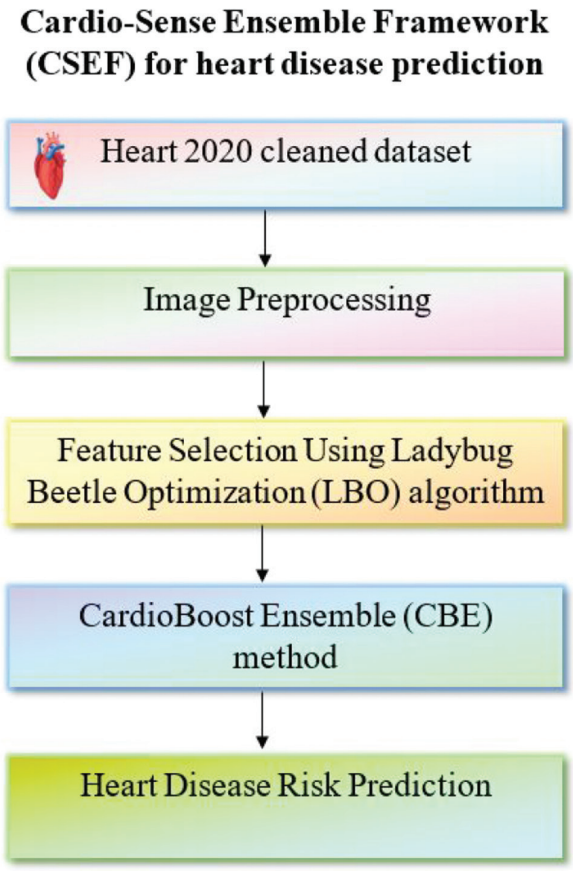
Cardiovascular Innovations and Applications

Vol. 11 (2026) 22
ISSN 2009-8618
DOI 10.15212/CVIA.2026.0007

RESEARCH ARTICLE

Performance Comparison of Machine Learning Classifiers and Cardio-Sense Ensemble for Heart Disease Detection

Graphical Abstract



Highlights

- A robust Cardio-Sense Ensemble Framework is proposed for accurate and reliable heart disease prediction.
- A combined SMOTE–ENN preprocessing strategy effectively addresses missing values, noise, and class imbalance while preserving data distribution.
- Ladybug Beetle Optimization (LBO) is employed for feature selection to reduce dimensionality and eliminate redundant and irrelevant features.
- A CardioBoost Ensemble (CBE) integrating LR, RF, KNN, and Gradient Boosting enhances baseline predictive performance.
- LightGBM meta-classification achieves superior accuracy by handling feature overlap and improving decision boundary learning.
- The proposed framework ensures high prediction accuracy, reduced overfitting, and strong clinical applicability for early heart disease screening.

Authors

Hemalata Nawale and Mangesh Nikose

Correspondence

hnawale@gmail.com (H. Nawale)

In brief

The proposed study introduces a robust Cardio-Sense Ensemble Framework (CSEF) for accurate heart disease prediction by integrating advanced preprocessing, feature selection, and ensemble learning techniques. It combines SMOTE–ENN for data balancing, Ladybug Beetle Optimization for feature selection, and a hybrid ensemble with LightGBM for classification. The framework achieves high predictive performance (≈98% accuracy and AUC), reducing overfitting and improving clinical reliability. Overall, it provides an efficient and scalable solution for early cardiovascular disease screening and decision support.

©2026 The Authors. Creative Commons Attribution 4.0 International License

compuscript

Cardiovascular Innovations and Applications

RESEARCH ARTICLE

Performance Comparison of Machine Learning Classifiers and Cardio-Sense Ensemble for Heart Disease Detection

Hemalata Nawale[1] and Mangesh Nikose[1]

[1]*Department of Electrical and Electronics Engineering, Sandip University Nashik, Nashik, MH, India*

Received: 15 September 2025; Revised: 18 November 2025; Accepted: 6 March 2026

Abstract

Objective: Heart disease remains a leading cause of global mortality; consequently, accurate, reliable, and interpretable predictive models are needed for early diagnosis. This study was aimed at developing a robust hybrid ensemble learning framework that improves heart disease predictive accuracy while preserving clinical interpretability.

Methods: The Cardio-Sense Ensemble Framework (CSEF) for heart disease prediction was developed by using the publicly available Behavioral Risk Factor Surveillance System dataset. Laplacian binary optimization was used for optimized feature selection, to eliminate redundancy and enhance discriminative information. The refined feature set was used to train multiple baseline classifiers, including logistic regression, random forest, K-nearest neighbors, and XGBoost. Their outputs were combined with Light Gradient Boosting Machine as a meta-classifier. Model performance was assessed with accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC).

Results: The CSEF achieved an accuracy of 0.98 and an AUC of 0.98, thus outperforming conventional machine learning models while demonstrating improved sensitivity and specificity.

Conclusion: Our hybrid ensemble framework offers superior predictive performance and interpretability; therefore, it is suitable for early cardiovascular risk screening and clinical decision-support integration.

Keywords: heart disease prediction; cardiovascular diseases; SMOTE; edited nearest neighbor; ladybug beetle optimization Light Gradient Boosting Machine

Introduction

Heart disease is a major global health concern responsible for nearly 31% of all deaths annually [1]. According to the World Health Organization, every individual deserves access to adequate healthcare services, including timely screening and preventive diagnosis [2]. Despite medical advancements, the prevalence of cardiovascular disease (CVD) continues to increase, because of sedentary lifestyles, unhealthy diets, and delayed detection [3]. In developing countries, inadequate healthcare infrastructure, limited diagnostic resources, and a shortage of medical professionals further complicate early diagnosis [4]. CVD includes conditions such as coronary artery disease, heart failure, stroke, and peripheral arterial disease [5]. Factors such as arterial plaque buildup, irregular heart rhythms, and vessel narrowing often lead to severe complications, including angina, myocardial infarction, and

Correspondence: **Hemalata Nawale**, Department of Electrical and Electronics Engineering Sandip University Nashik, Maharashtra, lndia. E-mail: hnawale@gmail.com

cardiac arrest [6, 7]. The World Health Organization projects that, by 2030, heart disease and stroke will cause more than 23.6 million deaths annually [8]. Therefore, early detection and prediction of cardiac conditions are critical to decreasing mortality and minimizing healthcare burdens [9].

Identifying CVD is challenging because of overlapping clinical risk factors such as diabetes, hypertension, obesity, smoking, and elevated cholesterol levels [10–12]. These interrelated factors complicate diagnosis and often result in delayed or inaccurate detection. Moreover, medical data collected in real-world clinical environments are frequently incomplete, noisy, or inconsistent, thus making accurate prediction even more complex [13–15]. Traditional diagnostic methods, although effective, heavily rely on manual interpretation by healthcare professionals, are time-consuming and prone to human error, and lack effectiveness for early-stage identification [16].

