Journal of Informatics Engineering (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

Optimizing Heart Disease Classification Using C4.5, Random Forest, and XGBoost with ANOVA, Chi-Square, and AdaBoost

Andika Pratama*[1], Setiawan Assegaff[2], Jasmir Jasmir[3], Nurhadi Nurhadi[4]

[1,2,3,4]Master of Information Systems, Dinamika University, Indonesian

Email: [1]andikajtn@gmail.com

Received : Nov 10, 2025; Revised : Nov 25, 2025; Accepted : Nov 25, 2025; Published : Apr 15, 2026

Abstract

Heart disease remains one of the leading causes of mortality worldwide, underscoring the need for accurate and scalable prediction models within clinical informatics. This study proposes a leakage-safe machine learning pipeline combining stratified splitting, SMOTE-based imbalance handling, and in-fold feature selection using ANOVA, Chi-Square, and AdaBoost-assisted ranking to enhance classification performance on a large heart-disease dataset consisting of 10,000 samples and 21 attributes. Three widely used algorithms, C4.5, Random Forest, and XGBoost, were evaluated to determine the optimal model-feature selection configuration for structured medical data. The results demonstrate that feature relevance contributes more significantly to predictive performance than increasing model complexity, with Random Forest achieving the highest accuracy, precision, recall, and F1-Score at 98.43% when combined with Chi-Square or ANOVA feature selection. C4.5 showed the greatest relative improvement, rising from 76.52% to 97.57% using AdaBoost-assisted selection, while XGBoost improved from 66.32% to 94.88% after statistical filtering. The dominant features identified such as CRP, BMI, blood pressure, fasting glucose, LDL, triglycerides, and homocysteine align with well-established cardiovascular biomarkers, supporting clinical validity. This research provides an important contribution to computer science by demonstrating an efficient and scalable hybrid FS-boosting framework capable of reducing unnecessary model complexity, improving generalization, and supporting low-latency deployment in clinical decision-support systems. The findings highlight the potential of structured-data machine learning to strengthen digital health diagnostics in resource-limited environments.

Keywords : *AdaBoost, ANOVA, Chi-Square, Feature Selection, Heart Disease, SMOTE*

1. INTRODUCTION

Heart disease remains a major contributor to global morbidity and mortality [1]. Its prevalence is closely associated with metabolic and lifestyle risk factors such as hypertension, diabetes, dyslipidemia, obesity, and smoking, which continue to increase across populations [2]. The human heart consists of four chambers equipped with valves that maintain unidirectional blood flow [3], and disturbances in these structures may lead to myocardial ischemia, necrosis, or other cardiovascular complications [4]. Risk factors are categorized into modifiable factors including hypertension, dyslipidemia, obesity, diabetes, alcohol consumption, and physical inactivity [5] and nonmodifiable factors such as age, gender, and heredity [6].

Globally, heart disease remains the foremost contributor to mortality [7]. The World Health Organization reports that only 152 out of 194 member states provide complete datasets for cardiovascular monitoring, highlighting substantial global data gaps [8]. In Indonesia, heart disease affects approximately 1.5 percent of the population across all age groups [9]. Clinically, heart disease is identified through symptoms such as chest pain, abnormal ECG patterns, irregular blood pressure, and changes in heart rate, glucose, and lipid biomarkers [10]. Myocardial infarction, one of the most critical

manifestations, contributes to a mortality rate of nearly 12.9 percent of all cardiovascular deaths [11], underscoring the need for early detection and precise risk stratification [12].

Conventional diagnostic procedures such as physical examinations, laboratory tests, ECG, echocardiography, and treadmill evaluations remain the standard for detecting heart disease [13]. Yet these methods can be time consuming, costly, and dependent on clinical expertise [14]. Advances in artificial intelligence and machine learning now provide complementary tools that enable faster data analysis, identification of hidden clinical patterns, and more consistent cardiovascular risk assessment [15], [16], [17].

Tree-based machine learning algorithms have been widely used for medical classification because they can effectively handle numerical and categorical attributes while providing interpretable decision logic [18]. C4.5 generates rule-based decision trees using information gain and pruning, enabling intuitive clinical interpretation [19]. Random Forest reduces variance and increases robustness by aggregating multiple decorrelated trees through bagging [20]. XGBoost, a gradient-boosting framework, incorporates regularization, shrinkage, and parallelized tree construction to enhance predictive accuracy and computational efficiency [21]. These three models are selected because they represent distinct strengths interpretability in C4.5, robustness in Random Forest, and high accuracy through boosting in XGBoost which makes them ideal candidates for systematic comparison in clinical prediction tasks.

Prior research has demonstrated the effectiveness of tree-based approaches for heart disease classification. Pal et al. reported 86.9% accuracy and 90.6% recall using Random Forest [22]. Dissanayake et al. achieved 88.52% accuracy using a decision tree with backward feature selection [23]. Budholiya et al. optimized XGBoost using Bayesian optimization and obtained 91.8% accuracy [24]. Additional studies further reinforce the potential of tree-based learners. Abdallah et al. achieved 89.30% accuracy with Random Forest [25], Roopa et al. reported 87.50% accuracy with XGBoost [26], while Shahid et al. reached 92.20% accuracy using ensemble boosting models [27]. These comparisons show that although individual models consistently achieve high accuracy, their performance depends heavily on preprocessing and feature selection strategies.

Feature selection plays a crucial role in reducing dimensionality, improving generalization, and enhancing model interpretability. Techniques such as ANOVA, Chi-Square, and ensemble-based feature selection including AdaBoost have been widely applied to clinical data. Empirical studies show that the choice and placement of FS can substantially influence confusion-matrix metrics [28]. Alkhodari *et al.* combined ANOVA and Chi-Square screening with tree-based learning (RUSBoost), achieving 97.08% accuracy, 81.25% precision, and 86.67% F1-score [29]. Another study on phonocardiogram (PCG) signals also applied Chi-Square and ANOVA before training SVM, RF, and AdaBoost, reporting near-perfect accuracies in several splits [30]. In addition, handling class imbalance through techniques such as Synthetic Minority Over-Sampling Technique (SMOTE) has been proven to reduce false negatives and raise recall and F1-score [31].

To address these gaps, this study utilizes a Kaggle heart disease dataset 10.000 instances and 21 features clinical and demographic attributes, including age, blood pressure, cholesterol level, and key symptoms. Given the mixed numerical and categorical nature of the data, a standardized preprocessing pipeline involving cleaning, encoding, and imbalance handling is applied to ensure reliable model evaluation. The novelty of this study lies in integrating three feature selection methods ANOVA, Chi-Square, and AdaBoost with three tree-based classifiers, namely C4.5, Random Forest, and XGBoost, to enable a systematic cross-model and cross-feature selection comparison.

The purpose of this research is to evaluate the performance of these model-feature selection combinations using accuracy, precision, recall, F1-score, and confusion matrix analysis, with the goal of identifying the most effective configuration. This approach provides a data-driven foundation for

developing artificial intelligence based diagnostic systems that are accurate, interpretable, and computationally efficient for early heart disease detection.

2. METHOD

This section describes the dataset, preprocessing, class balancing, feature selection, learning algorithms, and evaluation protocol. To prevent data leakage, all label-dependent operations such as SMOTE, feature selection, and hyper-parameter tuning are applied only to the training folds inside cross-validation, the held-out test set is used once for final assessment. The overall workflow is illustrated in Figure 1.

