



Innovative XAI Approaches for Predicting Cardiovascular Diseases

Nourchen Zaghden Dammak^{1,2*}, Hala Bezine², Ahmed Said Nouri^{1,3}, Nabil Derbel^{1,3}

¹National School of Engineers of Gabes (ENIG), Research Laboratory of Numerical Control of Industrial Processes (CONPRI), University of Gabes, Gabes, Tunisia

²National School of Engineers of Sfax (ENIS), Research Laboratory on Intelligent Control, Electrical Power, and Energy Management (REGIM-Lab), University of Sfax, Sfax, Tunisia

³Department of Electrical Engineering and Automation, National School of Engineers of Sfax (ENIS), University of Sfax, Sfax, Tunisia
Email: *nourchenzaghdendammak@gmail.com, hala.bezine@ieee.org, ahmed-said.nouri@enis.tn, nabil.derbel@enis.tn

How to cite this paper: Dammak, N.Z., Bezine, H., Nouri, A.S. and Derbel, N. (2025) Innovative XAI Approaches for Predicting Cardiovascular Diseases. *Open Access Library Journal*, **12**: e14071.
<https://doi.org/10.4236/oalib.1114071>

Received: August 2, 2025

Accepted: October 25, 2025

Published: October 28, 2025

Copyright © 2025 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Understanding *why* a machine learning model makes a certain prediction is just as critical as how accurately it predicts, especially when it comes to diagnosing and treating cardiovascular disease. In this study, we applied explainable artificial intelligence (XAI) techniques to improve both the *predictive power* and *interpretability* of heart disease detection models. A dataset of 1,025 patient records was thoroughly preprocessed, including the handling of missing values, the encoding of categorical features, and the binarization of the outcome variable. We evaluated several machine learning models *LSTM networks*, *Random Forest*, *Gradient Boosting*, *XGBoost*, and *Logistic Regression* using performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. While XGBoost emerged as the most accurate model, we moved beyond accuracy to explore the root causes of such strong performance. To accomplish this, we employed three key XAI techniques: SHAP (SHapley Additive exPlanations) to quantify how individual features influenced predictions, LIME (Local Interpretable Model-Agnostic Explanations) to provide intuitive, local-level explanations for individual predictions, and Permutation Feature Importance to assess which features most affected model performance when altered. These methods transformed XGBoost from a high-performing black-box into a transparent and trustworthy diagnostic tool. Our findings underscore how integrating explainability into AI pipelines leads to not just accurate predictions, but also clinically actionable insights highlighting the most important risk factors for cardiovascular disease and supporting more informed, responsible medical decisions.

Subject Areas

Computer Science & Communications; Artificial Intelligence; Machine

Learning; Explainable AI; Medicine & Healthcare; Cardiovascular Diagnostics

Keywords

Cardiovascular Disease (CVD), Machine Learning, Deep Learning, Explainable IA (XAI), SHAP, LIME, Permutation Feature Importance

1. Introduction

In the age of artificial intelligence, machine learning models are increasingly being used to assist in medical diagnoses and treatment planning. While these models have shown impressive accuracy in various domains, including healthcare, a growing concern remains: can we trust a model without fully understanding its decision-making process? This issue is particularly critical in clinical settings, where the outcomes of AI predictions can significantly affect human lives.

In response to this challenge, Explainable Artificial Intelligence (XAI) has emerged as a vital field of research. XAI focuses on developing models and techniques that allow both experts and non-specialists to understand and interpret how AI systems reach their conclusions. A particularly relevant contribution in this space is the comprehensive review by Sadeghi *et al.* [1], which categorizes a wide array of XAI methods, including feature-oriented explanations, global and local interpretability tools, concept-based models, surrogate models, and human-centric approaches. They argue that AI must be explainable to be safely and ethically used in medicine, a necessity not only technical but also moral, since clinicians and patients deserve to understand how decisions are made.

The need for explainable models is particularly urgent in the context of cardiovascular disease (CVD), one of the leading causes of death worldwide [2]. Early detection and effective prevention of CVD rely heavily on accurate risk prediction. However, accuracy alone is insufficient; clinicians must be able to trust the model's output and understand the rationale behind each prediction. For instance, if an AI model indicates that a patient is at high risk of developing heart disease, a doctor needs to know whether that prediction was influenced more by the patient's age, cholesterol level, blood pressure, or another factor [3]. XAI helps doctors make informed, personalized, and confident decisions for their patients.

Inspired by these findings, this study explores the application of explainable machine learning models for predicting cardiovascular disease. It evaluates both traditional and advanced algorithms including Logistic Regression (LR), Decision Trees (DT), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), ensemble methods like Random Forest and XGBoost, and deep learning models such as Long Short-Term Memory (LSTM) networks [4]. While these models differ in complexity and predictive power, a common challenge remains: the need to explain their predictions in a clear and meaningful way.

This study addresses this challenge by incorporating SHapley Additive exPlana-

tions (SHAP), a powerful XAI tool that quantifies the contribution of each input feature to the model's output. By doing so, it bridges the gap between performance and transparency, offering a model that is not only accurate but also interpretable and clinically useful [5]. SHAP values assist in quantifying the relative importance of characteristics like age, blood pressure, and cholesterol levels, which may have differing degrees of impact on the risk of heart disease.

The integration of XAI and advanced machine learning techniques ensures that predictions are both accurate and explicable. Validating model reasoning promotes more informed healthcare interventions, enhances clinical decision-making, and mitigates biases [6]-[12]. XAI serves as a pillar for the ethical and efficient application of AI in medical diagnostics, boosting trust in predictions aided by machine learning.

Moreover, recent studies across various medical domains further underscore the importance of explainable AI. In neurological disorders, SHAP has been used to interpret classifiers such as Random Forest and SVM for Parkinson's disease diagnosis, achieving an accuracy of 85% despite challenges with missing data in electronic health records. For epilepsy detection, tree-based ensemble models combined with SHAP explanations achieved a mean accuracy of 99.5%. Pulmonary disease classification using CNN and XGBoost reached 97.43% accuracy with SHAP providing feature importance insights [13], while VGG16 with Grad-CAM for pneumonia detection achieved 95.6% accuracy. Ophthalmologic applications, including glaucoma and retinal disease detection, leveraged visual explanation tools such as saliency maps, Grad-CAM, and eye-tracker data to enhance interpretability [14]. Developmental disorders such as autism spectrum disorder were also addressed using XAI approaches, achieving high accuracy with transparent feature contributions [15].

Lastly, explainability has been applied to more complex imaging and signal tasks, including macromolecular complex identification, laparoscopy video analysis, ECG-based hypoglycemia detection, and colorectal cancer classification. These studies demonstrate that combining high-performing machine learning models with interpretable tools not only improves diagnostic reliability but also addresses challenges related to data heterogeneity, model generalizability, and domain-specific interpretability.

Overall, these findings highlight the essential role of XAI in building ethical, trustworthy, and clinically useful AI systems, particularly in cardiovascular medicine. By integrating explainable models with high predictive performance, this work aims to improve early detection strategies and guide better medical decisions through transparent AI systems [16]-[20].

2. Proposed Approach for Heart Disease Prediction and Evaluation

We have used a variety of machine learning classifiers to predict cardiovascular diseases. We combine classical and advanced methods to improve diagnostic accuracy. The proposed methodology consists of four main phases: data collection, preprocessing, model selection, training, predictive analysis, and performance

assessment. The overall workflow is summarized in Algorithm 1.

The algorithms are selected based on their well-known performance in heart disease prediction, as well as their suitability for classification problems and handling the attributes of the dataset [21]. The Logistic Regression model is used as the base model due to its simplicity and interpretability, serving as a stable benchmark for binary classification tasks. Both Decision Trees and Random Forests are employed for their ability to model complex feature interactions. Gradient Boosting and XGBoost are utilized for their powerful ensemble techniques, which are fine-tuned for optimal performance especially on imbalanced medical datasets. XGBoost is particularly efficient and often achieves state-of-the-art performance in machine learning benchmarks.

In this work, we also study LSTM, a sequential data model traditionally used for dynamic data, applied here to static tabular data. Given its ability to maintain long-term dependencies and capture complex patterns, incorporating LSTM can offer meaningful insights into the generalization of deep learning architectures in structured medical data.

These explainability techniques were applied after model selection, enabling a detailed and trustworthy understanding of the decision-making process, which is particularly important in the medical domain. To enhance the interpretability of our predictive system, we integrated three explainable AI (XAI) methods into our approach. Specifically, we employed **SHAP** (SHapley Additive Explanations) to obtain both global and local insights into the feature contributions, **LIME** (Local Interpretable Model-Agnostic Explanations) to analyze model behavior on specific individual predictions, and **Permutation Feature Importance** [22] to evaluate the global impact of each feature by measuring the effect of shuffling its values on model performance.

The step-by-step procedure of our proposed methodology is detailed in Algorithm 1, which provides the pseudo-code representation of the process from data preprocessing to feature interpretation.