Machine learning (ML) techniques have emerged as a powerful alternative for disease prediction, detection, and diagnosis [17]. ML algorithms can efficiently analyze large, complex datasets, identify hidden relationships among risk factors, and provide objective decision support for clinicians [18]. In recent years, data-driven approaches have been widely used to predict the likelihood of heart disease, thereby enabling faster and more accurate screening [19, 20]. Studies using ML models such as logistic regression (LR), support vector machine (SVM), random forest (RF), decision tree (DT), and Extreme Gradient Boosting (XGBoost) have demonstrated promising results in heart disease prediction [21–23].

However, these traditional models face several challenges, including the presence of missing values, class imbalance, redundant features, and high computational complexity [24]. These limitations can lead to model overfitting, diminished generalization capability, and suboptimal diagnostic accuracy in real-world settings [25]. Additionally, heuristic feature selection techniques often struggle to identify optimal feature subsets, whereas conventional ensemble methods might not be able to effectively handle imbalanced medical datasets [26]. These challenges highlight the need for a robust, optimized ML framework that can enhance diagnostic reliability, minimize false predictions, and provide interpretable outcomes suitable for clinical decision support.

This article is structured as follows.

- A hybrid strategy is used to address class imbalance, impute missing values, and eliminate noisy samples, thereby enhancing dataset quality and model generalization.
- A novel algorithm is introduced for efficient, low-overhead feature selection, thus reducing dimensionality while retaining key predictive attributes.
- A novel model is described, which combines multiple base classifiers and uses LightGBM as the meta-classifier to enhance accuracy, even in cases of overlapping feature distributions.
- The integrated approach is demonstrated to improve diagnostic precision, minimize false predictions, and provide a scalable, clinically applicable framework for early detection of heart disease.

The remainder of the article is structured as follows: Section 2 reviews the relevant literature, Section 3 explains the developed approach, Section 4 covers the outcomes of the experiments and comparative analysis, and Section 5 concludes the article with final remarks.

Literature Survey

Rath et al. [27] have examined the HD categorization ability of one ensemble model and three ML techniques. These methods were established through training on ECG recordings that were unbalanced. The AdaBoost outperformed the SVM and LR-based classification models, according to accuracy plots and three performance metrics. This finding held for both datasets. Compared with individual models, on the basis of the majority voting concept, the LR-AdaBoost ensemble model exhibited the highest accuracy, F1-score, and AUC values. However, the use of unbalanced ECG training datasets might potentially have led to biased categorization, because the model became inclined to predict the majority class and to overlook less frequent cases.

Kumar et al. [28] have classified individuals with heart disease by using LR and SVMs, to help physicians improve patients' lifestyle and overall health. Two datasets in CSV file format, a heart dataset and

CHD dataset, were used. The accuracy and graphical categorization of each algorithm were applied to both datasets and compared. The authors chose SVM and LR for classification, after considering several algorithms and techniques for predicting heart disease. This study was aimed at improving physicians' understanding and care of patients with heart disease. The models' limitations included their inability to focus on more than a few factors.

Asif et al. [29] have proposed a model using ML to forecast cardiac illness, on the basis of several preprocessing stages, ensemble learning methods, and hyperparameter optimization strategies. Three Kaggle datasets with comparable attributes were integrated to produce a comprehensive dataset for analysis, to evaluate model efficacy through cross-validation (CV) of grid search for hyperparameter optimization, data normalization, and application of the additional tree classifier. Study limitations included the focus on a particular set of ensemble learning strategies, which might have limited the model's predictive ability. Furthermore, overfitting might have resulted from the dataset's small size and lack of variety, and hindered the model's generalizability to larger populations.

Ghasemieh et al. [30], using ensemble learning, have developed a novel model for identifying individuals requiring emergency care before heart attacks or cardiac arrhythmias, acute heart failure, and unstable angina, as well as patients with heart disease requiring emergency readmission. To optimize detection performance and forecast whether a patient with cardiac issues requires hospitalization after a preliminary admittance, the Stacking Ensemble Learner (SEL) was developed for emergency cases. To ensure stability and excellent accuracy over several runs, the SEL model uses meta-learning with XGBoost. The study's strengths include its use of behavior-based features, a strong SEL with XGBoost, and an extensive private dataset from the MIT Laboratory for Computational Physiology. However, a potential drawback of the stacking model is that the high computational complexity due to training several models might be problematic in medical applications, in which time and computer resources are often scarce.

Nissa et al. [31] have provided a comprehensive framework using existing ML and boosting methods such as Cat boost, random forest, gradient boosting, Light Gradient Boosting Machine (LightGBM), and Ada boost to predict impaired heart function, arterial blockage, or disrupted blood circulation. This study was aimed at advancing the development of trustworthy models to precisely forecast the cardiovascular health hazards by focusing on early heart disease prediction via boosting techniques. However, the predictive potential was limited by the use of small secondary datasets, which might not have captured the full variability of the population, thus leading to overfitting or underfitting. The model's generalizability to novel, unseen data might therefore be restricted, thereby compromising the accuracy and reliability of heart disease predictions.

Mondal et al. [32] have introduced a new dual-stage stacking ML as the basis for a computational risk prediction model for heart conditions. The initial prediction model was created with five ML classifiers and a dataset comprising 11 important variables from 1190 patients from five sources. The classifiers underwent ten iterations of CV to ensure their robustness and generalizability. The hyperparameter tuning techniques of RandomizedSearchCV and GridSearchCV, whose goal is to determine ideal estimator values, were used to maximize the model's efficiency. Through a stacking ensemble technique, the highest-performing approaches, RF, XGBoost, and DT, were further refined. However, the model's lack of feature selection might result in duplicate or unnecessary features that would increase computational complexity and lower model performance. Moreover, the model might overfit and capture noise instead of significant patterns if features are not chosen carefully. High-dimensional data can contribute extraneous information and therefore impair the model's ability to generalize well.

Sultan et al. [33] have proposed the NCDG stacking ensemble model, using Gradient Boosting as the meta-learner, and naive Bayes, categorical boosting, and DTs as base learners. They performed preprocessing by applying the factorization method to convert string columns to integers. BorderLineSMOTE and SMOTE were applied to address data class imbalance. Additionally, hard and soft voting classifiers were implemented, and their performance was compared with the stacking model. The SHapley Additive exPlanations (SHAP) method was used to evaluate the explainability of the NCDG model's artificial intelligence decisions. However, managing

class imbalance might result in overfitting by oversampling instances of minority classes, as a result of the creation of synthetic samples from pre-existing data points, which might duplicate noise or outliers. The model's results would consequently appear manipulated, thus limiting its ability to generalize to untested data.

Ahmed et al. [34] have suggested a highly effective ML method for predicting heart failure survival by using a Gradient Boosting Machine (GBM) and adaptive inertia weight particle swarm optimization. SelectKBest and SMOTE were used to balance the dataset, and chi-square feature selection approaches were used to retain the most significant predictors and resolve the imbalance in survival outcomes. The adaptive inertia weight particle swarm optimization technique, which adaptively adjusted inertia weights to balance exploration and exploitation, was applied to determine the optimal hyperparameters for the GBM model. Information criteria such as the Bayesian information criterion and Akaike information criterion were used to further improve the model selection process. This process ensured selection of the ideal model, because of both predicted accuracy and model complexity. However, this study's small dataset, which excluded a variety of clinical and demographic characteristics, limited the model's ability to account for all variables affecting survival and might result in forecasts that lack precision and understandability. Furthermore, the model's ability to adapt to various clinical contexts might have been limited by the optimization strategy used.