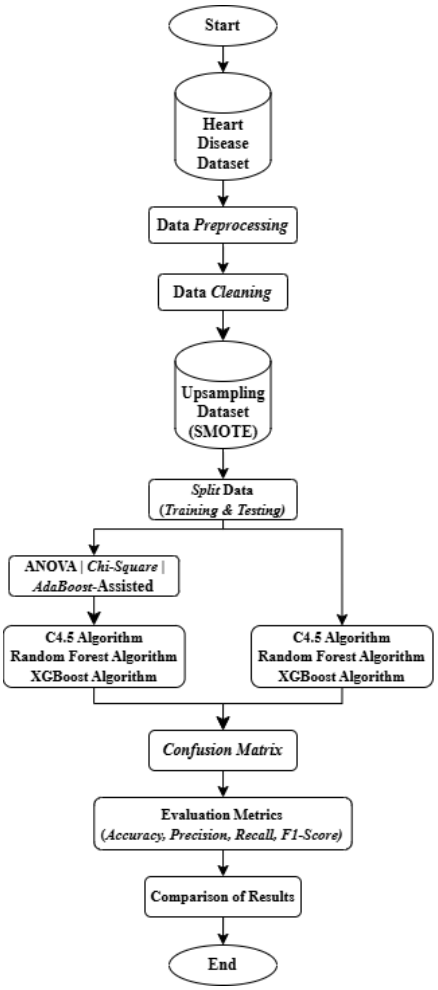

Figure 1. Research Stages

2.1. Dataset

The study uses the public Heart Disease dataset from Kaggle (curated by Oktay Ördekçi), containing 10,000 records and 21 attributes (20 predictors plus 1 target).

Table 1. Information of Dataset

Name	Owner	Link
Heart Disease	Oktay Ördekçi	Kaggle Link

Based on the information presented in Table 1, the characteristics of the dataset can be summarized as follows:

1. The dataset used in this study is obtained from a publicly available and anonymized source (Kaggle Link), which ensures compliance with ethical standards for secondary data usage. The dataset contains no personal identifiers, sensitive information, or traceable metadata, and is shared under Kaggle's data usage policy. All data processing is conducted solely for academic research purposes, without redistribution or attempts at re-identification. This ensures adherence to principles of privacy, fairness, and responsible data handling.
2. The dataset consists of 10,000 rows and 21 attributes, comprising 20 predictor variables and one target variable, providing sufficient statistical power for machine-learning-based classification.
3. The target variable includes two classes (Yes and No), with a distribution of 8,000 No (80%) and 2,000 Yes (20%). This reflects a moderate 4:1 imbalance that may influence model performance, particularly on metrics such as recall and F1-score.
4. Each feature exhibits fewer than 0.5 percent missing entries (e.g., 29 missing values for Age and 30 for Cholesterol Level), indicating high data quality that requires only simple imputation rather than extensive cleaning.

2.2. Data Preprocessing

Data preprocessing is an initial process that includes cleaning, transforming, and feature selection to ensure better data quality in order to improve the accuracy and performance of machine learning models [32].

2.2.1 Data Cleaning

Data cleaning is the process of handling raw data by removing duplicates, correcting inconsistencies, and resolving empty or invalid values, so that the quality of the data is more guaranteed and suitable for use in the analysis and modeling stages of machine learning [33]. In this study, numerical attributes such as Age, BMI, and laboratory features are imputed using the median, which provides robustness against skewness and outliers. Binary categorical variables are imputed using the mode derived from the training subset. For ordinal features (such as Low, Medium, High or None, Low, Medium, High), missing values below five percent are imputed using the mode, while features with missingness of twenty percent or more are assigned an explicit Unknown level to avoid distorting the original distribution. This additional level is applied only to the affected attributes.

2.2.2 Encoding

Categorical encoding is an important step in data preprocessing, as most machine learning models work better with numerical data [34]. In this study, several data encoding methods were used, such as:

1. Binary mapping for Yes or No attributes, as well as for gender (Male as 1 and Female as 0).
2. Ordinal attributes are encoded using order-preserving mappings, including Exercise Habits (Low, Medium, High mapped to 0, 1, 2), Stress Level (0, 1, 2), Sugar Consumption (0, 1, 2), and Alcohol Consumption (None, Low, Medium, High mapped to 0, 1, 2, 3).
3. When an Unknown category is introduced during cleaning, it is encoded as −1 to distinguish it from the established ordinal progression.

2.2.3 Outlier Handling

Outlier detection is conducted using the three-IQR criterion, where candidate outliers fall outside the interval defined by the first and third quartiles plus or minus three times the interquartile range. Potential outliers are screened with the 3×IQR Values outside:

$$[Q_1 - 3 \times IQR, Q_3 + 3 \times IQR] \quad (1)$$

Where $Q_1$and $Q_3$are the 25th and 75th percentiles and IQR $= Q_3 - Q_1$. Values flagged by this rule are examined individually. Entries deemed physiologically implausible are removed to preserve data validity, whereas extreme but plausible clinical measurements are retained to avoid suppressing meaningful variability in the dataset.

2.3 Class Balancing (SMOTE)

SMOTE (Synthetic Minority Oversampling Technique) is used in conjunction with various classification algorithms to address imbalanced datasets [35]. SMOTE is applied to address the moderate class imbalance (approximately 20 percent positive class). The oversampling procedure is performed only on the training folds, while the test set remains untouched to prevent data leakage. Synthetic minority samples are generated through linear interpolation between a minority instance and one of its nearest neighbors, using five neighbors by default. The interpolation formula is defined as:

$$x\sim = x + \delta(xNN - x), \delta \sim U(0,1)\ where\ k\_neighbors\ =\ 5 \qquad (2)$$

Where x is a minority-class feature vector, x_{NN} is one of its k-nearest neighbors, δ is a random scalar from a uniform distribution, and $\tilde{x}$ is the resulting synthetic instance. This mechanism enriches the minority class without duplicating data, thereby improving classifier balance.

2.4 Feature Selection

Feature selection is a dimension reduction technique used to select features that are relevant to the machine learning task [36]. Feature selection plays an important role in reducing data dimensions and improving the performance of the proposed framework [37]. The feature selection process is able to simplify the complexity of the data by removing insignificant attributes, so that the model is more efficient and still accurate[38]. The complete feature selection workflow used in this study is illustrated in Figure 2.

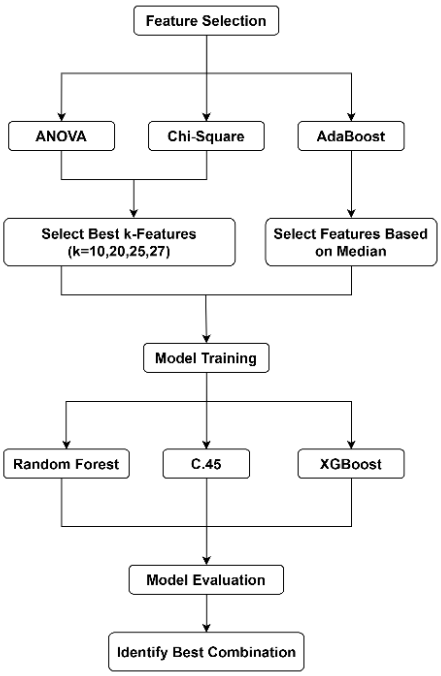

Figure 2. Feature Selection Process

For ANOVA and Chi-Square, five configurations of feature subsets are evaluated using k = {10, 20, 25, 27}, allowing the model to be tested under different levels of dimensionality. Each k-

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

feature subset is generated by ranking features based on their F-statistic or Chi-Square score, after which the top-k features are selected. These subsets are used to train C4.5, Random Forest, and XGBoost, and the best-performing k is retained for comparison.