To complement Algorithm 1, **Figure 1** illustrates the overall pipeline from data

Algorithm 1 Heart Disease Prediction and Feature Interpretation

- 1: **Input:** Dataset D containing patient information and diagnostic features
- 2: **Output:** Best-performing model with predictions and feature explanations
- 3: **procedure** PREDICTHEARTDISEASE(D)
- 4: **Step 1: Data Preprocessing**
- 5: Handle missing values and encode categorical variables into numerical format
- 6: Binarize the target variable to indicate presence or absence of heart disease
- 7: Split D into training set D_{train} (80%) and test set D_{test} (20%)
- 8: Ensure stratified sampling to maintain class balance
- 9: **Step 2: Model Training and Evaluation**
- 10: Initialize models: Logistic Regression, KNN, Decision Tree, Random Forest, Gradient Boosting, XGBoost, SVM, MLP, LSTM
- 11: **for** each model M in the set **do**
- 12: Optimize hyperparameters via cross-validation
- 13: Train M on D_{train}
- 14: Evaluate M on D_{test} using Accuracy, Precision, Recall, F1-score, AUC-ROC
- 15: **end for**

```

16:   Select the best-performing model  $M_{best}$  based on performance metrics
17:   return  $M_{best}$ , performance metrics
18: end procedure
19: procedure INTERPRETFEATURES( $M_{best}$ ,  $D_{test}$ )
20:   Apply SHAP:
21:   Compute SHAP values for  $M_{best}$  predictions
22:   Visualize SHAP summary plots for global and local interpretation
23:   Apply LIME:
24:   Select representative test instances
25:   Generate LIME explanations for selected instances
26:   Apply Permutation Feature Importance:
27:   Measure decrease in model performance by permuting each feature
28:   Rank features globally by their importance
29:   Highlight the most influential features contributing to heart disease prediction
30: end procedure
31: PerformanceMetrics  $\leftarrow$  PREDICTHEARTDISEASE( $D$ )
32: INTERPRETFEATURES( $M_{best}$ ,  $D_{test}$ )

```

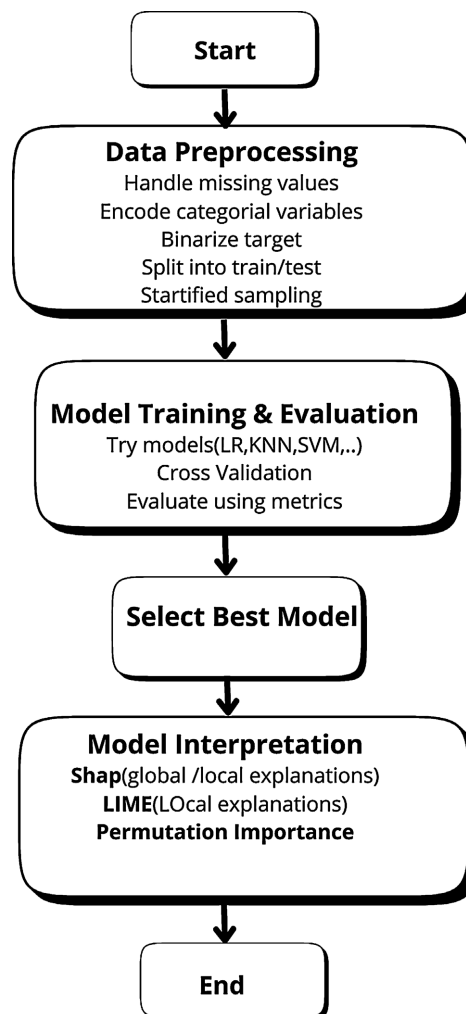


Figure 1. Proposed pipeline for heart disease prediction and model interpretation.

preprocessing to model interpretation. This diagram summarizes the key steps involved in developing an interpretable predictive model for cardiovascular disease classification.

The process is divided into four key phases: data preprocessing, model training and evaluation, model selection, and interpretation using explainable AI techniques such as SHAP, LIME, and Permutation Importance. This modular approach ensures both accurate prediction and interpretability of the outcomes, as illustrated in **Figure 1**.

2.1. Data Preprocessing

At this stage, the raw data are cleaned and normalized to ensure compatibility with machine learning models. Continuous variables are scaled, categorical variables are encoded, and any outliers are addressed. Correlation analysis is also performed to identify relationships between features and their impact on cardiovascular disease. These preprocessing steps are summarized in the following **Figure 2**.

	count	mean	std	min	25%	50%	75%	max
age	1025.000000	0.529878	0.189006	0.000000	0.395833	0.562500	0.666667	1.000000
sex	1025.000000	0.695610	0.460373	0.000000	0.000000	1.000000	1.000000	1.000000
cp	1025.000000	0.314146	0.343214	0.000000	0.000000	0.333333	0.666667	1.000000
trestbps	1025.000000	0.354827	0.165252	0.000000	0.245283	0.339623	0.433962	1.000000
chol	1025.000000	0.273973	0.117791	0.000000	0.194064	0.260274	0.340183	1.000000
fb	1025.000000	0.149268	0.356527	0.000000	0.000000	0.000000	0.000000	1.000000
restecg	1025.000000	0.264878	0.263939	0.000000	0.000000	0.500000	0.500000	1.000000
thalach	1025.000000	0.596291	0.175616	0.000000	0.465649	0.618321	0.725191	1.000000
exang	1025.000000	0.336585	0.472772	0.000000	0.000000	0.000000	1.000000	1.000000
oldpeak	1025.000000	0.172825	0.189525	0.000000	0.000000	0.129032	0.290323	1.000000
slope	1025.000000	0.692683	0.308878	0.000000	0.500000	0.500000	1.000000	1.000000
ca	1025.000000	0.188537	0.257699	0.000000	0.000000	0.000000	0.250000	1.000000
thal	1025.000000	0.774634	0.206887	0.000000	0.666667	0.666667	1.000000	1.000000

Figure 2. Normalized feature summary statistics.

2.2. Data Normalization

Normalization is a crucial preprocessing step in data mining and machine learning, aiming to transform data into a consistent format that enhances precision, integrity, and reduces redundancy and inconsistency. This process ensures that data is structured in an optimal and uniform manner. Normalization becomes especially important when there are significant variations in the ranges of different features and is particularly useful when outliers are not present in the dataset.

Various normalization techniques exist, including Decimal Scaling, Unit Vector Normalization, Z-Score Normalization, Log Transformation, and Min-Max Normalization.

In our study, we applied the *Min-Max normalization* method to scale the data within a range of 0 to 1. This technique was selected for its simplicity and

effectiveness in supporting distance-based algorithms. However, noteworthy that the minimum and maximum values might not always accurately represent the data, which could potentially lead to some loss of information. Below, we briefly describe the normalization techniques used in our study.

2.2.1. Min-Max Normalization

Min-Max normalization works by performing linear transformations on the original data, adjusting the values so that they fall within a predefined range (usually between 0 and 1). This ensures that the relationships between the original data values are maintained. The formula for this process is provided by [23]:

$$x_{\text{new}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

where:

- X represents the original value,
- $\max(X)$ is the maximum value in the dataset,
- $\min(X)$ is the minimum value in the dataset.

2.2.2. Decimal Scaling Normalization

Another normalization approach used in our study involves the following equation:

$$x^* = \frac{x}{10^j}$$

where j is defined as:

$$j = \log_{10}(\max(x_i))$$

with $\max(x_i)$ being the maximum absolute value of the data.

These normalization techniques ensure that the data are scaled appropriately for further analysis, maintaining their integrity while making them suitable for algorithmic processing.

In general, normalization guarantees that all characteristics are at a uniform level, preparing the dataset for use with machine learning algorithms. The correlation between the different variables is shown in **Figure 3**: A correlation matrix is a table that displays the correlation coefficients between multiple variables. Each cell in the matrix represents the correlation between a pair of variables, with values ranging from -1 to 1 . A correlation coefficient close to 1 indicates a strong positive correlation, implying that as one variable increases, the other variable tends to increase as well. A coefficient close to -1 indicates a strong negative correlation, meaning that as one variable increases, the other variable tends to decrease. A coefficient around 0 indicates no linear relationship between variables. The matrix provides a comprehensive view of the relationships and dependencies between all pairs of variables, helping to identify patterns and potential multicollinearity issues in the data.