The reviewed studies indicated that ML and ensemble approaches have markedly advanced heart disease prediction. However, several critical challenges remain unresolved. Most existing studies have drawbacks of class imbalance, missing values, and noisy data, thus leading to biased learning and poor generalization. Although oversampling methods such as SMOTE improve minority class representation, they can also introduce synthetic noise and overfitting. Similarly, traditional feature selection and optimization techniques often exhibit high computational complexity and are prone to local optima, thereby resulting in redundant or irrelevant features that degrade accuracy. Moreover, overlapping feature distributions between healthy and diseased individuals continue to cause misclassification and decrease diagnostic reliability.

These limitations highlight the need to develop a hybrid ensemble learning framework that can overcome prior shortcomings, by providing a balanced, optimized, and clinically reliable framework for early heart disease prediction.

Developed Method

Heart disease is a leading cause of death globally; therefore, accurate and timely forecasting mechanisms are needed. This study developed a robust Cardio-Sense Ensemble Framework (CSEF) for heart disease prediction, which was designed to increase predictive accuracy, efficiently manage missing data, balance datasets, and identify the most relevant features for optimal model performance. The methods for predicting heart disease focused on ensuring data quality, balancing class distributions, selecting relevant features, and applying robust classification techniques. Initial preprocessing for heart disease prediction can lead to information loss when imputation decreases data variability, or normalization distorts non-linear relationships. Bias can arise from improper handling of missing values or outlier removal, thus eliminating critical data. Additionally, SMOTE can generate unrealistic synthetic samples, and lead to noise and overfitting, particularly when outliers or overlaps exist between the minority and majority classes. To effectively address the challenges posed by missing values, outliers, and class imbalance, we combined use of SMOTE and ENN to provide a robust preprocessing strategy. The edited nearest neighbor (ENN) technique serves a dual purpose. First, it imputes missing values by leveraging the similarity of features among nearest neighbors, thereby preserving the underlying data distribution and decreasing the risk of variability loss or imputation bias. Second, it removes noisy and misclassified samples, and therefore improves data quality. In addition, SMOTE produces synthetic instances for the minority class by adaptively interpolating between hard-to-learn samples. This approach decreases the likelihood of introducing unrealistic or overlapping data points and effectively minimizes noise. Together, SMOTE and ENN not only balance the class distribution but also enhance the model's generalization ability by

ensuring a more representative, cleaner dataset for training. Feature selection methods for heart disease prediction face several challenges. Existing heuristic methods, although powerful, are computationally intensive and sensitive to parameter tuning, thus risking suboptimal feature sets due to convergence in local optima. Additionally, a lack of proper feature selection can introduce irrelevant or redundant features, thereby increasing computational complexity and causing overfitting by capturing noise instead of meaningful patterns. To overcome these challenges, the ladybug beetle optimization (LBO) algorithm, inspired by ladybugs' hunting habits, was developed. LBO efficiently balances exploration and exploitation, and can identify the most significant features while reducing dimensionality. By retaining only key predictive features, LBO increases model efficiency while also decreasing overfitting hazards and the computational overhead associated with irrelevant data. For the classification task, the feature distributions of healthy and diseased cases overlap; therefore, existing classification algorithms frequently encounter difficulties in predicting cardiac disease. Models struggle to establish distinct decision limits, because of the similarity of important features, such as blood pressure and cholesterol levels. This overlap causes misclassification, which in turn can lead to false negatives, in which real cases of cardiac disease are overlooked, and false positives, in which healthy patients might receive unnecessary therapies. These mistakes have the potential to impair predictive accuracy and have severe detrimental effects. To overcome these challenges, we used the CardioBoost Ensemble (CBE) method to classify heart disease prediction by using a range of ML models, including LR, RF, K-nearest neighbors (KNN), and Gradient Boosting Classifiers, to establish baseline performance. The final classification used LightGBM, an ensemble boosting model known for its speed, efficiency, and superior performance on large datasets. LightGBM's leaf-wise growth strategy, combined with hyperparameter tuning, decreases loss and increases predictive accuracy. This integrated method ensured a comprehensive approach to heart disease prediction with high accuracy, computational efficiency, and clinical relevance. Figure 1 presents the step-by-step architecture of the heart disease prediction framework, showing how the input undergoes SMOTE-ENN preprocessing, LBO-driven feature refinement, and final classification through the CBE with LightGBM.

Preprocessing with SMOTE ENN

The Behavioral Risk Factor Surveillance System (BRFSS) Heart Disease Dataset exhibits a significant class imbalance, in which the positive class (heart disease cases) represents only a small fraction of the total population. This imbalance leads to biased learning behavior, because most classification algorithms tend to prioritize the majority class, thereby resulting in high false-negative rates, poor sensitivity, and limited generalization. Additionally, the dataset contains noisy samples, overlapping class regions, and mislabeled instances, because its population-based, self-reported nature further degrades predictive accuracy.

To address these dataset-specific challenges, our framework used SMOTE-ENN, a hybrid resampling strategy that effectively balances minority-class representation while simultaneously reducing noise. SMOTE creates artificial minority class samples to rectify class imbalance through interpolation between existing instances. Specifically, new data points are created with the following formula (1):

$$x_{new} = x_i + \lambda \cdot (x_{NN} - x_i) \quad (1)$$

where x_i represents a minority class sample, x_{NN} is one of its k-nearest neighbors, and λ is a random number between 0 and 1. This controlled interpolation is particularly helpful for BRFSS, because it expands the sparse minority region without simply duplicating samples.

However, SMOTE alone may introduce borderline or noisy instances in regions where BRFSS data already have overlapping feature distributions. To mitigate this effect, we applied ENN immediately afterward, to remove samples whose class labels differ from those of the majority of their nearest neighbors. This filtering step eliminated ambiguous, mislabeled, or conflicting data points, which are common in large-scale survey datasets such as BRFSS. The SMOTE-ENN mechanism first increased the representation of minority heart-disease cases and then removed noise introduced by oversampling and present in the original dataset. This combined approach produced a more balanced, cleaner, and better-separated feature space,