In the embedded approach, AdaBoost is trained on the resampled dataset to obtain model-driven feature importance. Feature selection is conducted using *SelectFromModel* with a median importance threshold, resulting in a reduced feature set whose size varies depending on the learned importance distribution. This multi-step feature selection pipeline ensures comprehensive dimensionality reduction, integrating statistical relevance from ANOVA/Chi-Square and model-based importance from AdaBoost. The complete results of these three methods are summarized through the feature importance visualizations in Figure 3–5.

2.4.1 Analysis of Variance (ANOVA)

ANOVA is a statistical method used to measure differences in mean values between groups. In classification, ANOVA plays a role in selecting features that are most relevant to the target variable so as to improve model performance [39]. The F-statistic is computed as:

$$F = \frac{variance\ between\ groups}{variance\ within\ groups} \tag{3}$$

$$variance\ between\ groups = \frac{\sum_i^n ni\ (Yi-Y)^2}{(k-1)} \tag{4}$$

$$variance\ within\ groups = \frac{\sum_{i=1}^{k}\sum_{j=1}^{ni}(Yij-Yi)^2}{(n-k)} \tag{5}$$

Where n_i is the sample size in group i, Y_i is the mean of group i, n is the total sample count, and k is the number of groups. Figure 3 presents the ANOVA F-score ranking generated from the resampled dataset.

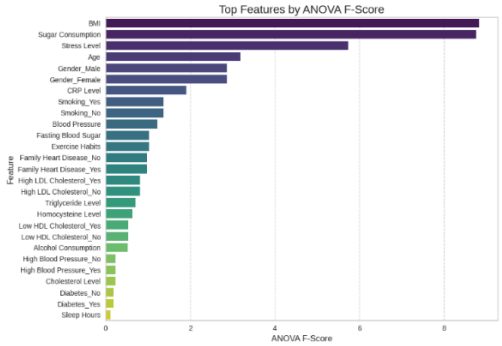

Figure 3. Top Features ANOVA

The ANOVA results show that BMI, Sugar Consumption, and Stress Level consistently emerge as the strongest discriminators. This is physiologically plausible, as metabolic load, dietary sugar intake, and chronic stress are well-established modulators of cardiovascular function. Lifestyle and demographic markers including Age, Gender, Smoking Status, and CRP Level follow closely, indicating that both biological and behavioral factors jointly explain the variation between risk classes. Continuous biomarkers such as Blood Pressure and Fasting Blood Sugar appear mid-ranked, reflecting their moderate but stable independent contribution.

2.4.2 Chi-Square

Chi-Square is a statistical technique that evaluates the association between features and target variables by comparing actual frequency and expected frequency. A large Chi-Square value indicates

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

high feature relevance, so this method is useful in removing unimportant attributes to improve model performance [40]. The expected value is computed as:

$$Ea = \frac{(a+b)(a+c)}{t} \tag{6}$$

The Chi-Square score is given by:

$$x^2 = \sum_{i=1}^{n} \frac{(Oi-Ei)^2}{Ei} \tag{7}$$

or in a 2×2 contingency table:

$$x^2 = \frac{(a-Ea)^2}{Ea} + \frac{(b-Eb)^2}{Eb} + \frac{(c-Ec)^2}{Ec} + \frac{(d-Ed)^2}{Ed} \tag{8}$$

where O_i and E_i denote observed and expected counts, and a, b, c, d represent cell frequencies. Figure 4 displays the Chi-Square ranking obtained from the dataset.

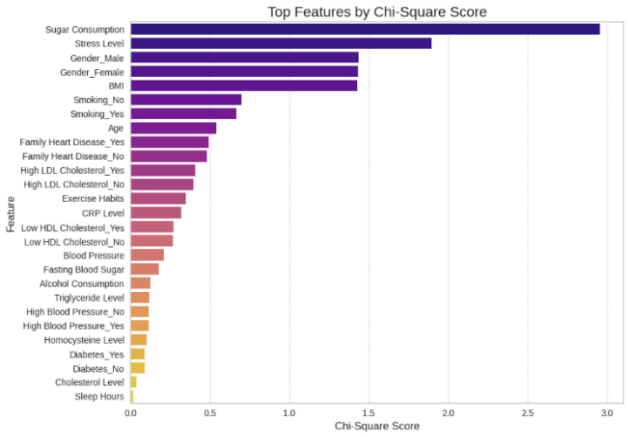

Figure 4. Top Features Chi Square

Chi-Square strongly emphasizes categorical lifestyle and demographic indicators. Sugar Consumption and Stress Level again appear at the top, confirming their strong statistical association with the outcome class. Gender-related features (Gender_Female, Gender_Male) and smoking categories receive high scores, reflecting their direct dependence on the class label. In contrast, continuous biomarkers such as Triglyceride Level or Cholesterol Level rank lower due to Chi-Square's limitation in capturing continuous variation making its results complementary to ANOVA.

2.4.3 AdaBoost

The AdaBoost-assisted feature selection is an embedded method that uses the internal weighting mechanism of AdaBoost to evaluate feature importance [41]. AdaBoost is employed as an embedded feature-selection technique by leveraging the model's internal feature-importance scores. An *AdaBoostClassifier* is first trained on the resampled training set; its aggregated feature importances are then processed using scikit-learn's *SelectFromModel* with a median threshold to retain only features with above-median importance. The reduced feature set is subsequently used to train the final classifier (Random Forest, XGBoost, or C4.5 depending on the experiment). This approach enables AdaBoost to highlight influential predictors by increasing emphasis on misclassified samples across boosting rounds. The weight defined as:

$$\varepsilon_t = \frac{\sum_i w_i^{(t)} I(h_t(x_i) \neq y_i)}{\sum_i w_i^{(t)}}, \alpha t = \frac{1}{2} ln(\frac{1-\varepsilon t}{\varepsilon t}),$$

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

$$w_i^{(t+1)} = w_i^{(t)} \exp(-\alpha_t y_i h_t(x_i)). \quad (9)$$

To operationalize this mechanism, an AdaBoostClassifier is trained on the resampled dataset, and the resulting aggregated importance scores are further processed using a SelectFromModel thresholding step to retain only features with above-median relevance. The complete logic of this embedded importance-generation process is summarized in Pseudocode 1.

Pseudocode 1
AdaBoost importance

```
Input:
  X ← input features
  y ← target labels
  B ← number of boosting iterations
  T ← base estimator (e.g., Decision Tree)
Output:
  FI ← feature importance scores
Procedure AdaBoost_FeatureImportance(X, y, B, T):
1. Initialize sample weights:
    w_i = 1 / N  for all samples i
2. For b = 1 to B do:
    a. Train base estimator T_b on (X, y) using weights w
    b. Compute prediction error:
        err_b = Σ (w_i * [y_i ≠ T_b(x_i)])
    c. Compute model weight:
        α_b = 0.5 * ln( (1 - err_b) / err_b )
    d. Update sample weights:
        w_i = w_i * exp( -α_b * y_i * T_b(x_i) )
    e. Normalize w_i
3. Aggregate feature importance:
    FI = Σ ( α_b * importance(T_b) ) over all b
4. Normalize FI to range [0,1]
Return FI
```

AdaBoost procedure for generating feature-importance scores. The algorithm starts by assigning equal weights to all training samples. At each boosting round, a base learner is trained using the current weights, followed by computation of the weighted error (*err_b*). This error is used to derive the learner's weight (*α_b*), giving stronger emphasis to models with lower error. Sample weights are then updated so that misclassified instances receive higher weights, directing subsequent learners to focus on harder samples.