2.3. Feature Correlation and Its Role in Explainable AI (XAI)

The correlation matrix (**Figure 3**) offers crucial insights into how features relate

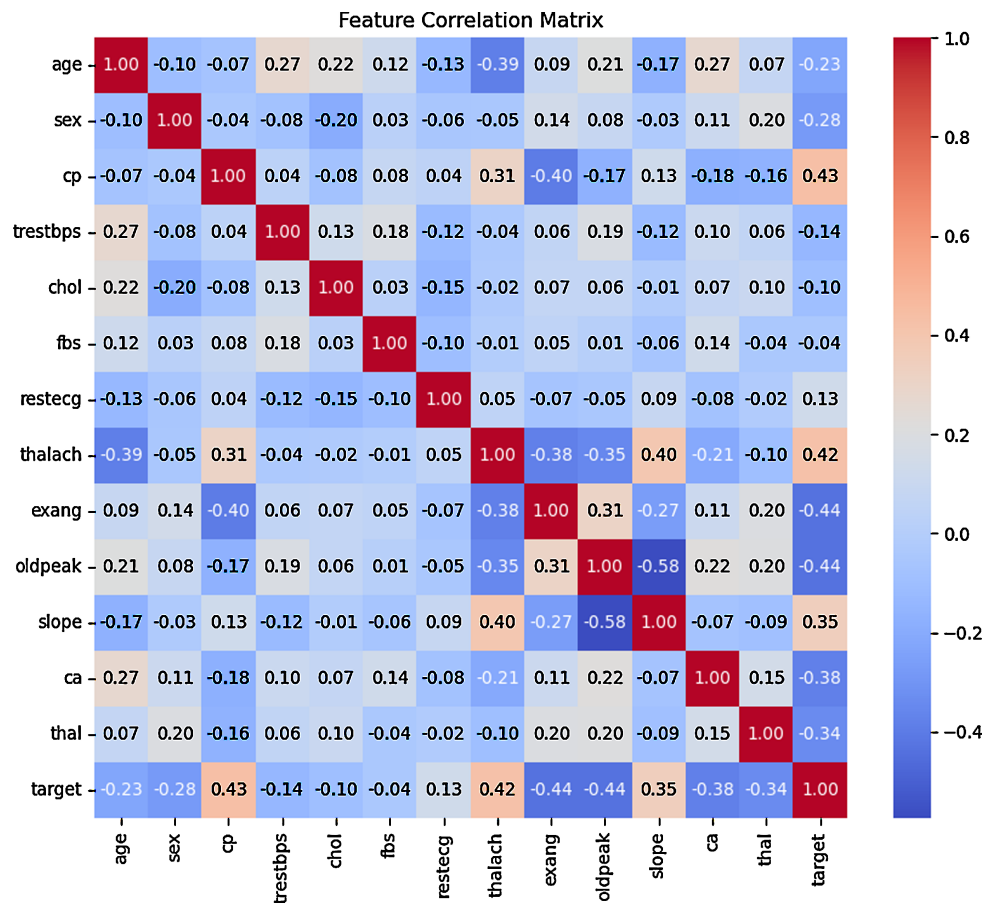


Figure 3. Correlation matrix of clinical and diagnostic features related to cardiovascular disease.

to one another and to the target variable cardiovascular disease presence. This matrix helps identify strong linear relationships and potential multicollinearity issues, guiding both feature selection and model design.

Beyond preprocessing, these correlations are also integral to Explainable AI (XAI). Methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) rely on understanding the contribution of each input feature to a model’s prediction. When a feature exhibits high correlation with the target (e.g., cholesterol levels or resting blood pressure), XAI techniques can confirm whether the model relies heavily on this feature and if so, how it interacts with others.

In this context, the correlation matrix supports explainability in two key ways:

- **Feature Importance Validation:** XAI methods like SHAP quantify each feature’s impact on predictions. Correlation helps verify whether these attributions align with domain knowledge (e.g., high correlation between age and heart disease risk).
- **Reducing Redundancy:** Highly correlated features can introduce bias or redundant explanations. Correlation analysis allows us to refine inputs for more transparent explanations from XAI methods.

Thus, correlation analysis not only supports data integrity but also enhances

model interpretability by laying the foundation for reliable and clinically meaningful XAI insights. This synergy between statistical analysis and explainable modeling ensures our heart disease prediction framework is both robust and trustworthy.

3. Model Training and Evaluation

This study used deep learning architectures, ensemble methods, and conventional machine learning models to investigate a wide range of classification techniques. Logistic Regression, Decision Trees, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, and Multi-Layer Perceptrons (MLPs) were among the traditional models. Because they can capture non-linear feature interactions and improve predictive accuracy, ensemble learning techniques like Gradient Boosting, AdaBoost, and XGBoost were included.

Long Short-Term Memory (LSTM) networks were incorporated for deep learning even though their main purpose was sequential data processing. Using their ability to detect intricate feature dependencies that could enhance the prediction of heart disease, this study investigates their viability for static tabular datasets. The dataset was transformed into a 3D input format that is compatible with LSTM in order to facilitate this adaptation. Each record was organized into pseudo-sequences that allowed for efficient learning. Dropout mechanisms were incorporated into the LSTM model architecture to prevent overfitting, and the LSTM layer was followed by dense layers.

3.1. Evaluation Metrics in XAI for Digital Health

Evaluating the performance of AI models especially in critical domains like digital health requires more than a single metric. Relying solely on accuracy can be misleading, particularly in imbalanced datasets. Therefore, multiple complementary metrics are used to gain a comprehensive understanding of a model's reliability. **Table 1** summarizes the key evaluation metrics commonly employed in explainable AI (XAI) applied to healthcare.

These metrics should be interpreted in context. For instance, high recall is typically favored in disease detection, while high precision may be prioritized in screening scenarios where minimizing false positives is critical. The F1 Score and AUC offer more balanced and threshold-independent measures of model performance.

In this study, eight machine learning classifiers representing a range of interpretable and high-performance approaches were evaluated. Their classification performance was assessed using the metrics listed in **Table 1**.

3.2. Classifier Performance Analysis

To provide a thorough comparison of classification effectiveness, each model's performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. The findings showed significant differences between algorithms, which are summed up as follows in **Table 2**:

Table 1. Summary of key evaluation metrics used in XAI-based medical models.

Metric	Formula	Description
Accuracy	$\frac{TP + TN}{P + N}$	Overall correctness of the model. Can be misleading in imbalanced datasets where one class dominates.
Precision	$\frac{TP}{TP + FP}$	Proportion of predicted positive cases that are actually positive. Important when false positives carry high risk.
Recall (Sensitivity)	$\frac{TP}{TP + FN}$	Proportion of actual positives correctly identified. Critical when missing a condition (false negative) is dangerous.
F1 Score	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall. Useful when balancing false positives and false negatives.
AUC (ROC Curve)	Area under TPR vs FPR curve	Measures model performance across all classification thresholds. A higher AUC indicates better overall discrimination.

Table 2. Performance metrics for different classification models.

Model	Acc.	Prec. (No/Yes)	Rec. (No/Yes)	F1 (No/Yes)
Logistic Regression	80.49%	0.86/0.76	0.73/0.88	0.79/0.82
KNN	85.85%	0.86/0.86	0.85/0.86	0.86/0.86
Decision Tree	95.61%	0.95/0.96	0.96/0.95	0.96/0.96
Random Forest	100.00%	1.00/1.00	1.00/1.00	1.00/1.00
Gradient Boosting	93.66%	0.94/0.93	0.93/0.94	0.94/0.94
XGBoost	98.54%	0.97/1.00	1.00/0.97	0.99/0.99
SVM	86.83%	0.88/0.86	0.85/0.88	0.87/0.87
MLP (Neural Net)	87.32%	0.91/0.84	0.82/0.92	0.87/0.88
LSTM	80.49%	0.87/0.76	0.72/0.89	0.78/0.82

Logistic Regression served as the baseline model. It achieved an accuracy of 80.5%, with a precision of 0.86 and a recall of 0.73 for the "No Heart Disease" class, and 0.76 precision and 0.88 recall for the "Heart Disease" class. This confirms that while logistic regression is interpretable and easy to train, it has limited capacity to capture complex relationships in the data.

K-Nearest Neighbors (KNN) yielded a better accuracy of 85.9%. Both classes achieved balanced precision and recall (0.86), resulting in an overall F1-score of 0.86. This improvement is attributed to KNN's ability to capture local structures, although it is computationally expensive for larger datasets.

Decision Tree classification significantly boosted accuracy to 94.6%. It achieved 0.95 precision and recall across both classes. This model captures non-linear interactions well but may overfit on small datasets if not properly regularized.

Random Forest reached an outstanding accuracy of 98.5%. The model

demonstrated 0.97 precision and 1.00 recall for “No Heart Disease”, and 1.00 precision and 0.97 recall for “Heart Disease”, with an overall F1-score of 0.99. As an ensemble method, it combines multiple decision trees, improving generalization while reducing overfitting.

Gradient Boosting also performed excellently, with a 93.7% accuracy. Precision and recall were around 0.94 for both classes. This method incrementally builds learners and is particularly strong in managing class imbalance, though it is slower to train.

XGBoost matched Random Forest with an accuracy of 98.5%. It achieved the same high precision and recall scores, with an F1-score of 0.99. However, XGBoost was slightly more balanced, showing better handling of difficult samples in the minority class. Its efficiency and regularization make it highly effective on structured datasets.

