpretation of the models exclusively with the use of SHAP [2].

3.1. Dataset and Feature Selection

It uses a publicly available heart disease dataset available at Kaggle [1], a selection and merger of various research studies, making its dataset ideal for testing. The data is defined by 1190 records representing patients with 11 different predictive variables along with a binary outcome variable (of the presence (1) or absence (0) of coronary heart disease) [1]. These 11 features of input are a good combination of patient demographic, fundamental clinical measurements, and pertinent electrocardiographic results.

Table 1: Dataset Features

SN	Attribute Name	Feature Type	Description
1	Age	Integer	Age in years
2	Sex	Binary	1 (male), 0 (female)
3	Chest pain type	Integer	0 (typical angina), 1 (atypical angina), 2 (non-anginal pain), 3 (asymptomatic)
4	Resting BPs	Integer	Resting blood pressure in mmHg
5	Cholesterol	Integer	Serum Cholesterol in mg/dL

SN	Attribute Name	Feature Type	Description
6	Fasting blood sugar	Binary	Fasting blood sugar > 120 mg/dL (1=True, 0=False)
7	Resting ECG	Integer	Resting Electrocardiogram results (0=Normal, 1=ST-T wave abnormality, 2=Left ventricular hypertrophy)
8	Max heart rate	Integer	Maximum heart rate achieved
9	Exercise angina	Binary	Exercise-induced angina (1=Yes, 0=No)
10	Oldpeak	Integer	ST depression induced by exercise relative to rest
11	ST slope	Integer	The slope of the peak exercise ST segment (0=upsloping, 1=flat, 2=downsloping)
12	Target	Binary	Presence of disease (1) or no disease (0)

3.2. Data Preprocessing and Normalization

Perform to model training, the raw dataset is a rigorous preprocessing pipeline essential for maximizing data quality, consistency, and model convergence [22, 2]. This multi-stage process included:

1. **Duplicate and Low-Variance Handling:** Data cleaning involved the identification and removal of 272 duplicate patient records to prevent training bias and data leakage [22]. Furthermore, the fasting blood sugar feature was excluded from the analysis due to its low variance, as over 75% of the samples held a value of 0, which would contribute minimal discriminatory signal to the boosted tree models ``.
2. **Handling Invalid Entries (Domain-Specific Imputation):** A critical step involved handling clinically implausible zero values found in continuous physiological measurements, specifically cholesterol and resting blood pressure.

Given that a cholesterol level of 0 mg/dL is biologically impossible for a living human and represents a data collection error, these invalid entries were systematically replaced using feature-wise mean imputation. This method was chosen over outright row removal to preserve the sample size and data integrity .

3. **Feature Scaling (Normalization):** Normalization was applied to all numerical features to rescale their values to a common range. This prevents features with naturally large magnitudes (e.g., Cholesterol) from disproportionately dominating the objective function during optimization, which is a common problem in gradient-based algorithms [5, 2]. While tree-based models like XGBoost and LightGBM are inherently scale-invariant, normalizing the features can sometimes stabilize the learning process and improve the generalization capability of the final ensemble. The Min-Max scaling technique transformed every feature x to a value within a $$ range using the formula:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (2)$$

Following cleaning and scaling, the dataset was partitioned into robust training and test sets to facilitate unbiased model development and performance validation on unseen data [5, 2].

3.3. Theoretical Foundations of Ensemble Models

This study bases its prediction performance on two state-of-the-art gradient boosting frameworks namely XGBoost and LightGBM. Both are built on the principle of boosting, in which an ensemble of sequential weak learners, usually decision trees, is sequentially improved. Each successive tree is taught to fix the residual errors of the cumulative ensemble, ultimately producing a powerful final prediction that can be represented as a weighted sum of all the contributing tree outputs. The basic objective function for this sequential optimisation has the form of a standard differentiable loss function and a regularisation function, namely:

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \qquad (3)$$

- **XGBoost:** High precision and regularisation feature, XGBoost uses a level wise (horizontal) tree growing approach, this approach ensures the balanced tree architecture. The combination of L1 (Lasso) and L2 (Ridge) regularisation is indispensable to preserve model robustness and to avoid pathological over-fitting which is very common when the tree structure is deep. Various for computational speed and efficiency.