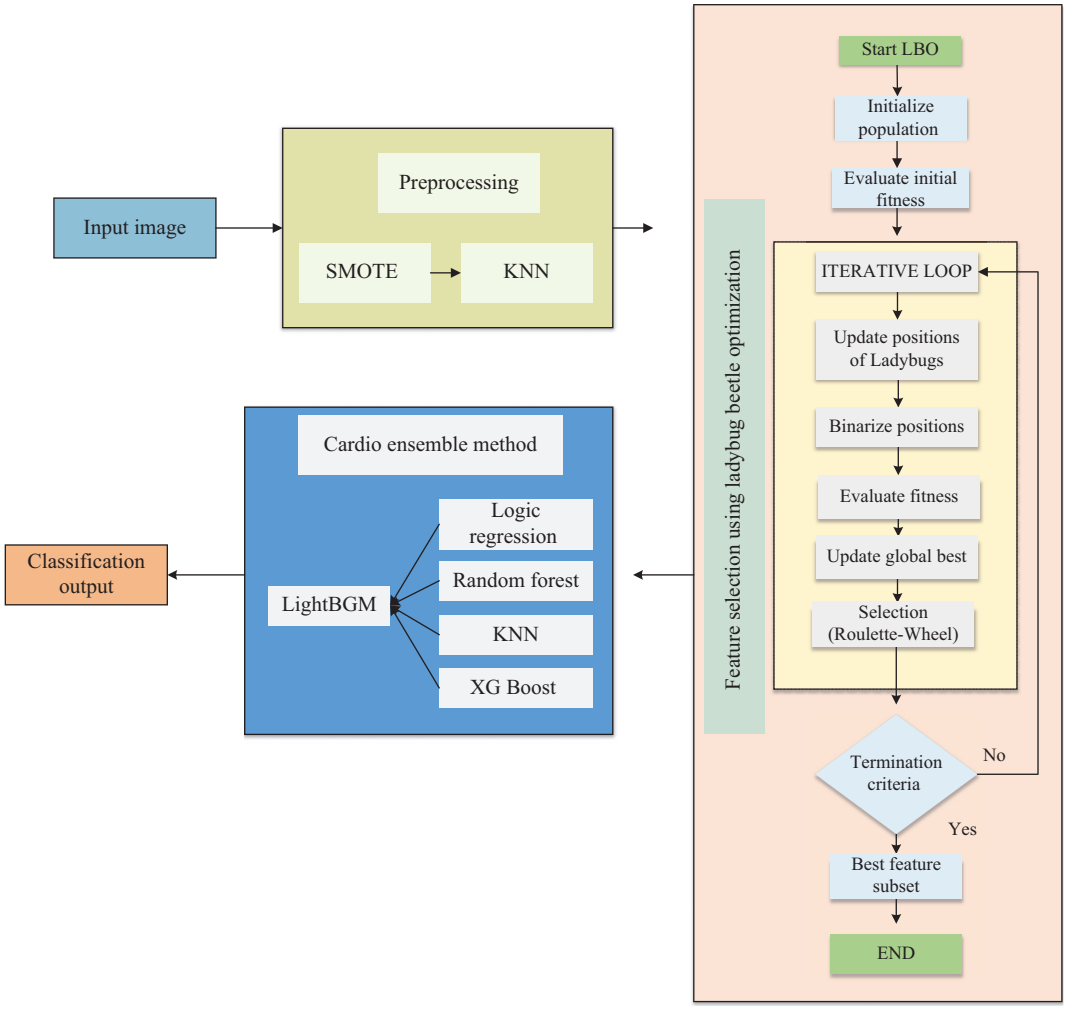


Figure 1 Overall Workflow of the Cardio-Sense Ensemble Framework.

and ultimately improved classification performance and decreased overfitting.

Feature Selection with the Ladybug Beetle Optimization Algorithm

Feature selection is crucial in heart disease prediction because population-based datasets such as BRFSS contain many correlated, redundant, and weakly informative attributes that can decrease model accuracy and increase computational burden. Conventional metaheuristic algorithms, such as the genetic algorithm (GA), particle swarm optimization (PSO), and ant colony optimization (ACO), often struggle with premature convergence, high parameter sensitivity, and difficulty in maintaining an effective balance between global exploration and local exploitation. These limitations frequently result in suboptimal or unstable feature subsets. To address these challenges, we used the LBO algorithm, because of its biologically inspired mechanisms, including attraction-repulsion behavior, Lévy-flight–based wide-range exploration, and adaptive movement strategies. LBO simulates how ladybugs alternate between broadly searching for food sources (exploration) and intensively exploiting promising regions (exploitation), thereby enabling efficient navigation through complex, multimodal feature-selection spaces. This ability is particularly beneficial for BRFSS, in which noisy and overlapping regions require robust global search to avoid local minima. By optimizing feature subsets that maximize classification accuracy, LBO yields a compact, discriminative set of predictors tailored to heart disease detection.

In the context of LBO for feature selection, each individual (ladybug) in the population is represented by a binary vector, wherein each bit corresponds to

a feature. A bit value of 1 indicates that the feature is selected, whereas a bit value of 0 indicates that the feature is not selected. The total number of features is denoted d, and the binary vector has a length of d, representing the selection status of all features. Let X_i^t denote the binary vector of the i^{th} ladybug at iteration t, where each element $x_{ij}^t \in \{0,1\}$ is either 0 or 1, indicating whether the j^{th} feature is selected or not.

Objective (Fitness) Function

The goal of feature selection is to identify a subset of features that maximizes predictive accuracy while minimizing the number of features used, thus increasing model efficiency. The objective function, $f(X_i^t)$, is designed to reflect this goal, and incorporates a weighted sum of the classification accuracy and the feature subset size:

$$f(X_i^t) = \omega \cdot (1 - Acc(X_i^t)) + (1-\omega) \cdot \frac{|X_i^t|_1}{d} \quad (2)$$

where $Acc(X_i^t)$ is the classification accuracy with the selected features, $|X_i^t|_1$ represents the number of selected features, and ω is a weighting factor between accuracy and simplicity. This fitness function encourages the algorithm to balance both high accuracy and minimum feature selection, and consequently ensures that the chosen features are both highly predictive and efficient.

Initialization phase: The optimization method starts by setting up a population of N ladybugs, in which each X_i^0 represents a candidate solution in the form of a binary vector. Each element $x_{ij}^0 \in \{0,1\}$ in the vector denotes the selection status of the j^{th} feature, where 1 indicates that the feature is chosen, and 0 denotes that the feature is not chosen. This binary representation corresponds to potential configurations of clinical features (e.g., age, blood pressure, or cholesterol).

Each individual in the population is evaluated with a fitness function $f(X_i^0)$, which measures the classification performance of an ML model. A starting population is created with a random N ladybugs:

$$X = \{X_1^0, X_2^0, \ldots, X_N^0\} \quad (3)$$

Iterative update mechanism: During each iteration, the position of each ladybug is updated by incorporation of three biologically inspired behavioral components. These vectors simulate the ladybugs' movement strategies in search of optimal solutions, thus mimicking the natural processes of food foraging and environmental interaction.

Each ladybug X_i^t updates its position according to the following three behavioral vectors.

The attraction vector A_{ij}^t models the influence of other ladybugs on the movement of the i^{th} ladybug, thereby guiding the population toward regions of higher fitness (social attraction).

The visibility vector V_{ij}^t represents the "visibility" or awareness of promising solutions in the current landscape, thus facilitating convergence.

The Lévy flight vector L_{ij}^t introduces stochastic exploratory behavior through long-range random steps, thereby enabling escape from local optima and promoting global exploration. The updated position of the i^{th} ladybug in the j^{th} dimension at iteration $t+1$ is calculated with the following equation:

$$Z_{ij}^{t+1} = X_{ij}^t + \alpha \cdot (A_{ij}^t + V_{ij}^t + L_{ij}^t) \quad (4)$$

where Z_{ij}^{t+1} is the updated continuous position before binarization, and X_{ij}^t denotes the current binary position. α denotes a step-size coefficient that controls the update magnitude, balancing exploration (via L) and exploitation (via A and V).