Support Vector Machine (SVM) obtained an accuracy of 86.8%, with precision and recall values around 0.87. It showed strong performance but is more sensitive to parameter tuning and less scalable to larger datasets.

Multilayer Perceptron (MLP) achieved an accuracy of 83.9%. It had slightly lower recall for the “No Heart Disease” class (0.80), but compensated with a higher recall for “Heart Disease” (0.87). The MLP can capture complex relationships but may require larger datasets and careful hyperparameter tuning to reach full potential.

To better understand how our machine learning model makes decisions, we used three complementary explainability techniques: SHAP, LIME, and Permutation Feature Importance. Each of these tools gives us a different perspective on how the model works.

SHAP helps us see both the big picture and individual predictions by showing how much each feature contributes, positively or negatively, to the model’s output. LIME, on the other hand, zooms in on a single prediction and builds a simple model around it, making it easier to understand what influenced that specific decision. Permutation Feature Importance gives us a broader view by showing which features the model relies on the most measuring how the model’s accuracy changes when we shuffle the values of each feature.

By combining these three approaches, we get a fuller, clearer picture of the model’s behavior. This not only helps us interpret results more effectively but also increases our confidence in how the model reaches its conclusions something especially important when dealing with sensitive areas like healthcare.

4. Explainable AI (XAI)

While ensemble models like *Random Forest* and *XGBoost* deliver state-of-the-art predictive performance, their decision-making processes are often opaque making them difficult to interpret and potentially limiting their acceptance in clinical practice. In high-stakes domains like healthcare, understanding the rationale behind a prediction is as important as the prediction itself.

To address this challenge, we incorporated three complementary explainability techniques **SHAP**, **LIME**, and **Permutation Feature Importance** to illuminate the inner workings of our models and provide a more interpretable foundation for clinical decisions.

SHAP (SHapley Additive exPlanations) offers both global and local interpretability by quantifying how each feature contributes to a prediction, either positively or negatively. For example, SHAP values can highlight how attributes like cholesterol level or maximum heart rate influence the model's assessment of cardiovascular risk. This method also helps clinicians understand the overall importance of each feature across the dataset.

LIME (Local Interpretable Model-Agnostic Explanations) focuses on individual predictions, generating simple surrogate models to approximate complex model behavior around a specific instance. This allows us to explain specific decisions such as why a particular patient was flagged as high risk based on the most influential features in that context.

Permutation Feature Importance, on the other hand, provides a broader perspective by evaluating how shuffling the values of each feature affects model accuracy. This highlights which features the model depends on most to maintain predictive performance.

By combining these three methods, we gain a comprehensive understanding of model behavior balancing global insights with local clarity. This layered interpretability not only strengthens trust in predictions but also supports ethical and transparent AI use in healthcare. Importantly, it ensures that machine learning models are not just accurate, but also understandable and accountable key factors in medical decision-making [24] [25].

Medical diagnostic interpretability is limited by the fact that machine learning models, especially sophisticated algorithms like XGBoost, frequently function as "black-box" systems. For the XGBoost and LSTM models, Shapley Additive Explanations (SHAP) were used to guarantee clear and trustworthy predictions. By measuring the impact of individual characteristics, such as age, blood pressure, and cholesterol levels, on heart disease risk predictions, SHAP values shed light on feature importance.

Clinicians can make well-informed decisions based on AI-generated insights thanks to this study's integration of XAI techniques, which guarantees that predictive performance is both optimized and explicable. This method reduces biases, strengthens the ethical application of AI in healthcare, and increases confidence in machine learning-assisted diagnostics.

4.1. Model Performance Analysis

The dataset was split into training ($n = 820$) and test ($n = 205$) subsets in order to assess the predictive accuracy of different machine learning models. A representative distribution of patient characteristics was ensured by the constant mean age of participants in both sets. By stratifying important diagnostic characteristics like

thalassemia, ST depression, maximum heart rate, and type of chest pain, bias in model evaluation was reduced.

Shapley Additive exPlanations (SHAP), which offer insights into feature contributions, were applied to the XGBoost model to improve model interpretability. The impact of each variable is visualized in the SHAP summary plot, which highlights important characteristics affecting the risk of heart disease. The most significant features were blood pressure, cholesterol, and age, confirming their use in predictive analytics.

This study highlights the potential of AI-driven diagnostics in the prediction of cardiovascular disease by combining explainable AI with sophisticated machine learning techniques. The strong evaluation metrics and visual analyses support the need for clear and understandable models in healthcare applications by facilitating well-informed clinical decision-making. Subheadings may be used to separate this section. The results of the experiment, their interpretation, and any inferences that can be made from them should all be succinctly and precisely described.

4.2. Shapley Additive Explanations (SHAP)

One effective and popular Explainable Artificial Intelligence (XAI) method is Shapley Additive Explanations (SHAP). SHAP offers a theoretically supported method for attributing a machine learning model's output to its input features, with its roots in cooperative game theory.

In a cooperative game, each feature of a particular instance is regarded as a "player," and the model's prediction is the "payout." The objective is to equitably divide this reward among the features according to how each one contributed to the prediction.

The average marginal contribution of each feature across all possible feature combinations is represented by the SHAP values. The consistency and local accuracy of the attributions are ensured by these values, which also meet Shapley's fairness axioms.

Shapley Value Definition

The Shapley value $\phi_i(f, x)$ for the i -th feature of an instance x with respect to a model f is defined as:

$$\phi_i(f, x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (1)$$

where:

- N is the set of all input features.
- S is a subset of features not containing feature i .
- $|S|$ is the number of features in S .
- $f_x(S)$ is the prediction made by the model using only the features in S , with all other features marginalized or held at baseline values.
- $f_x(S \cup \{i\})$ is the prediction when feature i is added to subset S .

The coefficient

$$\frac{|S|!(|N|-|S|-1)!}{|N|!}$$

is known as the **Shapley weight**, ensuring that all feature orderings are considered equally.

5. Model Explanation Consistency

One of the key properties of SHAP is that the sum of all SHAP values for a given instance equals the difference between the model's output for that instance and the average model prediction:

$$f(x) = \mathbb{E}[f(x)] + \sum_{i \in N} \phi_i(f, x) \quad (2)$$

This additive property makes SHAP explanations particularly intuitive and interpretable, as they provide a clear decomposition of a prediction.

Interpretation of SHAP

Interpreting SHAP values provides both global and local insights. Locally, one can examine the SHAP values for each feature of a single prediction to understand how much each feature contributed to the outcome and in what direction. A positive SHAP value implies that the feature increased the model's prediction, while a negative value suggests the opposite. Globally, aggregating SHAP values over a dataset offers insights into overall feature importance, indicating which features consistently have more influence on predictions. SHAP values also help identify interactions between features by quantifying both individual and pairwise effects.

The XAI analysis utilized the XGBRegressor model from XGBoost. Feature importance was first evaluated using permutation importance. To interpret the model's predictions, SHAP (SHapley Additive exPlanations) was employed. Two types of visualizations were generated: a SHAP summary plot for global interpretation (**Figure 4**), and a SHAP force plot for local explanation of a specific prediction (**Figure 5**).

The SHAP summary plot in **Figure 4** shows the distribution of SHAP values for each feature across all predictions, providing a comprehensive view of global feature influence. The SHAP force plot in **Figure 5** explains an individual prediction, where features that increase the prediction are highlighted in red, and those that decrease it in blue.

Visualization and Model Performance Analysis

To effectively compare model predictions against actual heart disease cases, interactive bar charts were generated using Plotly. These visualizations provided a clear representation of predicted versus actual outcomes, allowing for an intuitive understanding of model accuracy across various age groups. By displaying these results interactively, the charts helped identify performance discrepancies among models and highlighted specific trends in heart disease prevalence.