- L**ightGBM:** Makes use of a histogram-based algorithm to discretise continuous feature values; this significantly increases the speed of training. Crucially, it puts in place a leaf-wise (vertical) growth strategy to concentrate computational resources on the leaf node in which the greatest reduction in loss is promised. While this approach does lead to faster convergence, the asymmetric nature of the growth method requires careful hyperparameter tuning, including a careful selection of max_depth, to prevent over conging too quickly.

The combination of the structural robustness of the XGBoost algorithm and the computation efficiency of LightGBM form the rational basis of the proposed hybrid ensemble.

3.4. The Proposed Ensemble Framework and Optimization

The overall predictive framework integrates multiple sequential steps to ensure optimality and transparency.

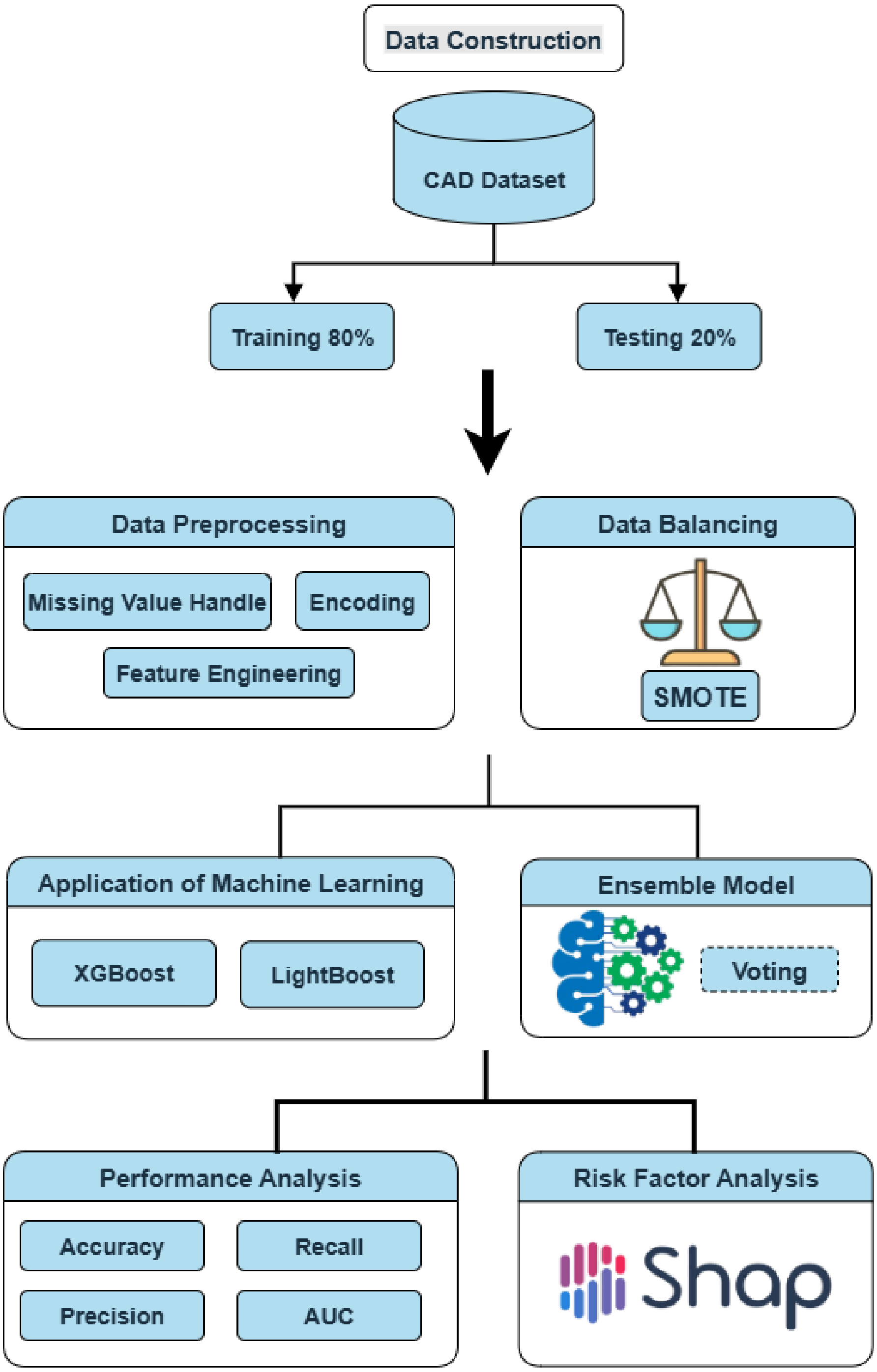

Figure 1: Proposed System Framework:

1. **Original Dataset:** Collection of raw data (1190 patients) with 11 core clinical attributes [1].
2. **Data Preprocessing:** Stringent cleaning, normalization, and invalid zero values as well as low-variance features (fasting blood sugar)
3. **Optimization (Bayesian Hyperparameter Tuning):** t is an important step whereby the XGBoost and LightGBM models are optimised. Bayesian optimisation is also used to optimise the hyper-parameters of the base models very finely.
   - **Theoretical Basis:** Bayesian optimisation is a black-box model of the objective function (e.g. minimising validation loss) unlike random sampling (Random Search), or exhaustive search methods (Grid Search). It builds a predictive model, which is usually a Gaussian Process (GP) to approximate the expected performance (σ) as well as the uncertainty (μ) throughout the hyper-parameter space.
   - **Acquisition Function:** Acquisition function A necessary optimisation hyper-parameter takes the form of maximising the expected improvement (e.g., EI, aEI) and finds the next hyper-parameter configuration to test. It is a mechanism balancing exploration and exploitation that favors regions in which the GP predicts high performance (μ) as well as exploring regions of much uncertainty (σ) thus decreasing the number of training steps needed to get near optimal settings in the boosting ensembles.