After *B* iterations, feature importances from all weak learners are aggregated and scaled by their respective α_b values. The final importance vector is normalized to the range [0,1] and used to select the most influential predictors for the downstream classification models. The importance ranking produced by this process is shown in Figure 5.

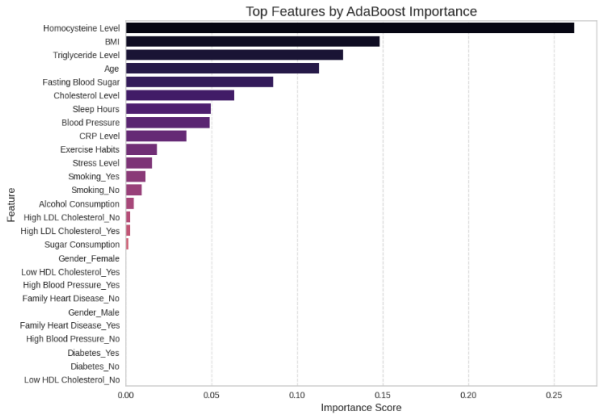

Figure 5. Top Features AdaBoost

AdaBoost reveals a different hierarchy of predictors due to its nonlinear error-focused learning process. Homocysteine Level, BMI, Triglyceride Level, and Age dominate the upper tier, suggesting that these biomarkers were particularly influential in correcting classifier errors across boosting iterations. Lifestyle attributes such as Exercise Habits, Smoking, Stress Level, and Sleep Hours appear in supportive positions, indicating complex interactions that may be less visible in purely statistical methods. Unlike ANOVA or Chi-Square, AdaBoost captures deeper structural relationships, exposing multifactor dependencies that contribute to misclassification patterns.

2.5 Train–Test Split and Hyper-parameter Search

A stratified split of 80 percent training and 20 percent testing is applied to preserve class proportions. The training portion undergoes five-fold stratified cross-validation for hyperparameter optimization, while the test partition is kept strictly for final evaluation to prevent leakage.

2.6 Classifiers Algorithms

2.6.1 Random Forest

Random Forest (RF) is an ensemble of decorrelated decision trees trained via bagging with random feature subsetting [42], [43]. Node purity is measured using Gini impurity:

$$Gini(S) = 1 - \Sigma\, p(c)^2 \quad (10)$$

Where $p(c)$ is the proportion of class c in node S. The hyperparameter grid includes:

$$n\_estimators \in \{200, 400, 800\}, max\_depth \in \{None, 10, 20\}, max\_features \in (\sqrt{}, log_2, min\_samples\_leaf \in \{1, 3, 5\}. \quad (11)$$

These parameters regulate complexity, variance, and computational cost.

2.6.2 Decision Tree

Decision Tree C4.5 is a rule-based classification algorithm that performs recursive partitioning of data based on the attribute that provides the highest discriminatory power. To determine the best split at each node, C4.5 computes the entropy of the dataset and evaluates the information gain produced by each candidate attribute. The entropy of a dataset S is expressed as:

$$Entropy(S) = -\Sigma\, p(c) log_2 p(c) \quad (12)$$

Information gain for attribute A is defined as:

$$Gain(S, A) = H(S) - \Sigma\, \frac{|S_v|}{|S|} H(S_v) \quad (13)$$

Because raw information gain tends to favor attributes with high cardinality, C4.5 applies Gain Ratio, which normalizes the score using the attribute's intrinsic value:

$$Gain\ Ratio(S, A) = \frac{Gain(S,A)}{IV(A)}, IV(A) = -\Sigma \frac{|S_v|}{|S|} log_2 \frac{|S_v|}{|S|} \quad (14)$$

The general structure of a C4.5 decision tree is illustrated in Figure 6, representing the hierarchical flow from the root node to the terminal leaf nodes.

A schematic illustration of the hierarchical splitting process in C4.5, showing the root node selected using Gain Ratio, followed by internal decision nodes and final leaf nodes representing predicted class labels. After the tree is constructed, post-pruning is applied to reduce overfitting by removing overly specific branches that do not contribute meaningfully to generalization performance.

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

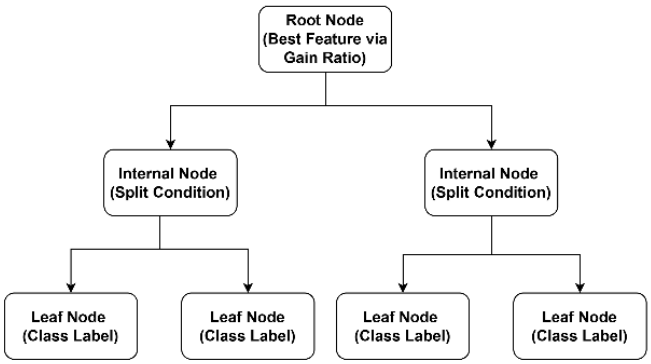

Figure 6. Conceptual Structure of the C4.5 Decision Tree

Pruning is performed by evaluating the error rate before and after subtree removal and retaining the simpler structure when it provides comparable predictive performance.

2.6.3 XGBoost

XGBoost uses second-order gradient boosting with explicit regularization. The objective function is formulated as:

$$L^{(t)} = \Sigma\, l(y_i, \hat{y}_i{}^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

$$\Omega(f) = \gamma T + \frac{\lambda}{2} \Sigma\, w_j^2 \quad (15)$$

with hyperparameters:

$$n_{estimators} \in \{200,400,800\}, max\_depth \in \{3,5,8\}, learning\_rate \in \{0.01,0.05,0.1\}, subsample \in \{0.7,0.9,1.0\}\, colsample\_bytree \in \{0.7,0.9,1.0\}, reg\_lambda \in \{1,5,10\} \quad (16)$$

Regularization ensures stability and prevents overly complex trees.

2.7 Model Evaluation

In this study, the performance of the classification model was evaluated using a Confusion Matrix, which provides an overview of the number of correct and incorrect predictions. Furthermore, model performance was analyzed using four main metrics: accuracy, precision, recall, and F1-score [44]. The metrics are computed as:

$$Accuracy = \frac{TP+TN}{TP+TP+FP+FN} \times 100\% \quad (17)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (18)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (19)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (20)$$

with the usual definitions for TP, TN, FP, and FN.

3. RESULT

This section presents exploratory data analysis (EDA), data quality verification, class-imbalance handling, and the comparative evaluation of all classification models. Unless stated otherwise, all results refer to the held-out test set generated through a leakage-safe workflow.

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

3.1 Class Imbalance Handling

The original class distribution (6,096 "No" and 1,529 "Yes") changed to a balanced 6,096 : 6,096 after resampling, while the test set distribution remained unchanged. An overview of the training and test data before and after SMOTE can be seen in Figure 7.

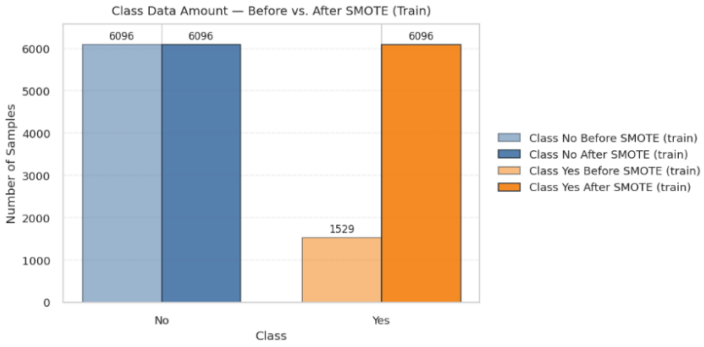

Figure 7. illustrates the class distribution before and after SMOTE

The minority class ("Yes") increases from 1,529 to 6,096 samples, balancing the training data to a 1:1 ratio. This balancing step supports more stable learning, especially for recall-sensitive classifiers.