The integration of both traditional and advanced machine learning techniques

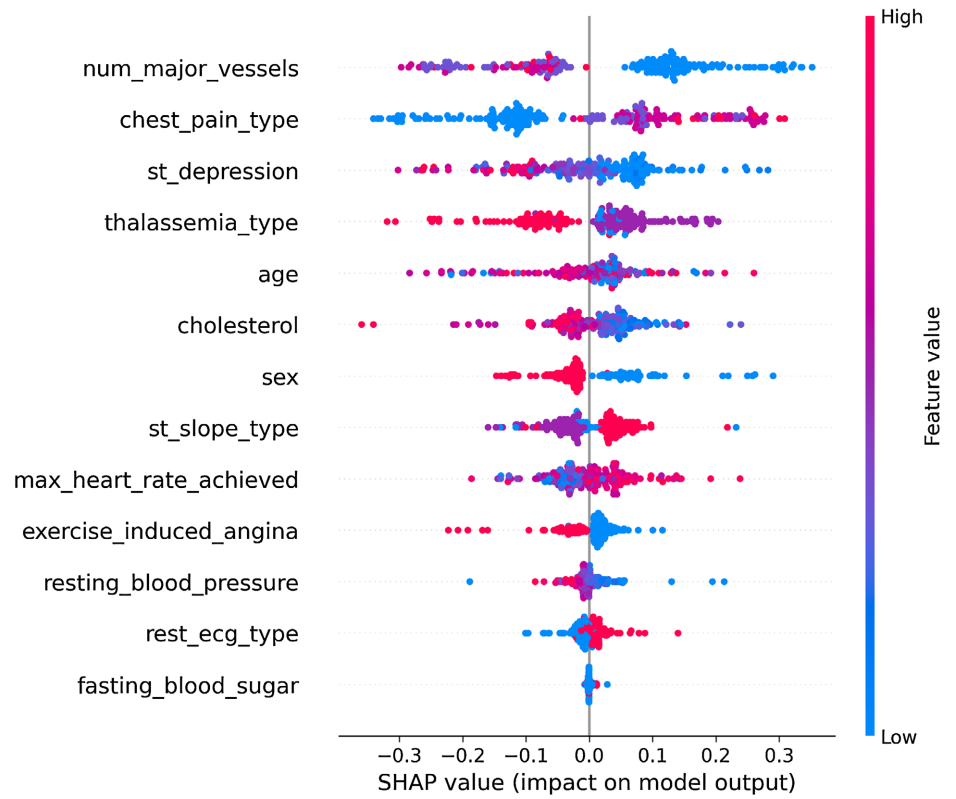


Figure 4. SHAP summary plot showing the global feature importance and the distribution of SHAP values for each feature across the dataset.

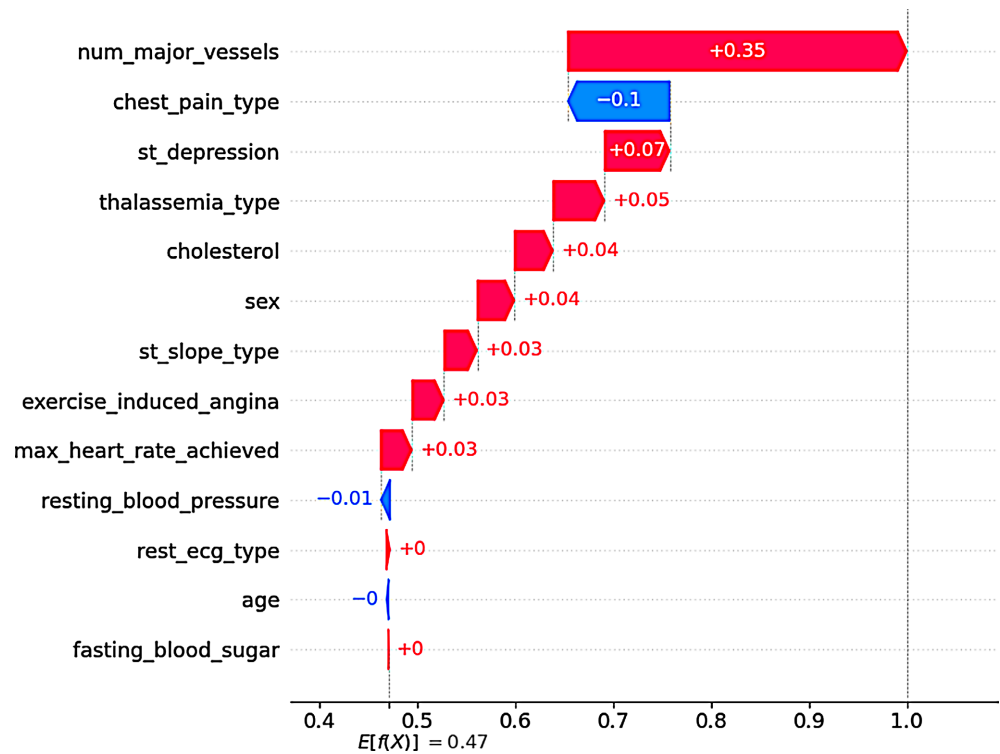


Figure 5. SHAP force plot visualizing the impact of each feature on a specific prediction. Features pushing the prediction higher appear in red, while those reducing it appear in blue.

established a robust predictive framework for heart disease detection. Evaluation metrics, combined with visual analysis, provided crucial insights into each model’s strengths and limitations, guiding future refinements and clinical applications. This data-driven approach reinforces the significance of machine learning in healthcare, particularly in improving diagnostic precision and early intervention strategies. **Figure 6** and **Figure 7** illustrate the SHAP-based interpretation of the XGBoost model. Specifically, **Figure 6** shows the mean absolute SHAP values across all features, identifying the most influential variables in the model’s decisions. **Figure 7** presents a force plot explaining a single prediction, highlighting how specific features push the model’s output toward or away from a certain classification.

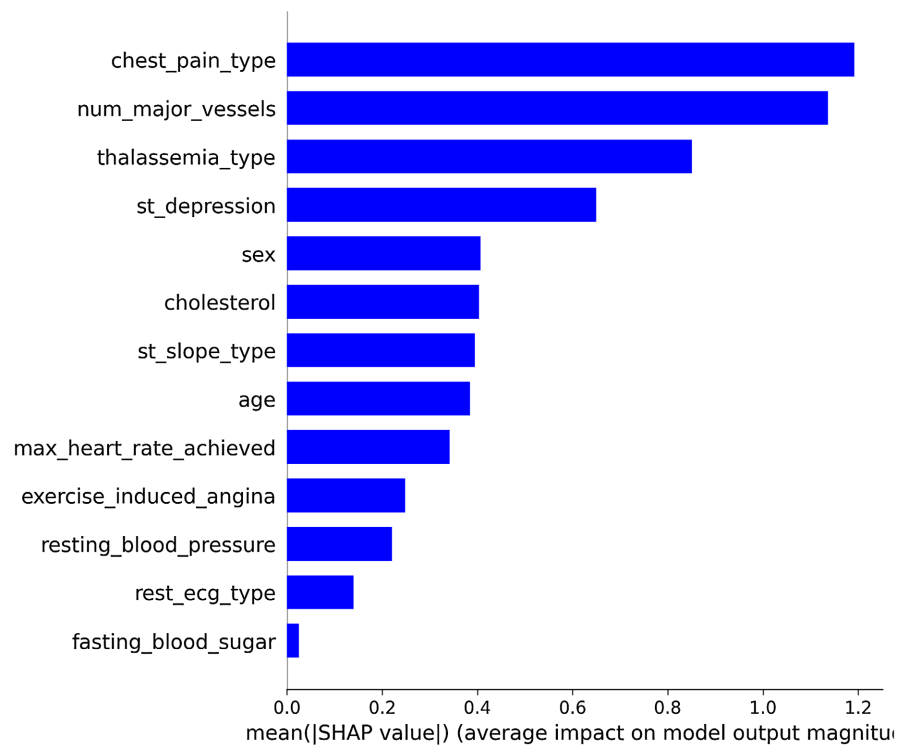


Figure 6. Mean absolute SHAP values: XGBoost on the data.

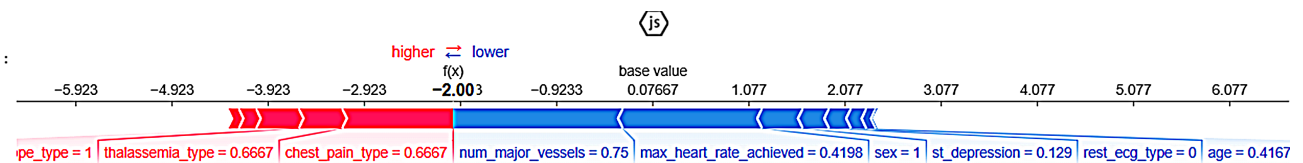


Figure 7. Mean absolute SHAP values.