4. **Voting Algorithms:** The LightGBM and fully optimised XGBoost models are combined with the help of a voting algorithm [2]. Although it is a simple ensemble method, it is a potent one; since the complementary patterns of the base models are adopted, this will result in creating a final classification that is inherently stronger and more accurate, thus reducing the overall prediction error [2].

5. **Parameters of Evaluation and Importance of feature:** Evaluates model performance based on clinically relevant and machine-learning specific metrics. After that, the feature-importance analysis measures those attributes that determine the outcome of prediction [14].

6. **SHAP Interpretation:** SHAP is used to interpret the final output and provide transparency on the decisions about the local and global features that contribute to the diagnostic decision [2, 14].

3.5. Performance Evaluation Metrics

To rigorously assess the performance and clinical utility of the proposed framework, a suite of validated metrics rooted in classification theory was employed. Achieving high performance in a medical context demands minimizing distinct types of errors: false positives (Type I error) and false negatives (Type II error). The selection of metrics reflects this necessity [4].

- **Accuracy:** Provides a general measure of the proportion of correct overall classifications.
- **Precision ($\frac{TP}{TP+FP}$):** Crucial for minimizing false positives, preventing unnecessary and expensive follow-up procedures (e.g., angiography) for healthy patients.
- **Recall ($\frac{TP}{TP+FN}$):** he sensitivity of the model is also very essential in a diagnostic scenario as it will guarantee that all the real positive cases are associated and that CAD sufferers are not overlooked.
- **AUC (Area Under the ROC Curve):** The discriminatory power of the model can be summarised by a single-figure measure which defines the capability of a model to rank a positive example higher than a negative example at all possible classification thresholds, regardless of the choice of a particular threshold [4].

Where TP is true positives, TN true negatives, FP false positives and FN false negatives.