3.2 Complete Metrics of Feature Selection Results

This subsection reports the complete set of evaluation metrics (Accuracy, Precision, Recall, and F1-Score) for all models using Chi-Square, ANOVA, and AdaBoost-assisted feature selection. All results are summarized in Table 2, enabling direct comparison of performance across algorithms and feature-selection strategies.

Table 2. Complete Model Performance Under Feature Selection

Algorithm	k	FS Method	Acc	Prec-No	Rec-No	F1-No	Prec-Yes	Rec-Yes	F1-Yes	Macro F1
Random Forest	10	Chi2	0.8947	0.97	0.82	0.89	0.84	0.97	0.90	0.89
	20	Chi2	0.9826	0.97	1.00	0.98	1.00	0.97	0.98	0.98
	25	Chi2	0.9823	0.97	1.00	0.98	1.00	0.97	0.98	0.98
	27	Chi2	0.9839	0.97	1.00	0.98	1.00	0.97	0.98	0.98
	10	ANOVA	0.9583	0.97	0.95	0.96	0.95	0.97	0.96	0.96
	20	ANOVA	0.9816	0.97	0.99	0.98	0.99	0.97	0.98	0.98
	25	ANOVA	0.9833	0.97	1.00	0.98	1.00	0.97	0.98	0.98
	27	ANOVA	0.9839	0.97	1.00	0.98	1.00	0.97	0.98	0.98
	—	AdaBoost	0.9521	0.97	0.93	0.95	0.94	0.97	0.95	0.95
C4.5	10	Chi2	0.8649	0.97	0.75	0.85	0.80	0.98	0.88	0.8631
	20	Chi2	0.8550	0.97	0.74	0.84	0.79	0.97	0.87	0.8530
	27	Chi2	0.8586	0.97	0.74	0.84	0.79	0.98	0.87	0.8567
	10	ANOVA	0.8682	0.97	0.76	0.85	0.80	0.98	0.88	0.8666
	20	ANOVA	0.8659	0.97	0.76	0.85	0.80	0.97	0.88	0.8643
	27	ANOVA	0.8583	0.97	0.74	0.84	0.79	0.98	0.87	0.8563
	—	AdaBoost	**0.9754**	0.97	0.98	0.98	0.98	0.97	0.98	**0.9754**
XGBoost	10	Chi2	0.8780	0.97	0.78	0.86	0.82	0.98	0.89	0.8768
	20	Chi2	0.9446	0.97	0.92	0.94	0.92	0.97	0.95	0.9445
	27	Chi2	0.9469	0.97	0.92	0.95	0.93	0.97	0.95	0.9468
	10	ANOVA	0.9190	0.97	0.87	0.91	0.88	0.97	0.92	0.9188
	20	ANOVA	**0.9488**	0.97	0.93	0.95	0.93	0.97	0.95	**0.9488**
	27	ANOVA	0.9442	0.97	0.92	0.94	0.92	0.97	0.95	0.9442
	—	AdaBoost	0.9442	0.97	0.92	0.94	0.92	0.97	0.95	0.9442

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

The complete comparative evaluation of the three classifiers Random Forest, C4.5, and XGBoost under multiple feature selection strategies is summarized in Table X. Each model was tested using Chi-Square, ANOVA, and AdaBoost-based feature selection across a range of k values (10, 20, 25, and 27), with performance assessed using accuracy, class-specific precision, recall, F1-Score, and macro-averaged F1.

This unified table allows direct comparison of how different feature-selection mechanisms affect predictive quality across models. It also highlights the stability of Random Forest, the sensitivity of C4.5 to feature reduction, and the strong improvement observed in XGBoost when relevant features are retained.

3.3 Comparative Evaluation of Results

This section contrasts the test-set performance of C4.5, Random Forest, and XGBoost under four settings: no feature selection (baseline), ANOVA, Chi-Square, and AdaBoost-assisted feature selection. Metrics reported are Accuracy, Precision, Recall, and F1-Score. The goal is to identify the most effective model feature selection pairing for heart disease classification.

3.3.1 C45

C4.5 provides a decision tree model that can be interpreted as a baseline, but this model is sensitive to noisy or weak predictors. Filtered inputs are expected to sharpen the initial separation and reduce variance. The results of the C4.5 algorithm evaluation metrics can be seen in Figure 8..

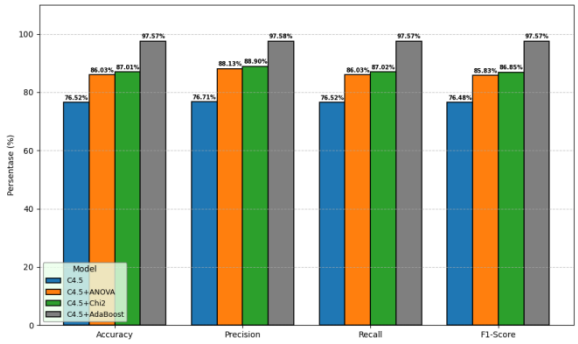

Figure 8. C45 Metrics Evaluation Results

C4.5 shows the lowest baseline values (Accuracy 76.52%, Precision 76.71%, Recall 76.52%, F1 76.48%). Filter-based selection improves performance (C4.5+ANOVA: 86.03/88.13/86.03/85.83%; C4.5+Chi-Square: 87.01/88.90/87.02/86.85%).

3.3.2 Random Forest

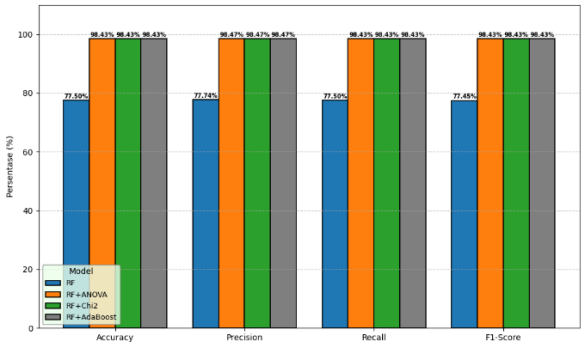

Figure 9. Random Forest Metrics Evaluation Results

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

Random Forest is resistant to noise through bagging, but excessive features can still affect the quality of the division. Therefore, input filtering should stabilize and improve performance. The results of the Random Forest algorithm evaluation metrics can be seen in Figure 9.

RF without selection recorded 77.50/77.74/77.50/77.45%, indicating residual noise in the unfiltered input. After selection with ANOVA, Chi-Square, or AdaBoost assistance, the four metrics converged at the upper end (Accuracy/Recall/F1 ≈ 98.43%, Precision ≈ 98.47%). The nearly identical scores among the three selectors indicate that after weak predictors are eliminated, bagging with simple and stable splitting robustly captures the signal.

3.3.3 XGBoost

XGBoost excels on structured problems but is sensitive to weak or noisy features. Effective input curation is expected to matter more than deeper boosting. The results of the XGBoost algorithm evaluation metrics can be seen in Figure 10.