5.1. Understanding LIME (Local Interpretable Model-Agnostic Explanation)

LIME, is a powerful method used to explain predictions made by complex machine learning models. Its primary goal is to provide local, understandable insights into why a model made a specific prediction. Rather than trying to explain

the model as a whole, LIME focuses on explaining individual predictions in a way that humans can interpret. The way LIME works is by creating small, local explanations around the specific data point in question. It does this by generating **perturbations** (or variations) of the instance being explained. Essentially, this involves modifying some of the data features while keeping others, and then passing these altered instances through the original “black-box” model. By observing how the model reacts to these changes, LIME is able to approximate the decision-making process of the model with a simpler, more interpretable model, such as a linear model or decision tree. What makes LIME particularly valuable is its ability to take any black-box model no matter how complex or opaque it is and to produce explanations that help to understand the reasoning behind the model’s predictions. This is similar to another popular technique, **SHAP**, which also aims to shed light on how machine learning models make their decisions. The difference is that LIME works locally, focusing on individual predictions, which provides a more granular level of insight. The method is not dependent on the inner workings of the model. Instead, it works by observing how the model behaves with different inputs and outputs. The key here is that it creates a **local approximation**, meaning it only looks at the region around the specific instance we are interested in. This makes the explanations more **relevant** and **understandable**, as it narrows the focus down to a specific scenario instead of trying to explain the entire model. LIME approach is driven by a formula that balances **accuracy** and **simplicity**. The goal is to build an interpretable model that closely mimics the black-box model’s behavior for the instance at hand, while avoiding overly complex explanations. Essentially, LIME tries to find a trade-off between accuracy (how well the simple model approximates the black-box model) and simplicity (how easy it is for humans to understand the explanation). The formula can be expressed as follows [26]:

$$\mathcal{E}_{LIME}(x) = \underset{g \in \mathcal{G}}{\operatorname{argmin}} (L(f, g, \pi(x)) + \Omega(g)) \quad (3)$$

In the formula:

- f is the original complex model.
- g is the simpler, interpretable model that we aim to build.
- $L(f, g, \pi(x))$ is the loss function that measures how closely the simple model g approximates the predictions of the original model f for the instance x .
- $\Omega(g)$ represents the complexity of the model g , which we want to minimize to ensure interpretability.
- $\pi(x)$ defines the neighborhood of the data point x and how perturbations are created around it.

One of the key strengths of LIME is its **model-agnostic** nature. This means that LIME can be used with **any type of model**, regardless of how complicated or how little we know about its internal workings. It doesn’t require access to the model’s internals, which makes it an incredibly flexible tool for explaining predictions across a wide range of machine learning algorithms.

In summary, LIME is a versatile and insightful tool for explaining the decisions made by machine learning models, especially those that are complex and difficult to interpret. By focusing on local explanations, LIME helps bridge the gap between complex model predictions and human understanding, ensuring that machine learning models are not only accurate but also transparent and trustworthy.

5.1.1. LIME Visualization and Interpretation

To enhance the transparency of our model predictions, we applied the LIME (Local Interpretable Model-Agnostic Explanations) technique, which helps explain individual predictions by approximating the model locally with a simpler, interpretable model. This makes it possible to understand which input features influenced a specific decision, even for complex models.

Each LIME visualization is composed of three main parts:

- The **left panel** shows the predicted probability for the data instance.
- The **middle section** lists the most important features that contributed to the prediction.
- The **right panel** displays the actual values of those input features.

In the **True Negative** case, we observe that only one feature, *ca*, slightly supports the positive class, while all other variables contribute to the negative prediction. This suggests that fewer features are involved in confirming negative outcomes.

On the other hand, the **True Positive** prediction involves a greater number of features (except *trestbps* and *lbs*) that actively support the positive class. This shows that multiple factors are contributing to the model's confidence in identifying a positive case.

By incorporating LIME, we gain valuable insight into the inner workings of our models, bridging the gap between predictive performance and interpretability.

5.1.2. Interpretation of LIME Explanations

The LIME visualizations generated for three test instances offer a clear and intuitive understanding of the model's decision-making process. In the first and second cases, both classified correctly as *Cardio* (*i.e.*, presence of heart disease), the model predicted high probabilities of 97.93% and 97.85%, respectively. Several features played a decisive role in these predictions. Notably, the absence of major vessels ($\text{num_major_vessels} \leq 0.00$), low levels of thalassemia, and specific cholesterol ranges had strong positive contributions. These were complemented by other supportive factors such as low ST depression and certain heart rate ranges. While some features like chest pain type slightly pulled the prediction away from a Cardio outcome, their impact was comparatively minor.

In the third instance, the model correctly predicted a *No Cardio* outcome (absence of heart disease) with a very low probability for Cardio (0.14%). Here, protective features such as moderate age and low ST depression negatively influenced the likelihood of a Cardio classification. Although some risk-related features, including high cholesterol and the presence of visible vessels, showed positive

contributions toward the Cardio class, the overall influence of protective indicators led the model to a confident and accurate prediction.

These explanations confirm that the model does not operate as a black box: it relies on clinically relevant features and their nuanced interactions. LIME serves as a powerful interpretability tool, offering transparency in how predictions are formed, which is especially valuable in sensitive domains like healthcare. An illustration of these LIME-based explanations for the three selected test instances is shown in **Figure 8**.

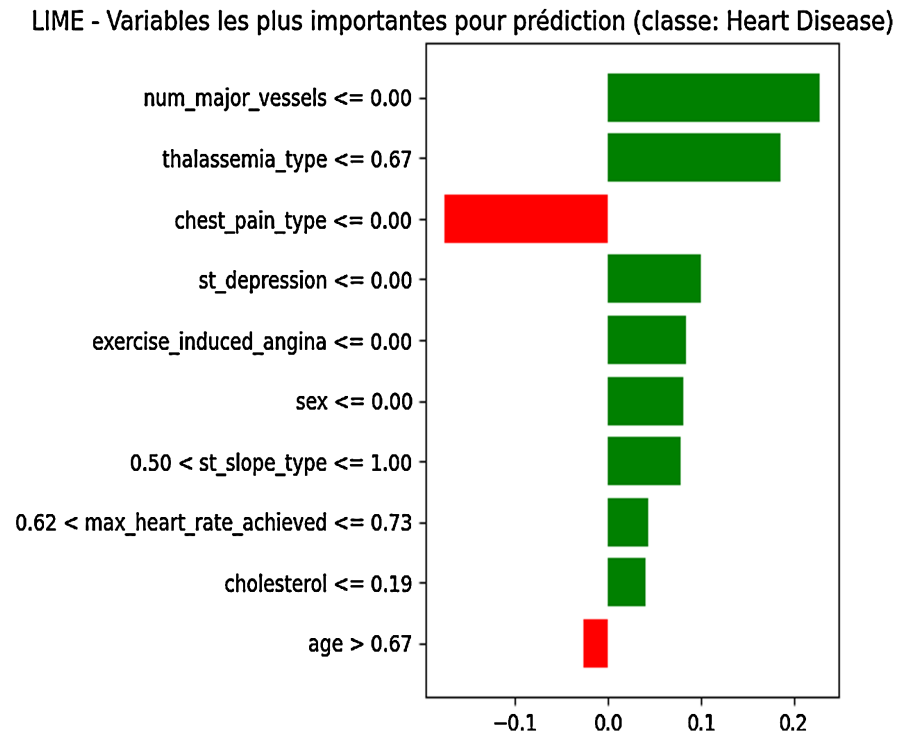


Figure 8. Comparison of feature importance methods.

5.2. Permutation Feature Importance

Permutation Feature Importance provides a practical and intuitive way to understand which input variables have the greatest impact on a model's performance. The approach works by randomly shuffling the values of each feature and observing how this affects the model's accuracy. If shuffling a feature causes a noticeable drop in performance, that feature is likely important.

In our study, we applied this method to the XGBoost classifier to evaluate the role of each feature in predicting cardiovascular disease. Among the top contributors were num_major_vessels, chest_pain_type, age, and thalassemia_type, all of which are consistent with known medical risk factors. These results validate the model's clinical relevance and also highlight how permutation importance can uncover globally impactful features, even those that may not be frequently used in internal model splits [7]. The ranking of feature importance obtained through this method is illustrated in **Figure 9**.