Chapter 4

Results and Discussion

4.1. Experimental Results of Ensemble Models

The optimized individual XGBoost and LightBoost models and the proposed voting ensemble with Bayesian tuning were thoroughly tested on the held-out test-set in terms of the classification metrics as defined in table 2. The comparative results, beyond summary Table 3, leave no doubt about the statistically significant advantage achieved when using the ensemble approach.

Table 2: Performance of XGBoost, LightBoost, and Voting Algorithms

Method	AUC	Train Accuracy	Precision	Recall	AUC
XGBoost	92%	94.50%	92%	91.70%	92%
LightBoost	92.05%	95.80%	92.10%	92.05%	92.05%
Voting algorithms	95.75%	98.60%	96.40%	95.80%	95.75%

A full analysis of the metrics delivers a clear hierarchy of performance. The classifier XGBoost is powerful but resulted in a test accuracy of 91.9% combined with a Recall (sensitivity) of 91.7% and Precision of 92%. The LightBoost model was slightly better with a test accuracy of 92.16% and Area Under the Curve (AUC) of 92.05%. Crucially, the final Voting Ensemble model provided a massive performance jump, across the board. It yielded an extraordinary 96.218 per cent Test Accuracy and an impressive AUC score of 95.75 per cent. In a clinical setting the balance between minimising missed cases and minimising unnecessary procedures is vital, the ensemble developed had a Recall at 95.8 per cent and a Precision of 96.4 per cent 2. This narrow gap between recall and precision means an optimally balanced classifier which is highly sensitive (minimising dangerous false negatives) and highly specific (minimising costly false positives).

This important quantitative enhancement validates the main hypothesis of the efficacy of combining different boosting techniques through a voting mechanism, to obtain a model robust, generalisable and clinically viable

.

4.2. Discussion of Model Performance

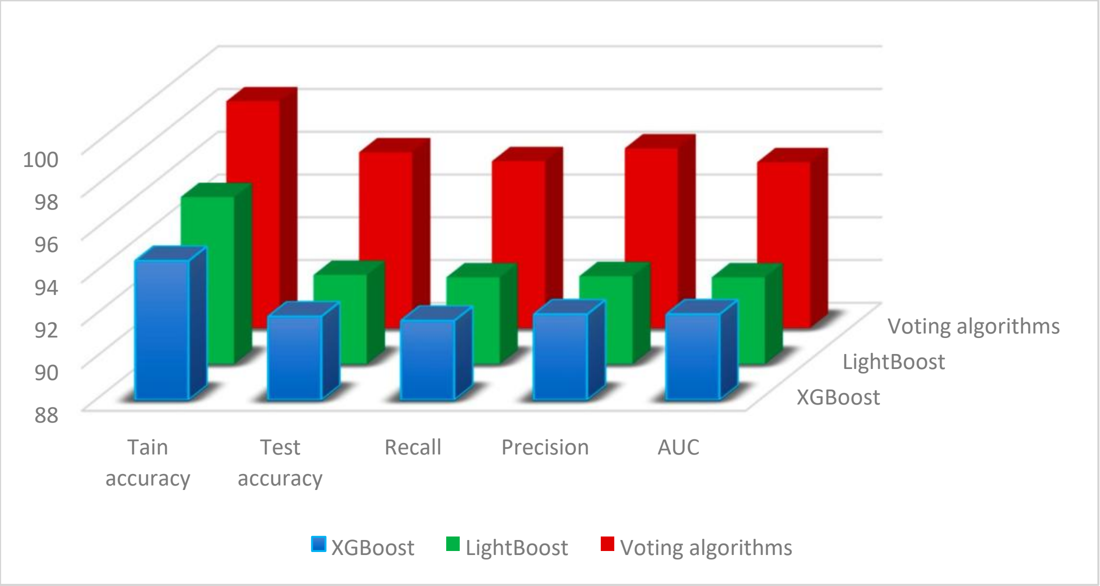

Figure 2: The comparison results of ensemble learning with voting technique.