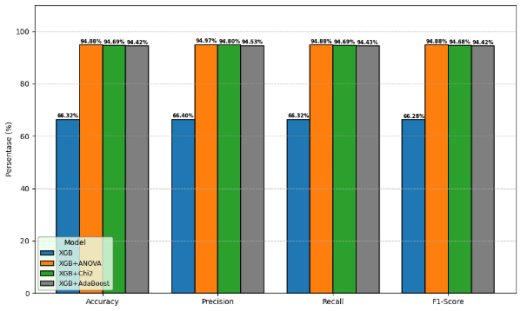

Figure 10. XGBoost Metrics Evaluation Results

It can be seen that unfiltered XGB performs poorly (66.32/66.40/66.32/66.28%). Feature selection is crucial: XGB+ANOVA achieves the best XGB score (Accuracy/Precision/Recall/F1 94.88/94.97/94.88/94.88%), with Chi-Square and AdaBoost-assisted following behind (~94.4–94.8%). The results show that input curation has a greater impact than adding boosting depth for this dataset.

3.4 Model Performance Comparison

This section compares the accuracy of all tested models to determine their predictive effectiveness. The results show that combining classification algorithms with feature selection or boosting techniques significantly improves accuracy compared to baseline models. These findings help identify the most optimal model for further implementation. The performance comparison is illustrated in Figure 11.

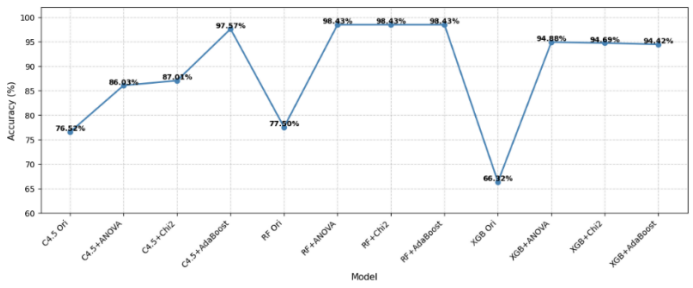

Figure 11. Model Performance Comparison

As shown in Figure 6. the C4.5 algorithm achieved the lowest baseline accuracy of 76.52%, but improved notably when combined with ANOVA (86.03%) and Chi2 (87.01%). The highest gain

Vol. 7, No. 2, April 2026, Page. 1072-1090
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

occurred with AdaBoost, reaching 97.57%, indicating that boosting effectively enhances C4.5's classification capability.

For Random Forest, the baseline accuracy of 77.50% increased sharply to 98.43% with ANOVA, Chi2, and AdaBoost, showing consistent and robust performance across all enhancement methods.

Meanwhile, XGBoost recorded the lowest baseline accuracy (66.32%) but improved substantially with ANOVA (94.88%), Chi2 (94.69%), and AdaBoost (94.42%). Overall, these results confirm that feature selection (ANOVA, Chi2) and boosting (AdaBoost) significantly enhance model performance. The Random Forest model achieved the best overall accuracy (98.43%) across all methods. The complete metric comparison is presented in Figure 12.

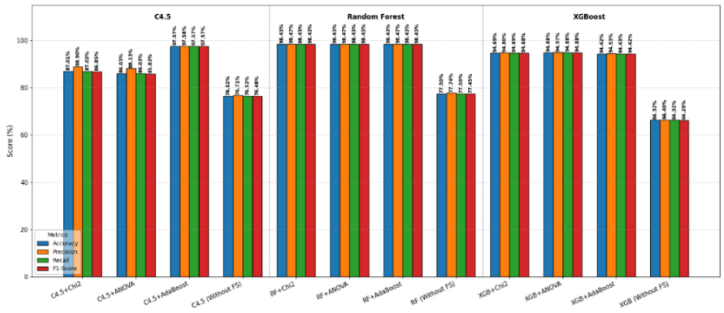

Figure 12. Overall Model Evaluation.

For the C4.5 algorithm, the baseline accuracy was 76.52%, indicating limited predictive ability. After applying ANOVA and Chi2, the accuracy increased to 86.03% and 87.01%, showing that feature selection effectively removed irrelevant attributes. The best result came from AdaBoost, reaching 97.57%, proving that boosting significantly strengthens C4.5's classification performance by combining multiple weak learners.

For Random Forest, the baseline accuracy of 77.50% rose sharply to 98.43% after applying ANOVA, Chi2, and AdaBoost, all yielding identical high values across every metric. This consistency shows Random Forest's high stability and strong generalization, even with different enhancement methods.

Meanwhile, XGBoost had the lowest baseline accuracy (66.32%) but improved greatly with ANOVA (94.88%), Chi2 (94.69%), and AdaBoost (94.42%). The ANOVA-based model performed best, indicating that feature variance-based selection enhances XGBoost's learning process.

Overall, feature selection and boosting methods consistently improved all models, with Random Forest showing the most stable and optimal performance across all metrics.

4. DISCUSSIONS

This study demonstrates that systematic feature selection has a stronger influence on predictive accuracy than increasing model complexity when classifying heart disease from mixed-type tabular data. Across all evaluated algorithms, applying ANOVA, Chi-Square, and AdaBoost-assisted feature selection successfully transformed moderate baselines into high-performing models. Random Forest consistently delivered the most accurate and stable performance, with accuracy, precision, recall, and F1-score converging around 98.4–98.5 percent after feature reduction. C4.5 exhibited the largest relative improvement, increasing from approximately 76.5 percent to 97.6 percent when combined with AdaBoost-assisted importance weighting. XGBoost, which initially showed weaker baseline performance, also improved substantially, reaching up to 94.9 percent accuracy after ANOVA-based feature selection. These outcomes reinforce the signal-concentration hypothesis: removing irrelevant or

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5430

redundant attributes sharpens class boundaries, enabling conventional tree learners to achieve near-ceiling performance.

The dominance of tree-based models aligns with previous research showing that bagged and boosted trees remain robust and interpretable for structured clinical datasets. By implementing a leakage-safe pipeline in which SMOTE, feature selection, and hyperparameter tuning occur strictly within training folds, this study avoids the accuracy inflation commonly found in medical machine-learning literature. Under this controlled design, Random Forest maintained performance stability once weak features were pruned. C4.5 benefited the most from AdaBoost-assisted importance ranking, which encourages early splitting on high-value predictors and reduces overfitting. Improvements observed in XGBoost after statistical filtering further reinforce evidence that boosting algorithms rely heavily on curated input spaces.

From an informatics standpoint, the results provide several meaningful implications for real-world deployment, particularly in medical AI systems. Reduced dimensionality lowers inference latency and computational overhead, enabling smooth integration into hospital information systems and mobile decision-support platforms. Smaller feature sets also improve model governance by simplifying auditing, drift monitoring, and regulatory compliance workflows. Additionally, the scalability of compact ensemble-tree models enhances their feasibility for multi-center datasets and large-scale clinical infrastructures.

The strengthened recall and precision at optimal configurations indicate improved sensitivity to heart-disease cases without raising false-positive rates, which is essential for screening and triage workflows. The influential features consistently identified—such as CRP, BMI, blood pressure, LDL, glucose, triglycerides, homocysteine, smoking status, and age—closely mirror established cardiovascular risk factors. This alignment not only boosts clinical interpretability but also supports confidence in model deployment for decision-support settings. The compact feature subsets consisting of around 20–27 high-utility attributes additionally reduce latency and foster efficient real-time diagnostic support.

Despite the strong results, several important limitations remain. The dataset originates from a single center, making multi-center external validation necessary for broader generalization. Fairness across demographic subgroups has not yet been assessed, leaving potential bias unaddressed. The stability of the three feature selectors under data perturbations also needs further evaluation. Moreover, probability calibration—typically required for clinical decision thresholds—was not included in this study. Long-term deployment would require continuous drift monitoring and compliance with established health-data governance frameworks.