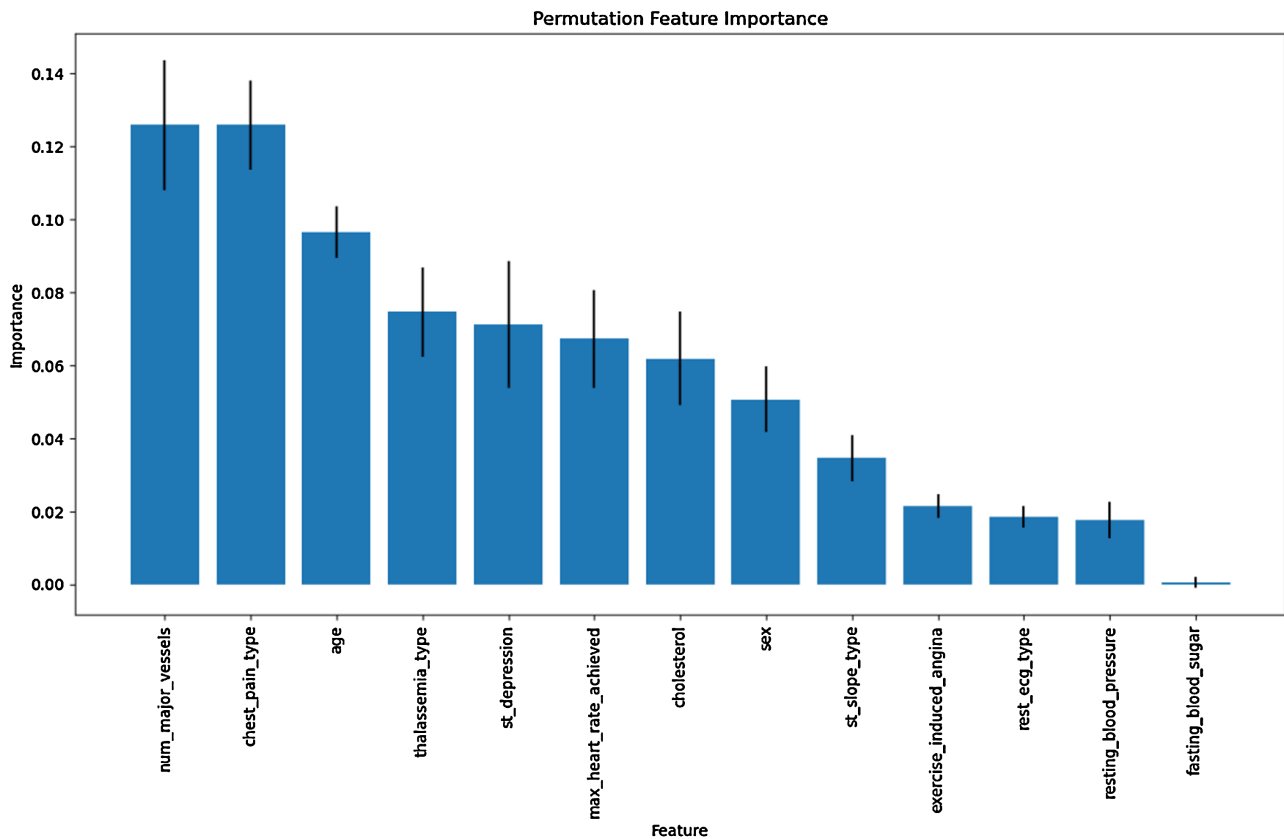


Figure 9. Permutation feature importance.

6. Feature Importance Comparison: XGBoost, Permutation, SHAP, and LIME

To better understand and trust the predictions made by our heart disease classification model, we incorporated four complementary explainability methods: XGBoost's built-in feature importance, Permutation Feature Importance, SHAP (SHapley Additive Explanations), and LIME (Local Interpretable Model-Agnostic Explanations). Each method offers a unique lens to examine the model's behavior ranging from global trends to local, instance-specific insights.

Explanation Techniques Overview

XGBoost Feature Importance reflects how frequently and effectively each feature is used to split data in the model's trees, providing a high-level view of what drives decision-making across the dataset.

Permutation Importance evaluates the importance of features by measuring how much performance drops when their values are randomly shuffled. Unlike XGBoost's internal metric, this method is model-agnostic and directly linked to predictive accuracy.

SHAP Values take a game-theoretical approach by calculating how much each feature contributes to an individual prediction, accounting for all possible feature interactions. SHAP is especially valuable for understanding both global and local

behavior. **LIME** focuses on specific predictions. It builds an interpretable model around a single instance by perturbing the data slightly and observing the effect on the prediction. LIME excels at explaining decisions in a way that's accessible, especially when dealing with binary or engineered features.

7. Results and Interpretation

Table 3 summarizes the top features' importance scores assigned by SHAP, LIME, and Permutation methods. Several patterns emerge across methods. Most notably, `num_major_vessels`, `thalassemia_type`, and `chest_pain_type` are consistently highlighted as influential features. This agreement across techniques adds robustness to our interpretation of these features as critical risk factors. Some interesting differences also appear. For example, `sex` and `exercise_induced_angina` are given greater weight in SHAP and LIME than in permutation importance or XGBoost. This suggests that while they may not be dominant at the global level, they significantly influence specific predictions a nuance that instance-level explanations can reveal.

Together, these methods provide a more complete and transparent view of the model's reasoning, which is particularly important in high-stakes domains like healthcare. By combining global and local perspectives, we ensure that the model's predictions are not only accurate but also explainable and trustworthy.

Table 3. Feature importance comparison: SHAP, LIME, and permutation.

Feature	SHAP Value	LIME Value	Permutation Importance
<code>num_major_vessels</code>	0.3269	0.2283	0.1259
<code>thalassemia_type</code>	0.0542	0.1862	0.0746
<code>chest_pain_type</code>	-0.1017	-0.1762	0.1259
<code>st_depression</code>	0.0881	0.1007	0.0712
<code>exercise_induced_angina</code>	0.0292	0.0843	0.0215
<code>sex</code>	0.0342	0.0817	0.0507
<code>st_slope_type</code>	0.0275	0.0787	0.0346
<code>max_heart_rate_achieved</code>	0.0344	0.0433	0.0673
<code>cholesterol</code>	0.0369	0.0412	0.0620
<code>age</code>	0.0022	-0.0260	0.0966

Comparison of Importance Methods

Each method brings unique strengths. XGBoost's [27]-[29] internal importance helps us understand the model's construction, highlighting the features it structurally depends on. Permutation importance reveals which features most affect overall accuracy, even if they aren't structurally central in the model. SHAP provides a detailed, consistent explanation framework, quantifying how each feature

pushes predictions toward higher or lower risk. Finally, LIME offers clear, instance-specific justifications especially valuable in clinical settings where understanding individual cases is crucial. **Feature Importance Comparison: SHAP, LIME, and Permutation**

By combining these tools, we gain a holistic view of model behavior. This multi-angle approach is especially important in healthcare, where transparency, fairness, and reliability are essential for deploying machine learning responsibly. A comparative overview of the interpretability outputs produced by SHAP, LIME, and Permutation Importance is presented in **Figure 10**, highlighting how each method provides complementary insights into model decisions.

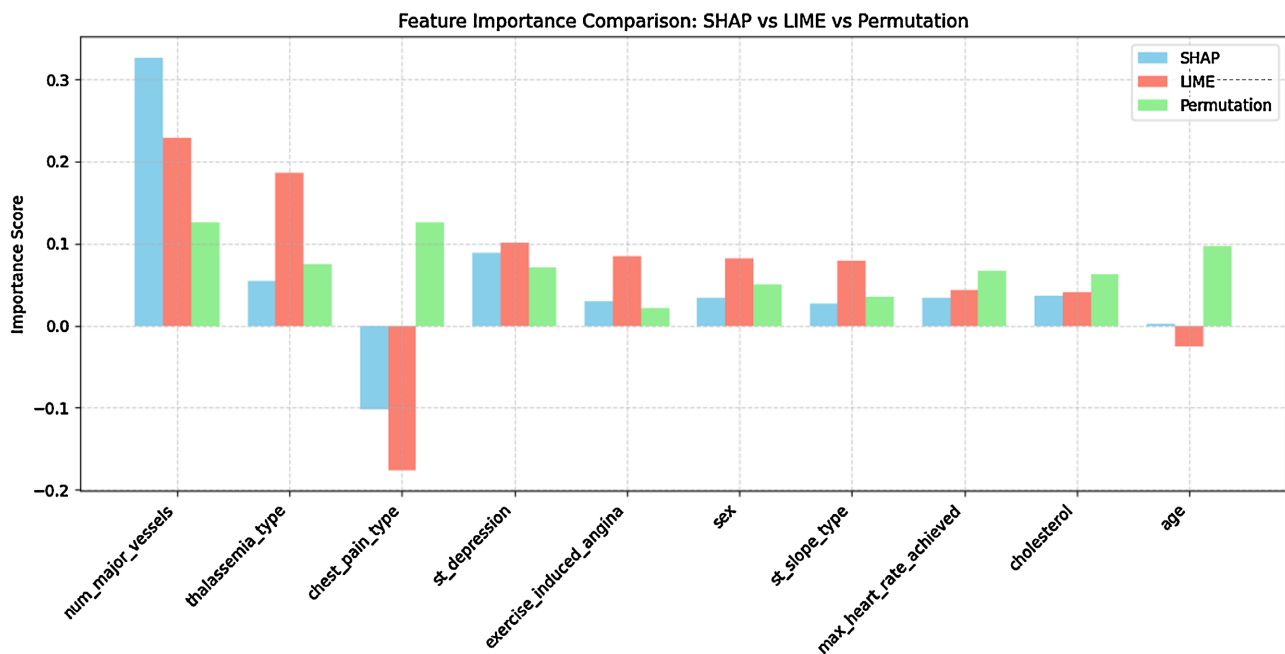


Figure 10. Comparison: SHAP, LIME, and permutation.

8. Conclusions

In this work, we investigated how explainable artificial intelligence (XAI) techniques can make powerful yet complex machine learning models more transparent and trustworthy in healthcare applications. While models like XGBoost offer excellent predictive performance, their inner workings often remain a black box, raising challenges for clinical adoption, where trust and accountability are critical. By comparing multiple explanation methods XGBoost importance, permutation importance, SHAP values, and LIME we uncovered how each technique brings its own perspective. Some, like permutation importance and SHAP, revealed influential features that XGBoost's structure did not fully highlight, while LIME provided interpretable decision rules that can be more easily understood by clinicians.