These vectors collectively enable the ladybugs to effectively navigate the search space by balancing the exploration of new regions with the exploitation of known high-quality solutions.

The real-value updates are converted to binary with a sigmoid transfer function:

$$S(Z_{ij}^{t+1}) = \frac{1}{1 + e^{-Z_{ij}^{t+1}}} \quad (5)$$

This process ensures that the updated positions of ladybugs remain binary, thereby representing feature selection decisions.

Fitness evaluation and global best update: After movement, each ladybug's new position is evaluated again. If the new solution offers better cost (e.g., higher predictive accuracy), it replaces the previous one. The global best solution representing the most accurate feature combination or parameter set yet found is continually updated throughout the iterations.

Selection via the roulette-wheel method: To determine which individuals are carried forward to the next generation, a roulette-wheel selection

Algorithm 1 LBO Algorithm.

Input:
Dataset D with d features
Population size N
Maximum iterations T_{max}
Step-size coefficient α
Weight ω (for fitness function)
Output:
Optimal feature subset X^*
Begin
Initialize population $X_i^0 \in \{0,1\}^d$ for $i = 1$ to N
Evaluate fitness $f(X_i^0)$ for all i
Set global best $X^* = argmin(f(X_i^0))$

For $t = 1$ to T_{max} do
For each ladybug $i = 1$ to N do
Compute attraction vector A_{ij}^t
Compute visibility vector V_{ij}^t
Compute Lévy flight step vector L_{ij}^t
Update continuous position:
$$Z_i^{t+1} = X_i^t + \alpha(A_i^t + V_i^t + L_i^t)$$
Apply sigmoid transfer function $S(Z_i^{t+1})$ with Eq. (5)
Generate binary solution X_i^{t+1}
End For
Evaluate $f(X_i^{t+1})$ for all i
Update global best X^* if $f(X_i^{t+1}) > f(X^*)$
Apply roulette-wheel selection to form next generation
End For
Return $X^*, f(X^*)$, $\mathcal{F}_{selected} = \{j \mid x_j^* = 1\}$
End

mechanism is used. This method assigns a selection probability to each ladybug according to its relative fitness, thereby favoring individuals with better performance while maintaining population diversity. The selection probability P_i for the i^{th} ladybug is computed with:

$$P_i = \frac{\rho_i}{\sum_{j=1}^{N} \rho_j} \tag{6}$$

where P_i represents the normalized probability of selection on the basis of relative fitness, and N is the total number of ladybugs in the population. This stochastic selection process ensures that individuals with higher relative fitness have a greater probability of being chosen; consequently, it encourages the propagation of superior solutions while allowing for the occasional selection of less-fit individuals to preserve diversity and avoid premature convergence.

Dynamic adaptation: LBOA optionally includes strategies such as adaptive population sizing, mutation, or local search refinements to enhance convergence speed and robustness. This aspect is particularly important in complex medical datasets, such as those used in heart disease diagnosis, in which overfitting and local minima are common.

Termination and output: The process repeats until the most iterations possible or function evaluations are reached. The process terminates by outputting the best solution discovered, which includes the most relevant features and/or the most effective parameter configuration for accurate heart disease prediction.

The LBO algorithm stops when any of the following subsequent circumstances are satisfied:

The maximum number of iterations $t = T_{max}$ is reached.

The maximum number of function evaluations $FE = FE_{max}$ is exceeded.

No significant improvement is found in the global best solution for κ consecutive iterations.

After the optimization process, the LBO algorithm yields the following outputs:

1. **Optimal feature subset**: The binary vector $X^* = [x_1^*, x_2^*, \ldots \ldots, x_d^*]$ represents the best solution found, wherein each element $x_j^* \in \{0,1\}$ indicates whether the j^{th} feature is selected $(x_j^* = 1$ or not $x_j^* = 1)$.
2. **Optimal fitness value**: The fitness of the best feature subset is given by:

$$f(X^*) = min\{f(X_i^t)\},\ \forall i \in [1,\ N],\ \forall t \in [1,\ T] \tag{7}$$

where T is the total number of iterations, and N is the population size.
3. **Reduced feature set**: The set of selected features can be extracted as:

$$\mathcal{F}_{selected} = \{j \mid x_j^* = 1\} \tag{8}$$

where $\mathcal{F}_{selected}$ denotes the indices of features included in the optimal subset. These outputs enable the construction of a simplified and high-performing model by reducing dimensionality while retaining predictive accuracy. Algorithm 1 shows the LBO algorithm.

CardioBoost Ensemble method

A key challenge in heart disease prediction is the substantial overlap in feature distributions between

healthy and affected individuals. Clinical indicators such as blood pressure, cholesterol, and resting heart rate often exhibit similar ranges across classes, thus hindering traditional classifiers from forming clear decision boundaries, and leading to elevated false positives and false negatives, which decrease model reliability in clinical settings. To overcome these limitations, the CBE framework integrated multiple baseline classifiers (LR, RF, KNN, and XGBoost) and uses LightGBM as the final classifier to enhance robustness and predictive accuracy.

K-nearest Neighbor

KNN was included as a simple distance-based classifier that makes predictions through majority voting among nearest neighbors. Because BRFSS features contain both continuous and categorical attributes with complex neighborhood structures, KNN helps evaluate how well proximity-based decision boundaries can capture underlying patterns. KNN accuracy is influenced by the choice of the parameter k, representing the number of neighbors considered. As k increases, the classification accuracy may improve. The class assignment is determined with a distance metric, typically the Euclidean distance:

$$d(x,\ y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{9}$$

where x and y are two feature vectors. The KNN algorithm assigns the class with the highest frequency among its k closest neighbors. The accuracy of KNN is extremely sensitive to the selection of k.

Random Forest Classifier

RF is a strong bagging-based tree ensemble. By training multiple DTs on bootstrapped subsets of samples and features, RF decreases variance and mitigates overfitting. Its inclusion allows for assessment of how well a diversity-driven ensemble handles the noisy and non-linear interactions commonly present in population-level health datasets. In RF, every tree produces a class prediction, and the model's forecast is the class that receives the most votes. The ultimate result is decided by majority voting. Given a dataset $=\{(x_i, y_i)\}_{i=1}^{N}$, RF constructs T trees $\{h_t(x)\}_{t=1}^{T}$. Mathematically, for a given input instance x, the prediction $\hat{y}$ is computed as:

$$\hat{y} = mode(h_1(x),\ h_2(x),\ \ldots\ \ldots,\ h_T(x)) \tag{10}$$

Where $h_t(x)$ denotes the prediction from the i^{th} DT, and T represents a forest's total number of trees. RF works particularly well on high-dimensional datasets, and it offers robust performance by balancing bias and variance. Its favorable generalizability makes it a reliable choice for medical prediction tasks such as heart disease classification.