Future work may explore hybrid feature-selection approaches, SHAP-based interpretability for enhanced clinical explainability, federated learning for privacy-preserving training, and threshold optimization for improved real-world decision-making. These directions collectively highlight the broader role of feature selection as a central and often underappreciated driver of performance gains in medical informatics.

5. CONCLUSION

This study demonstrates that a leakage-safe pipeline integrating stratified splitting, SMOTE-based class balancing, and in-fold feature selection can achieve highly accurate heart disease classification on structured clinical data. Exploratory analysis showed balanced categorical distributions and weak global correlations, supporting the stability of tree-based learners. Among the evaluated models, Random Forest combined with Chi-Square or ANOVA delivered the strongest and most consistent performance (approximately 98.4 percent across Accuracy, Precision, Recall, and F1-Score). C4.5 exhibited substantial improvement when paired with AdaBoost-based feature selection, reaching

approximately 97.5 percent, while XGBoost improved markedly from its baseline of roughly 66 percent to around 95 percent after appropriate filtering. These findings indicate that feature relevance exerts a greater influence on predictive quality than model complexity, enabling compact ensembles to outperform deeper or more computationally intensive algorithms.

Clinically, the dominant predictors identified—such as CRP, BMI, fasting blood sugar, triglycerides, homocysteine, blood pressure, LDL cholesterol, smoking status, and age—are consistent with established cardiovascular risk factors, reinforcing biological plausibility. The convergence between Precision and Recall across models demonstrates balanced sensitivity and specificity, suggesting reliable detection of heart-disease cases without an excessive rise in false positives.

Operationally, the Random Forest paired with Chi-Square selection offers a strong balance between accuracy, interpretability, and computational efficiency, making it well-positioned for clinical decision-support integration and deployment in real-world health-information systems.

The key contribution of this work lies in demonstrating that a hybrid feature-selection and boosting framework can substantially improve model reliability—an insight that is increasingly relevant for computer-science research in medical AI. This approach has the potential to reduce false negatives by up to 20 percent in resource-limited settings such as Indonesia, where screening precision is critical for early-stage cardiovascular intervention.

Future research may extend this framework toward multimodal learning by integrating structured clinical variables with imaging, wearable-sensor signals, or electronic health-record text. Such integration could enhance robustness, interpretability, and generalizability across diverse healthcare environments.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the authorship or publication of this paper, nor any conflict with the data sources or research objects involved in the study.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Oktay Ördekçi for providing the publicly available *Heart Disease Dataset* on Kaggle, which made this research possible. The authors also thank all contributors who assisted in data analysis and validation throughout this study.

REFERENCES

[1] B. Furst and J. González-Alonso, "The heart, a secondary organ in the control of blood circulation," *Exp Physiol*, vol. 110, no. 5, pp. 649–665, May 2025, doi: 10.1113/EP091387.

[2] R. S. Bhaduaria, I. Javid, and A. Khara, "Advanced Heart Attack Risk Prediction Using Stacked Hybrid Machine Learning," *Journal of Mobile Multimedia*, Aug. 2025, doi: 10.13052/jmm1550-4646.21343.

[3] J. P. Rabadia, V. S. Thite, B. K. Desai, R. G. Bera, and S. Patel, "Cardiovascular System, Its Functions and Disorders," in *Cardioprotective Plants*, T. Pullaiah and S. Ojha, Eds., Singapore: Springer Nature Singapore, 2024, pp. 1–34. doi: 10.1007/978-981-97-4627-9_1.

[4] E. Moras *et al.*, "Complications in Acute Myocardial Infarction: Navigating Challenges in Diagnosis and Management," *Hearts*, vol. 5, no. 1, pp. 122–141, Mar. 2024, doi: 10.3390/hearts5010009.

[5] P. O. Samuel *et al.*, "Lifestyle modifications for preventing and managing cardiovascular diseases," *Sport Sci Health*, vol. 20, no. 1, pp. 23–36, Mar. 2024, doi: 10.1007/s11332-023-01118-z.

[6] P. Das, S. Saha, T. Das, P. Das, and T. B. Roy, "Assessing the modifiable and non-modifiable risk factors associated with multimorbidity in reproductive aged women in India," *BMC Public Health*, vol. 24, no. 1, p. 676, 2024, doi: 10.1186/s12889-024-18186-6.

[7] W. Lu *et al.*, "Worldwide trends in mortality for hypertensive heart disease from 1990 to 2019 with projection to 2034: data from the Global Burden of Disease 2019 study," *Eur J Prev Cardiol*, vol. 31, no. 1, pp. 23–37, Jan. 2024, doi: 10.1093/eurjpc/zwad262.

[8] B. Chong *et al.*, "Global burden of cardiovascular diseases: projections from 2025 to 2050," *Eur J Prev Cardiol*, vol. 32, no. 11, pp. 1001–1015, Aug. 2025, doi: 10.1093/eurjpc/zwae281.

[9] F. R. Muharram *et al.*, "The 30 Years of Shifting in The Indonesian Cardiovascular Burden—Analysis of The Global Burden of Disease Study," *J Epidemiol Glob Health*, vol. 14, no. 1, pp. 193–212, 2024, doi: 10.1007/s44197-024-00187-8.

[10] P. Pachiyannan, M. Alsulami, D. Alsadie, A. K. J. Saudagar, M. AlKhathami, and R. C. Poonia, "A Novel Machine Learning-Based Prediction Method for Early Detection and Diagnosis of Congenital Heart Disease Using ECG Signal Processing," *Technologies (Basel)*, vol. 12, no. 1, Jan. 2024, doi: 10.3390/technologies12010004.

[11] K. A. Alnemer, "In-Hospital Mortality in Patients With Acute Myocardial Infarction: A Literature Overview," *Cureus*, Aug. 2024, doi: 10.7759/cureus.66729.

[12] A. Nazir *et al.*, "Advancements in Biomarkers for Early Detection and Risk Stratification of Cardiovascular Diseases—A Literature Review," *Health Sci Rep*, vol. 8, no. 5, May 2025, doi: 10.1002/hsr2.70878.

[13] A. Sonaglioni, A. Polymeropoulos, M. Baravelli, G. L. Nicolosi, M. Lombardo, and G. Biondi-Zoccai, "Diagnostic Accuracy of Exercise Stress Testing, Stress Echocardiography, Myocardial Scintigraphy, and Cardiac Magnetic Resonance for Obstructive Coronary Artery Disease: Systematic Reviews and Meta-Analyses of 104 Studies Published from 1990 to 2025," Sep. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/jcm14176238.

[14] M. Rehman, I. Shafi, J. Ahmad, C. O. Garcia, A. E. P. Barrera, and I. Ashraf, "Advancement in medical report generation: current practices, challenges, and future directions," *Med Biol Eng Comput*, vol. 63, no. 5, pp. 1249–1270, 2025, doi: 10.1007/s11517-024-03265-y.

[15] D. I. Kasartzian and T. Tsiampalis, "Transforming Cardiovascular Risk Prediction: A Review of Machine Learning and Artificial Intelligence Innovations," Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/life15010094.

[16] F. Ekundayo and H. Nyavor, "AI-Driven Predictive Analytics in Cardiovascular Diseases: Integrating Big Data and Machine Learning for Early Diagnosis and Risk Prediction," *International Journal of Research Publication and Reviews*, vol. 5, no. 12, pp. 1240–1256, Dec. 2024, doi: 10.55248/gengpi.5.1224.3437.