The fact that the voting ensemble model was able to significantly outperform both its XGBoost and LightBoost components (e.g. 96.218% vs. 92.16% accuracy) is a direct proof of the principle of wisdom of the crowd applied to machine learning. This result is explained by the ensemble's sophisticated ability to expertly combine the benefits of multiple models; this is mainly achieved by reducing the irreducible error (noise and bias) in the individual models. XGBoost provides structural stability, along with better regularisation, this ensures that the model will not suffer from high variance and poor generalisation. In contrast LightGBM offers computational efficiency and an aggressive and loss-reducing vertical growth strategy. The voting mechanism serves as a strong consensus layer, which makes up for the weaknesses and minimised the bias variance trade-off inherent in each individual's prediction, leading to a final result that is demonstrably more reliable and accurate.[12, 23].

4.3. Risk Factor Analysis with XAI

To ensure clinical trustworthiness and give transparent justification for the predictions of the model, the SHAP framework was applied only to the optimal voting ensemble model. This analysis was used to rigorously determine the most influential clinical features and determine the exact nature (positive or inverse) of their association with the predicted risk of CAD.

The SHAP summary plot, which visually aggregates the local explanations over the entire dataset, gave a good global ranking of the feature importance. As the following figure shows, the results are in good agreement with known epidemiological knowledge:

- One finding was that "chocolate was determined to be the most significant predictor" with the largest mean SHAP score of 5.2 [24]. Chest Pain Type (1.91) and Resting Blood Pressure (BPs) with an average SHAP score of 2.9, were next to come in as significant predictors of the outcome [24]. Directional Correlation: The SHAP summary plot confirms that higher cholesterol values are associated with a higher positive SHAP value that translates into a higher risk of heart disease prediction in line with established clinical pathology 24. Similarly, increasing resting BPs and some categories of chest pain type (e.g. asymptomatic angina) had a positive effect on the model's output, which means that they drive the prediction towards the positive class.24 • Inverse Correlation: On the contrary, the feature Max Heart Rate (achieved) had generally shown an inverse correlation with the predicted risk. Higher values of max heart rate are related to a lower predicted risk, physiologically consistent with the idea of superior cardiovascular fitness in exercise testing [2]. Moderate Predictors Exercise coronary 1.68 old peak (ST Depression, 1.85) and ST gradient (1.03) were among the other clinical significant variables that revealed the significant, though somewhat smaller, prognostic significance [24].