Extreme Gradient Boosting

XGBoost, a powerful gradient-boosted tree algorithm, was selected for its robustness in modeling complex relationships through sequential error correction. Its regularization mechanisms and weighted tree growth make it ideal for benchmarking against our model in scenarios involving multi-collinearity and non-linear risk patterns. Its unique approach combines weak ML algorithms with agile learning and gradient boosting to decrease residual errors from previous predictions. The algorithm continuously refines its model by adjusting according to these errors. Additionally, XGBoost uses a regularized loss function to evaluate model performance, as represented by the equation:

$$\mathcal{L}(\varnothing) = \sum_{i=1}^{n} l\left(y_i,\ \hat{y}_i\right) + \sum_{k=1}^{K}\Omega\left(f_k\right) \tag{11}$$

where l is a differentiable loss function (e.g., logistic loss), $\hat{y}_i$ is the prediction from the ensemble, $\Omega(\mathrm{f}) = \gamma \mathrm{T} + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$ denotes the regularization term, T is the number of leaves in a tree, w_j is the score on leaf $\boldsymbol{j}$, and γ and λ are regularization parameters.

Logistic Regression

LR, a widely used supervised learning method for binary classification, models the probability of class membership with a linear decision boundary. By serving as the classical interpretable baseline, it allows for assessment of how well simple linear

relationships capture the underlying patterns in the dataset, in contrast to more complex non-linear models. It models the relationship between the input characteristics and the target variable with the logistic (sigmoid) function, which maps real-valued inputs into the range [0, 1] and consequently enables probabilistic interpretation. The logistic function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{12}$$

Where $z = w^T x + b$, x is the feature vector, w represents the weight vector, b represents the bias term, and $\sigma(z)$ represents the positive class's expected probability.

For reproducibility, the baseline classifiers were trained with their key hyperparameters. XGBoost was configured with learning_rate = 0.547, max_depth = 5, and n_estimators = 338. The RF model used max_depth = 15, max_features = auto, min_samples_leaf = 1, and n_estimators = 500. The DT used criterion = entropy, max_depth = 11, and splitter = best. KNN was set with n_neighbors = 5 and distance-based weighting to ensure balanced predictive behavior. The predicted probabilities from LR, RF, KNN, and XGBoost were concatenated as four additional meta-features and were used as input to the optimized LightGBM classifier, thus forming a stacking-based ensemble.

Light GBM

LightGBM was chosen as the final classifier in the CBE, because it efficiently handles large-scale, high-dimensional, and heterogeneous clinical data while maintaining strong predictive accuracy. In contrast to level-wise boosting methods such as XGBoost, LightGBM uses a leaf-wise growth strategy that selects splits producing the maximum reduction in loss, thus enabling deeper and more expressive trees. Consequently, its faster training, better optimization, and superior modeling of complex feature interactions make it highly suited for heart disease prediction.

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, f(x_i;\theta)) + \lambda \|\theta\|^2 \tag{13}$$

where θ represents model parameters, l is the loss function, and λ is a regularization term.

Results and Discussion

This section provides the results and details of evaluation of our model. The dataset is discussed in detail, and a representation of the setup used for testing is provided. Numerous experiments were conducted to determine the effectiveness of the developed method.

System Configuration

The following system setups are required to run the suggested model in Python 3.9. For basic computing, an Intel Core i5 processor is frequently used, and 16 GB of RAM is sufficient for performing data-intensive work and running Python. Nvidia GPUs can accelerate computations in Python frameworks such as TensorFlow and PyTorch, and therefore are highly beneficial for ML and DL applications. A 1 TB hard disk drive provides ample space for Python scripts, libraries, and datasets, while also offering sufficient read/write speeds to enhance overall system performance. Python and its libraries are fully supported on the widely used Windows 10 operating system.

Dataset Description

The dataset used in this study was obtained from the Centers for Disease Control and Prevention through the BRFSS 2020 survey and is publicly available as the Heart 2020 Cleaned dataset on Kaggle. The original BRFSS dataset contains nearly 300 health-related variables, from which the cleaned version retained 13 relevant attributes associated with cardiovascular health (Table 1). To enhance model efficiency, we used an optimization-based feature selection method to identify the seven most significant predictors (Table 2). The dataset comprised 1,243,000 samples, divided into 80% for training (475,908 class 0; 518,428 class 1) and 20% for testing (118,977 class 0; 129,607 class 1). The binary target variable HeartDisease indicated whether an individual was diagnosed with a heart condition. This balanced and large-scale dataset provided a reliable foundation for robust and generalizable heart disease prediction.

Table 1 Categorization of Features in the Dataset.

Category	Features
Demographic and physiological	Sex, AgeCategory, Race
Lifestyle and behavioral	Smoking, AlcoholDrinking, PhysicalActivity
Clinical and health condition	HeartDisease, Stroke, DiffWalking, Diabetic, GenHealth, Asthma, KidneyDisease, SkinCancer

Table 2 Final Optimized Features Selected for Model Development.

	Selected feature
1	HeartDisease
2	AlcoholDrinking
3	Stroke
4	MentalHealth
5	DiffWalking
6	Race
7	KidneyDisease

Performance Metrics

Definitions of the metrics used to evaluate performance indicators, including the F1 score, recall, accuracy, and precision, are provided below.

$$\text{accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}}$$

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{F1 score} = 2 \times \frac{\text{pecision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{specificity} = \frac{\text{number of true neagtives}}{\text{number of true negatives} + \text{number of false positives}}$$

Performance Evaluation

Preprocessing

Figure 2 presents the class distribution before and after the SMOTE-ENN preprocessing technique. Initially, the dataset was heavily imbalanced, with the majority class (label 0) comprising 90.7% and the minority class (label 1) representing only 9.3%

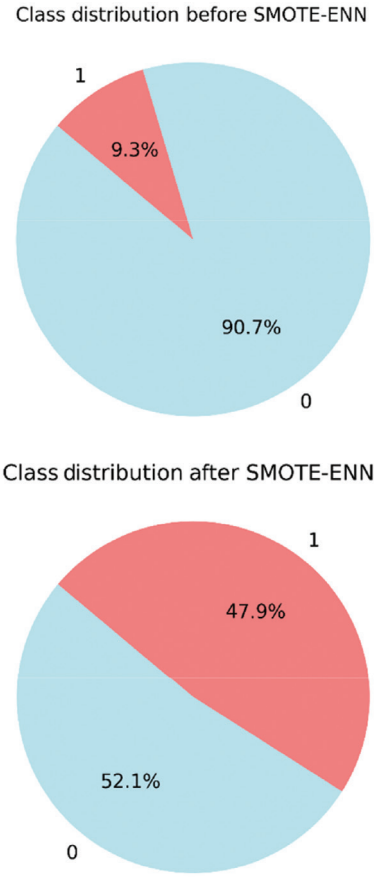


Figure 2 Preprocessing Output.

of samples. This imbalance would have caused classifiers to bias toward the majority class and would have often resulted in a high rate of false negatives. This issue is particularly serious in heart disease prediction, in which missed diagnoses have clinical consequences. After application of SMOTE-ENN, the right-side pie chart indicated a substantially improved class distribution, with 52.1% majority and 47.9% minority samples. SMOTE increased minority representation through synthetic sample generation, and ENN removed noisy and overlapping instances, thereby producing a cleaner and more balanced dataset. This refined distribution provided classifiers with more representative training data, thus leading to better decision boundaries,

improved sensitivity, and decreased misclassification of true heart-disease cases.