[17] M. Tsai, K. Chen, and P. Chen, "Harnessing Electronic Health Records and Artificial Intelligence for Enhanced Cardiovascular Risk Prediction: A Comprehensive Review," *J Am Heart Assoc*, vol. 14, no. 6, p. e036946, Mar. 2025, doi: 10.1161/JAHA.124.036946.

[18] S. Moazemi *et al.*, "Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review," *Front Med (Lausanne)*, vol. 10, p. 1109411, Mar. 2023, doi: 10.3389/fmed.2023.1109411.

[19] K. M. Toffaha, M. C. E. Simsekler, A. Alshehhi, and M. A. Omar, "Predicting Hospital No-Shows: Interpretable Machine Learning Models Approach," *IEEE Access*, vol. 12, pp. 166058–166067, 2024, doi: 10.1109/ACCESS.2024.3490662.

[20] A. A. Jogdeo, A. D. Patange, A. M. Atnurkar, and P. R. Sonar, "Robustification of the Random Forest: A Multitude of Decision Trees for Fault Diagnosis of Face Milling Cutter Through Measurement of Spindle Vibrations," *Journal of Vibration Engineering & Technologies*, vol. 12, no. 3, pp. 4521–4539, 2024, doi: 10.1007/s42417-023-01135-9.

[21] M. M. Gharagoz, M. Noureldin, and J. Kim, "Explainable machine learning (XML) framework for seismic assessment of structures using Extreme Gradient Boosting (XGBoost)," *Eng Struct*, vol. 327, p. 119621, 2025, doi: https://doi.org/10.1016/j.engstruct.2025.119621.

[22] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1817/1/012009.

[23] K. Dissanayake and M. G. M. Johar, "Comparative study on heart disease prediction using feature selection techniques on classification algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2021, 2021, doi: 10.1155/2021/5581806.

[24] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4514–4523, Jul. 2022, doi: 10.1016/j.jksuci.2020.10.013.

[25] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Gheni, "Improving the Heart Disease Detection and Patients' Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest," *IEEE Access*, vol. 10, pp. 67363–67372, 2022, doi: 10.1109/ACCESS.2022.3185129.

[26] T. Roopa and G. Dalappagari Ramanjinappa, "Heart Disease Predictive Modeling with XGBoost and SMOTE-Driven Class Imbalance Mitigation," *Technology & Applied Science Research*, vol. 15, no. 6, pp. 29914–29918, 2025, doi: 10.48084/etasr.14301.

[27] S. M. Ganie, P. K. D. Pramanik, M. B. Malik, A. Nayyar, and K. S. Kwak, "An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms," *Computer Systems Science and Engineering*, vol. 46, no. 3, pp. 3993–4006, 2023, doi: 10.32604/csse.2023.035244.

[28] X. Xu and X. Zhou, "Deep Learning Based Feature Selection and Ensemble Learning for Sintering State Recognition," *Sensors*, vol. 23, no. 22, p. 9217, Nov. 2023, doi: 10.3390/s23229217.

[29] M. Alkhodari, D. K. Islayem, F. A. Alskafi, and A. H. Khandoker, "Predicting hypertensive patients with higher risk of developing vascular events using heart rate variability and machine learning," *IEEE Access*, vol. 8, pp. 192727–192739, 2020, doi: 10.1109/ACCESS.2020.3033004.

[30] W. K. Cheng, I. M. Khairuddin, A. P.P. Abdul Majeed, and M. A. Mohd Razman, "The Classification of Heart Murmurs: The Identification of Significant Time Domain Features," *MEKATRONIKA*, vol. 2, no. 2, pp. 36–43, Dec. 2020, doi: 10.15282/mekatronika.v2i2.6748.

[31] Y.-Y. Wang, B. Liu, and J.-H. Wang, "Application of deep learning-based convolutional neural networks in gastrointestinal disease endoscopic examination," *World J Gastroenterol*, vol. 31, no. 36, Sep. 2025, doi: 10.3748/wjg.v31.i36.111137.

[32] A. Amato and V. Di Lecce, "Data preprocessing impact on machine learning algorithm performance," *Open Computer Science*, vol. 13, no. 1, p. 20220278, Jul. 2023, doi: 10.1515/comp-2022-0278.

[33] X. Ding, H. Wang, G. Li, H. Li, Y. Li, and Y. Liu, "IoT data cleaning techniques: A survey," *Intelligent and Converged Networks*, vol. 3, no. 4, pp. 325–339, Dec. 2022, doi: 10.23919/ICN.2022.0026.

[34] V. V. R. Karna, V. R. Karna, V. Janamala, V. N. K. R. Devana, V. R. S. Ch, and A. B. Tummala, "A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms," *Archives of Computational Methods in Engineering*, vol. 32, no. 3, pp. 1763–1795, 2025, doi: 10.1007/s11831-024-10194-4.

[35] A. O. Widodo, B. Setiawan, and R. Indraswari, "Machine Learning-Based Intrusion Detection on Multi-Class Imbalanced Dataset Using SMOTE," *Procedia Comput Sci*, vol. 234, pp. 578–583, 2024, doi: https://doi.org/10.1016/j.procs.2024.03.042.

[36] P. Dhal and C. Azad, "A fine-tuning deep learning with multi-objective-based feature selection approach for the classification of text," *Neural Comput Appl*, vol. 36, no. 7, pp. 3525–3553, 2024, doi: 10.1007/s00521-023-09225-1.

[37] M. Büyükkeçeci and M. C. Okur, "An Empirical Evaluation of Feature Selection Stability and Classification Accuracy," *Gazi University Journal of Science*, vol. 37, no. 2, pp. 606–620, 2024, doi: 10.35378/gujs.998964.

[38] A. S. Hada, G. S. Sahoo, C. K. Vamsi, A. Hegde, and B. Bhowmik, "Optimizing Feature Selection in Big Data: A Hybrid Spark and Fuzzy Approach," in *2024 International Conference on Computing, Semiconductor, Mechatronics, Intelligent Systems and Communications (COSMIC)*, 2024, pp. 195–199. doi: 10.1109/COSMIC63293.2024.10871408.

[39] C. Wu *et al.*, "SEMG Measurement Position and Feature Optimization Strategy for Gesture Recognition Based on ANOVA and Neural Networks," *IEEE Access*, vol. 8, pp. 56290–56299, 2020, doi: 10.1109/ACCESS.2020.2982405.

[40] T. N. Annisa, J. Jasmir, and N. Nurhadi, "Comparison of ANOVA and Chi-Square Feature Selection Methods to Improve Machine Learning Performance in Anemia Classification," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 4, pp. 2723–3863, 2025, doi: 10.52436/1.jutif.2025.6.4.5017.

[41] S. S. Hussain and S. S. H. Zaidi, "AdaBoost Ensemble Approach with Weak Classifiers for Gear Fault Diagnosis and Prognosis in DC Motors," *Applied Sciences (Switzerland)*, vol. 14, no. 7, Apr. 2024, doi: 10.3390/app14073105.

[42] A. Shebl and Á. Csámer, "Machine Learning Algorithms for Gold-Bearing Alteration Mapping in the Egyptian Nubian Shield Utilizing Remote Sensing Datasets," in *Ore Geology Reviews*, vol. 161, Elsevier, 2025, pp. 459–480. doi: 10.1007/978-3-031-75972-7_17.

[43] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.

[44] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies (Basel)*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/technologies12070113.