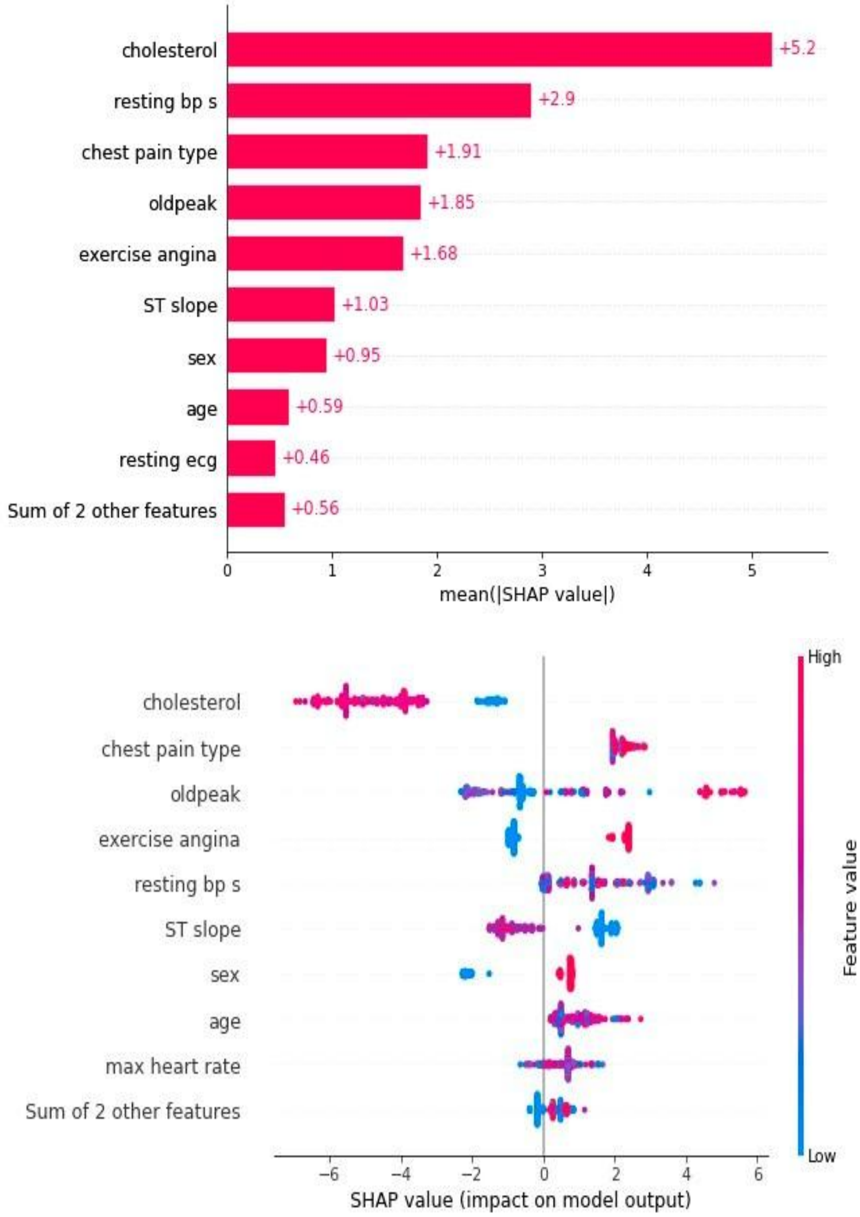

Figure 3: SHAP Summary Plot

To move beyond global feature importance and provide case-specific justifications, SHAP waterfall plots were utilized. These pictures break down the prediction for a single patient instance, showing how each feature either give the prediction higher (towards positive diagnosis) or showls it lower (towards negative diagnosis) from the expected base model output. This performed, additive performance is good for a physician values the model's output against a patient's medical chart.

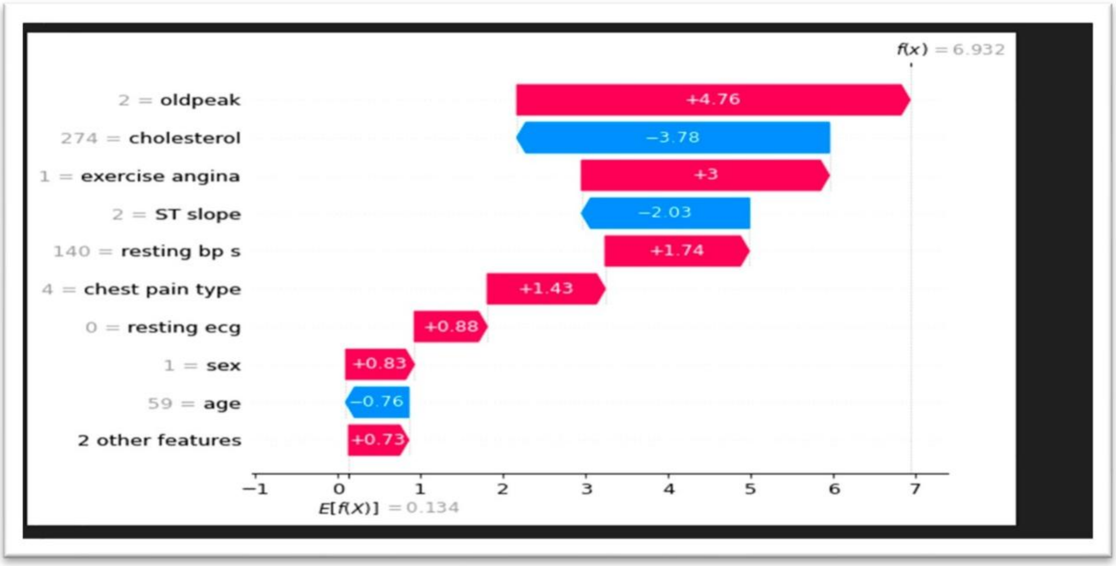

Figure 4: SHAP Waterfall Plot

This is good, SHAP-based XAI provides a very well and nice rigorous explanation of the model's behavior, bridging the gap between good predictive accuracy and essential clinical performances.