Feature Selection

During the optimization process, each solution (ladybug) was represented as a binary vector, wherein a feature that was selected was denoted by 1, and a feature that was not selected was denoted by 0 (Figure 3). The fitness of each solution was evaluated with a hybrid objective function balancing classification accuracy (evaluated via a LightGBM classifier) and the number of selected features, thereby ensuring both performance and parsimony. After convergence of the LBBO algorithm, the optimal feature subset was passed to the LightGBM model, which underwent hyperparameter tuning with a grid or randomized search strategy embedded in the LBOA cycle.

The best hyperparameters obtained during this optimization phase were as follows: colsample_bytree: 1.0, learning_rate: 0.3, max_depth: 15, min_child_samples: 100, n_estimators: 1000, num_leaves: 150, reg_alpha: 0.0, reg_lambda: 0.0, subsample: 1.0. These hyperparameters represented the optimal configuration found by LBO to obtain the best categorization results with the chosen features. The tuned model demonstrated superior predictive power, thus validating the effectiveness of LBO in both feature selection and hyperparameter tuning. By decreasing irrelevant and redundant features while optimizing model parameters, the LBO-based feature selection framework enhanced diagnostic accuracy and computational efficiency, and therefore is suitable for practical deployment in heart disease prediction systems.

```
Best Parameters:
OrderedDict([
('colsample_bytree', 1.0),
('learning_rate', 0.299),
('max_depth', 15),
('min_child_samples', 100),
('n_estimators', 1000),
('num_leaves', 150),
('reg_alpha', 0.0),
('reg_lambda', 0.0),
('subsample', 1.0)
])
```

Figure 3 Feature Selection with LBBO.

Figure 4 compares the convergence behavior of four optimization algorithms, LBO, PSO, ACO, and GA, over 50 iterations. The LBO algorithm achieved the highest accuracy and the fastest convergence rate. This superior performance arose from LBO's adaptive attraction-repulsion mechanism, which effectively balanced exploration and exploitation, and consequently was able to escape local optima and identify the most discriminative feature subset more efficiently than conventional population-based algorithms such as PSO, ACO, and GA.

Classification Report

Confusion matrices comparing the performance of several baseline and ensemble classifiers (Figure 5) highlighted the clear superiority of the LightGBM model. LightGBM achieved the most balanced and accurate predictions, with exceptionally low misclassification (only 1184 false positives and 2209 false negatives), thus indicating strong sensitivity and specificity. This performance was driven by LightGBM's leaf-wise tree growth, which enabled

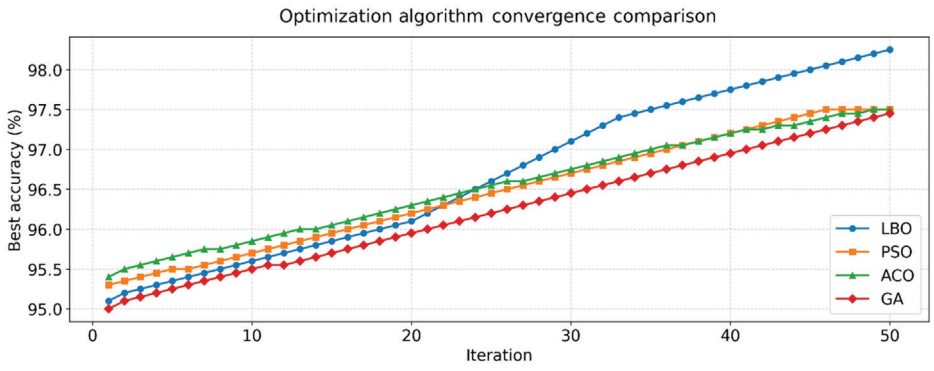


Figure 4 Optimization Algorithm Convergence Comparison.

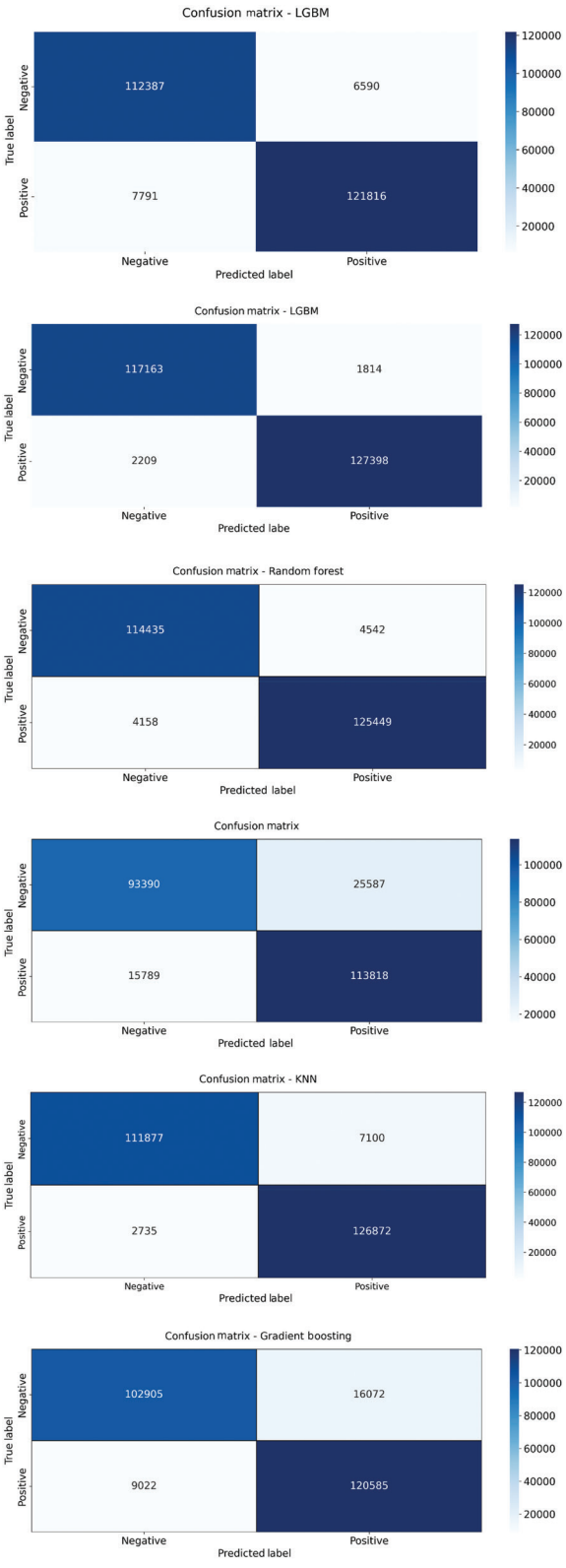


Figure 5 Confusion Matrix.

deeper, more discriminative splits and allowed the model to capture subtle nonlinear patterns present in heart-disease features. RF also performed well but showed slightly higher misclassification, thereby illustrating the limitations of variance-reduction approaches in handling overlapping clinical attributes. Gradient Boosting exhibited further degradation, reflected by a noticeable increase in false predictions, thus suggesting weaker adaptability to noisy or borderline cases. XGBoost demonstrated the highest misclassification rates among ensemble methods, given its difficulty in clearly separating closely clustered health-risk profiles in this dataset. Baseline models such as LR and KNN showed moderate results but still had high errors due to restrictions in linear separability for LR and sensitivity to local neighborhood fluctuations for KNN. Overall, the results confirmed that boosting-based ensembles, particularly LightGBM, provided the most reliable and clinically meaningful classification performance for heart-disease prediction.