Each method has its own strengths: XGBoost shows how the model is built, permutation identifies features with real impact, SHAP gives consistent and theory-backed insights, and LIME delivers user-friendly explanations. Taken

together, these complementary tools offer a more complete understanding of model behavior. Ultimately, integrating multiple XAI methods helps bridge the gap between predictive accuracy and clinical transparency, ensuring that machine learning models not only perform well but are also interpretable, reliable, and ready for real-world healthcare use.

Author Contributions

N.Z.D. conceptualized the study, developed the methodology, and wrote the main manuscript text. H.B. contributed to data analysis and validation. A.S.N. and N.D. supervised the project and reviewed the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Sadeghi, Z., Alizadehsani, R., Cifci, M.A., Kausar, S., Rehman, R., Mahanta, P., *et al.* (2023) A Review of Explainable Artificial Intelligence in Healthcare. *Computer Methods and Programs in Biomedicine*, **226**, Article ID: 107194.
- [2] Swathy, M. and Saruladha, K. (2022) A Comparative Study of Classification and Prediction of Cardio-Vascular Diseases (CVD) Using Machine Learning and Deep Learning Techniques. *ICT Express*, **8**, 109-116.
<https://doi.org/10.1016/j.icte.2021.08.021>
- [3] Ahmad, I., Yao, C., Li, L., Chen, Y., Liu, Z., Ullah, I., *et al.* (2024) An Efficient Feature Selection and Explainable Classification Method for EEG-Based Epileptic Seizure Detection. *Journal of Information Security and Applications*, **80**, Article ID: 103654.
<https://doi.org/10.1016/j.jisa.2023.103654>
- [4] Amini, M., Bagheri, A., Piri, S. and Delen, D. (2024) A Hybrid AI Framework to Address the Issue of Frequent Missing Values with Application in EHR Systems: The Case of Parkinson's Disease. *Proceedings of the Annual Hawaii International Conference on System Sciences*, Hawaii, 3-6 January 2024, 2.
<https://doi.org/10.24251/hicss.2024.126>
- [5] Barr Kumarakulasinghe, N., Blomberg, T., Liu, J., Saraiva Leao, A. and Papapetrou, P. (2020) Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models. 2020 *IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, 28-30 July 2020, 7-12.
<https://doi.org/10.1109/cbms49503.2020.00009>
- [6] Peng, J., Zou, K., Zhou, M., Teng, Y., Zhu, X., Zhang, F., *et al.* (2021) An Explainable Artificial Intelligence Framework for the Deterioration Risk Prediction of Hepatitis Patients. *Journal of Medical Systems*, **45**, Article No. 61.
<https://doi.org/10.1007/s10916-021-01736-5>
- [7] Sabol, P., Sinčák, P., Hartono, P., Kočan, P., Benetinová, Z., Blichárová, A., *et al.* (2020) Explainable Classifier for Improving the Accountability in Decision-Making for Colorectal Cancer Diagnosis from Histopathological Images. *Journal of Biomedical Informatics*, **109**, Article ID: 103523. <https://doi.org/10.1016/j.jbi.2020.103523>
- [8] Kletz, S., Schoeffmann, K. and Husslein, H. (2019) Learning the Representation of

- Instrument Images in Laparoscopy Videos. *Healthcare Technology Letters*, **6**, 197-203. <https://doi.org/10.1049/htl.2019.0077>
- [9] Porumb, M., Stranges, S., Pescapè, A. and Pecchia, L. (2020) Precision Medicine and Artificial Intelligence: A Pilot Study on Deep Learning for Hypoglycemic Events Detection Based on ECG. *Scientific Reports*, **10**, Article No. 170. <https://doi.org/10.1038/s41598-019-56927-5>
- [10] Izadyazdanabadi, M., Belykh, E., Cavallo, C., Zhao, X., Gandhi, S., Moreira, L.B., et al. (2018) Weakly-Supervised Learning-Based Feature Localization for Confocal Laser Endomicroscopy Glioma Images. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C. and Fichtinger, G., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, Springer, 300-308. https://doi.org/10.1007/978-3-030-00934-2_34
- [11] Deperlioglu, O., Kose, U., Gupta, D., Khanna, A., Giampaolo, F. and Fortino, G. (2022) Explainable Framework for Glaucoma Diagnosis by Image Processing and Convolutional Neural Network Synergy: Analysis with Doctor Evaluation. *Future Generation Computer Systems*, **129**, 152-169. <https://doi.org/10.1016/j.future.2021.11.018>
- [12] Colin, J. and Surantha, N. (2025) Interpretable Deep Learning for Pneumonia Detection Using Chest X-Ray Images. *Information*, **16**, Article 53. <https://doi.org/10.3390/info16010053>
- [13] Wani, N.A., Kumar, R. and Bedi, J. (2024) DeepXplainer: An Interpretable Deep Learning Based Approach for Lung Cancer Detection Using Explainable Artificial Intelligence. *Computer Methods and Programs in Biomedicine*, **243**, Article ID: 107879. <https://doi.org/10.1016/j.cmpb.2023.107879>
- [14] Muddamsetty, S.M., Jahromi, M.N.S. and Moeslund, T.B. (2021) Expert Level Evaluations for Explainable AI (XAI) Methods in the Medical Domain. In: Del Bimbo, A., et al., Eds., *Pattern Recognition. ICPR International Workshops and Challenges*, Springer, 35-46. https://doi.org/10.1007/978-3-030-68796-0_3
- [15] Biswas, M., Kaiser, M.S., Mahmud, M., Al Mamun, S., Hossain, M.S. and Rahman, M.A. (2021) An XAI Based Autism Detection: The Context behind the Detection. In: Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q. and Zhong, N., Eds., *Brain Informatics*, Springer, 448-459. https://doi.org/10.1007/978-3-030-86993-9_40
- [16] Alamatsaz, N., Tabatabaei, L., Yazdchi, M., Payan, H., Alamatsaz, N. and Nasimi, F. (2024) A Lightweight Hybrid CNN-LSTM Explainable Model for ECG-Based Arrhythmia Detection. *Biomedical Signal Processing and Control*, **90**, Article ID: 105884. <https://doi.org/10.1016/j.bspc.2023.105884>
- [17] Mahmoudi, M.R., Akbarzadeh, H., Parvin, H., Nejatian, S., Rezaie, V. and Alinejad-Rokny, H. (2020) Consensus Function Based on Cluster-Wise Two Level Clustering. *Artificial Intelligence Review*, **54**, 639-665. <https://doi.org/10.1007/s10462-020-09862-1>
- [18] Suh, J., Yoo, S., Park, J., Cho, S.Y., Cho, M.C., Son, H., et al. (2020) Development and Validation of an Explainable Artificial Intelligence-Based Decision-Supporting Tool for Prostate Biopsy. *BJU International*, **126**, 694-703. <https://doi.org/10.1111/bju.15122>
- [19] Yilmaz, R. and Yağın, F.H. (2022) Early Detection of Coronary Heart Disease Based on Machine Learning Methods. *Medical Records*, **4**, 1-6. <https://doi.org/10.37990/medr.1011924>
- [20] Sigut, J., Fumero, F., Estévez, J., Alayón, S. and Díaz-Alemán, T. (2023) In-Depth Evaluation of Saliency Maps for Interpreting Convolutional Neural Network Decisions

- in the Diagnosis of Glaucoma Based on Fundus Imaging. *Sensors*, **24**, Article 239. <https://doi.org/10.3390/s24010239>
- [21] Sengar, P.S. (2023) Heart Attack Prediction Dataset. Kaggle.
- [22] Joly, N.A. and Arif, A.S.M. (2024) Permutation Feature Importance-Based Cardiovascular Disease (CVD) Prediction Using Ann. In: Arefin, M.S., Kaiser, M.S., Bhuiyan, T., Dey, N. and Mahmud, M., Eds., *Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning*, Springer, 1039-1053. https://doi.org/10.1007/978-981-99-8937-9_69
- [23] Han, J., Kamber, M. and Pei, J. (2011) Data Mining: Concepts and Techniques. 3rd Edition, Morgan Kaufmann.
- [24] Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A.L., Manikandan, R. and Gandomi, A.H. (2024) Classification Models Combined with Boruta Feature Selection for Heart Disease Prediction. *Informatics in Medicine Unlocked*, **44**, Article ID: 101442. <https://doi.org/10.1016/j.imu.2023.101442>
- [25] Lundberg, S.M. and Lee, S.I. (2017) A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 4768-4777.
- [26] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 1135-1144. <https://doi.org/10.1145/2939672.2939778>
- [27] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [28] Li, J.P., Haq, A.U., Din, S.U., Khan, J., Khan, A. and Saboor, A. (2020) Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access*, **8**, 107562-107582. <https://doi.org/10.1109/access.2020.3001149>
- [29] Ahsan, M.M., Luna, S.A. and Siddique, Z. (2022) Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. *Healthcare*, **10**, Article 541.