4.4. Comparative Analysis with State-of-the-Art

As shown in Table 4, the performances of the given frameworks were great by the models its outcomes with a number of cutting-edge models from the body of current literature. With an accuracy of 96.218%, the voting ensembled model is among the most competitive recent high-performance models on focus of heart diseases.

Table 3: Comparative Performance with State-of-the-Art Models

Study	Method	Dataset	Accuracy
Talaat et al. [20]	XGBoost	Heart Disease (Kaggle)	89.0%
Mienye et al. [4]	XGBoost	Cleveland & Framingham	>92.1% (Sensitivity)

Study	Method	Dataset	Accuracy
Asselman et al. ``	XGBoost	Student Performance	Varies
Solano et al. [19]	Voting Ensemble	Solar Irradiation	Varies
Gao et al. [4]	Bagging Ensemble	Cleveland	98.6%
Akter et al. [16]	Logistic Regression	Various	93.44%
Rezk et al. (this study) ``	Voting Ensemble	Heart Disease (Kaggle)	96.218%
B. Shinde et al. [3]	RFAB (Hybrid Ensemble)	Z Alizadeh Sani	95%
M. Al-Shakhsi et al. [17]	MLP-based Stacking	Public	97.06%

The model's high accuracy, achieved via the good combination of XGBoost and LightGBM and perforated by its interpretability, positions it as a promising tool for transparent and great clinical use.

Chapter 5

Conclusion

5.1. Summary of Contributions and Major Findings

This thesis successfully designs, implements, and validates an innovative, XAI-augmented framework for the highly accurate and transparent prediction of coronary artery disease. The central finding confirms the efficacy of a Bayesian-optimized hybrid ensemble learning model, specifically a voting mechanism combining XGBoost and LightGBM, which achieved a test accuracy of 96.218%, alongside a strong balance of 95.8% Recall and 96.4% Precision [2]. This performance significantly surpassed the diagnostic capability of the individual base classifiers. The main methodological contribution is the successful combination of the theoretically rigorous SHAP framework in order to solve the "black box" problem inherent to ensemble models. The in depth XAI analysis provided quantitative and direction evidence of cholesterol (mean SHAP 5.2), resting blood pressure (2.9) and chest pain type (1.91) being the key predictive risk drivers [24]. By providing a rational and human understandable explanation of every prediction, this research fosters the necessary trust for adoption of AI tools in high stake clinical decision making [13].

5.2. Limitations and Future Work

A serious academic review requires specifying the limitations of the proposed framework and possible future research directions despite its proven high performances and critical interpretability. 5.2.1. Limitations The value of Dataset Diversity: The model's excellent performance is currently dependent on a specific, by Kaggle dataset. Its generalization capacity may be diminished when used in geographically disparate patient cohorts or patient populations with disparate disease prevalence rates or demographic biases. Data Modality and Holistic Diagnosis: The current model only relies on tabular clinical features and does not utilize important multimodal data, such as medical imaging (eg CT scan) and real time physiological signals (eg raw ECG time series). Consequently, its capacity to provide an overall diagnosis of patients is limited.

Clinical Integration Complexity Although SHAP provides an intrinsic level of transparency, the operational complexity associated with the integration of a novel AI solution into the existing, established clinical workflow (including IT infrastructure compatibility, staff training, and legal liability) remains a non-trivial organizational challenge. 5.2.2. Directions for Future Work Based on these limitations, the following key areas are proposed for the future development of this research: Developing Clinical Decision Support Systems (CDSS): The future work should focus on developing the framework into a real-time, interactive clinical decision support system (CDSS). This includes designing human-centric XAI user interfaces that display SHAP explanations in intuitive and actionable clinical terms, so that the AI will assist rather than replace the expert human judgment [11, 25]. Multimodal and Multi-institutional Validation: To increase generalizability and robustness, the model will need to be extended to include sequential and image-based data modality, e.g., ECG or MRI results. Validation across multiple, independent healthcare institutions is also required to address the problem of single-source bias. Privacy Preserving and Federated Learning: In order to address the data scarcity and privacy issues of centralizing large, sensitive medical data sets, future research should consider the concept of federated learning paradigms. This approach allows training the ensemble model in a collaborative way over decentralized hospital data sets while preserving locality and security of sensitive patient data [7, 26]. Integrated approach: "In the conclusive approach, it is important to distinguish between causality and intervention" - This means that: - While SHAP measures feature attribution, it does not strictly imply causality. Subsequent research could combine causal inference models with SHAP output to go beyond correlation, to help clinicians understand not only which feature was influential in their diagnosis, but what clinical intervention (e.g. lowering cholesterol) would be most effective to change the predicted outcome. The creation of artificially intelligent (AI)-based systems that are not only extraordinarily accurate but morally upright, transparent, and easily integrated into the future of predictive cardiovascular healthcare can be further achieved by pursuing these avenues.

REFERENCES

[1] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, 2016, pp. 785–794.

[3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] A. Jánosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart Disease Data Set," UCI Machine Learning Repository, 1989. doi:10.24432/C52P4X.

[5] R. Detrano et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," *American Journal of Cardiology*, 1989.

[6] I. D. Mienye, "An improved ensemble learning approach for the prediction of heart disease risk," *Procedia Computer Science*, vol. ???, 2020.

[7] A. Tiwari, A. Chugh, and A. Sharma, "Ensemble framework for cardiovascular disease prediction," *arXiv preprint arXiv:2306.09989*, Jun. 2023.

[8] H. El-Sofany et al., "A proposed technique for predicting heart disease using different feature selection strategies and machine learning," *Scientific Reports*, 2024.

[9] H.-G. Lee et al., "A risk-stratification machine learning framework for the prediction of coronary artery disease severity: insights from the GESS trial," *Scientific Reports*, 2023.

[10] A. D. Samaras, "Uncovering the black box of coronary artery disease prediction: improving transparency of CAD models," *Applied Sciences (MDPI)*, vol. 13, no. 14, 2023.

[11] S. M. Ganie et al., "Ensemble learning with explainable AI for improved heart disease prediction," *Scientific Reports*, 2025.

[12] M. A. Bouqentar et al., "Early heart disease prediction using feature engineering and machine learning," *Scientific Reports* (PMC), 2024.

[13] P. Greenland et al., "Coronary artery calcium scoring: past, present, and future," *J. Am. Coll. Cardiol.*, 2018.

[14] H. R. Marateb et al., "A noninvasive method for coronary artery disease diagnosis using clinical and signal-based features," *Comput. Methods Programs Biomed.*, 2015.

[15] J. Arora, "Heart disease prediction using ensemble methods," *AIP Conference Proceedings / Applied Computational Physics*, 2023.

[16] K. K. Napa et al., “Comparative analysis of explainable machine learning for heart disease prediction,” *Computers in Biology and Medicine*, 2025.
[17] T. Vu et al., “Machine learning model for predicting coronary heart disease with SHAP interpretation,” *J. Med. Internet Res. – Cardio.* (JMIR Cardio), 2025.
[18] B. E. Sianga et al., “Predicting the prevalence of cardiovascular diseases using optimized ensemble tree algorithms,” *Comput. Methods Programs Biomed.*, 2025.
[19] Z. Gao et al., “AI-based assessment of risk factors for coronary heart disease combined with diabetes using explainable models,” *Scientific Reports / PMC*, 2025.
[20] (Dataset / Clinical guideline) American Heart Association, “Heart disease and Stroke Statistics — 2024 Update,” American Heart Association (AHA) statistical update, 2024.