Figure 6 illustrates the ROC curve of our tuned LightGBM-based CSEF for heart disease prediction. The model achieved an AUC of 0.98, indicating near-perfect discrimination between positive and negative cases. This performance reflected the effectiveness of the optimized boosting framework in capturing complex relationships among clinical and behavioral risk factors. The ROC curve confirmed that the model was well suited for reliable decision support in heart disease risk assessment, in which minimizing false negatives is clinically crucial.

Explainable AI Analysis Using SHAP

To ensure the transparency and interpretability of our CBE model, we applied SHAP to analyze how different features influenced model predictions. SHAP provides both global and local explanations quantifying the extent and direction of each feature's contribution to the outcome. This approach helps validate the model's reasoning process and supports its reliability for clinical decision-making.

The SHAP summary analysis (Figure 7) highlighted that Race had the greatest overall influence on prediction outcomes and was followed by MentalHealth and DiffWalking. Higher feature values of these variables were strongly associated with higher predicted risk, whereas lower values

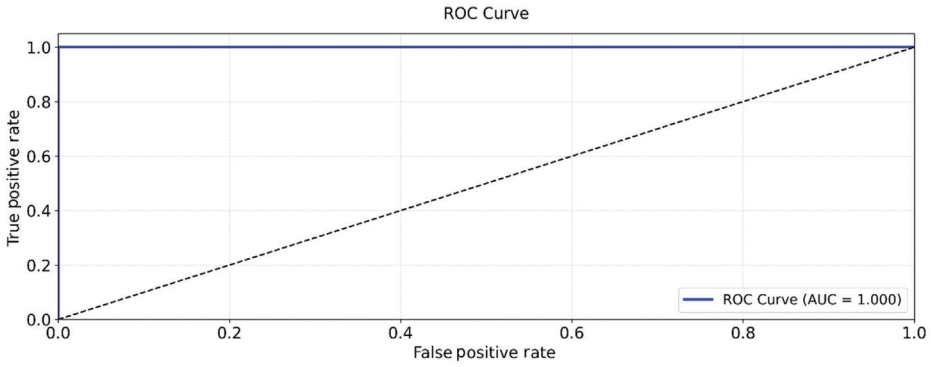


Figure 6 ROC.

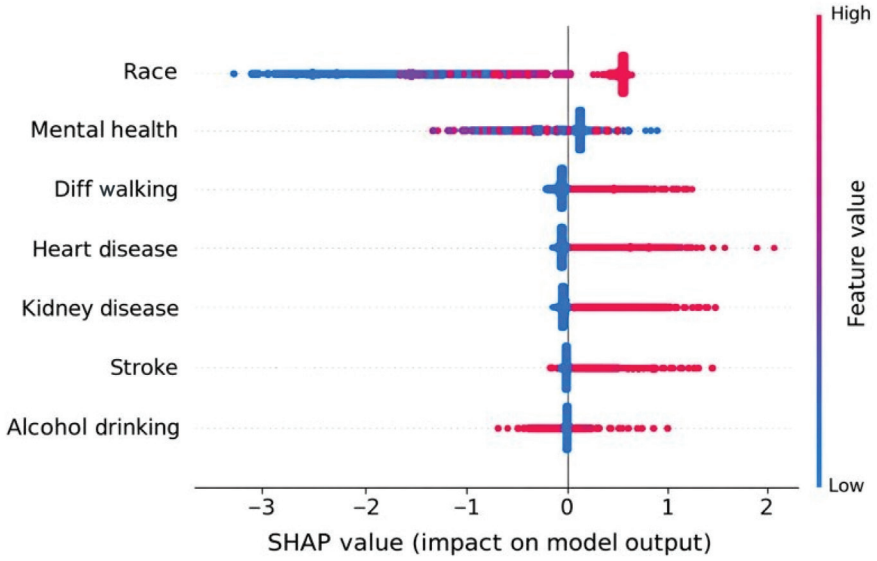


Figure 7 SHAP Summary Plot Indicating Global Feature Influence.

tended to be associated with lower likelihood of the outcome. The distribution indicated that the model effectively captured the nonlinear relationships between clinical and lifestyle factors contributing to heart disease prediction.

Feature importance analysis (Figure 8) confirmed that Race was the most dominant predictor and had the largest average effect on the model's output. MentalHealth, DiffWalking, and HeartDisease also substantially contributed to the predictive process, whereas KidneyDisease, Stroke, and AlcoholDrinking exhibited comparatively minor influences. This ranking validated that the model prioritized the most clinically relevant features in assessing cardiovascular risk.

The local explanation (Figure 9) illustrated how individual feature values interacted in producing a specific prediction for one instance. In this case, pre-existing HeartDisease, Race, and DiffWalking conditions increased the predicted risk, whereas good MentalHealth and absence of Stroke lowered the predicted risk. This interpretation confirmed that the model's prediction aligned with established clinical reasoning and demonstrated reliable interpretability at the individual level.

Comparative Analysis

To comprehensively evaluate the performance of various classification models, we present two bar charts comparing six models (LGBM, LR, RF, KNN, GB, and our tuned LGBM model) across the key metrics of accuracy, precision, recall, F1 score, specificity, and AUC. This comparison highlighted the relative strengths of ensemble-based models against traditional classifiers.

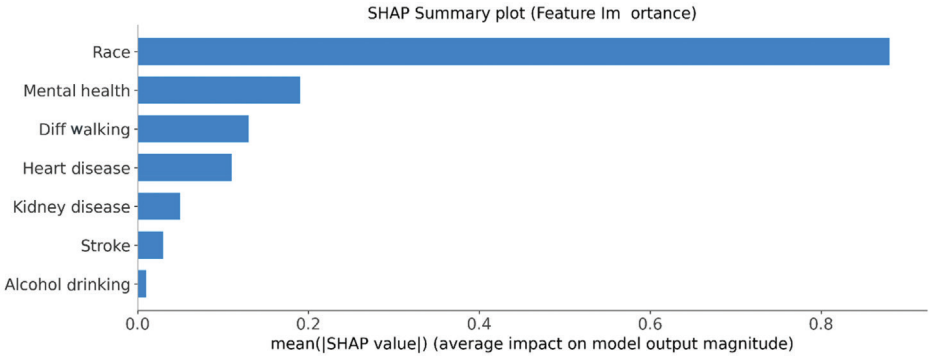


Figure 8 SHAP Bar Plot Indicating Feature Importance Ranking.

Figure 9 SHAP Force Plot Providing Local Instance Explanation.

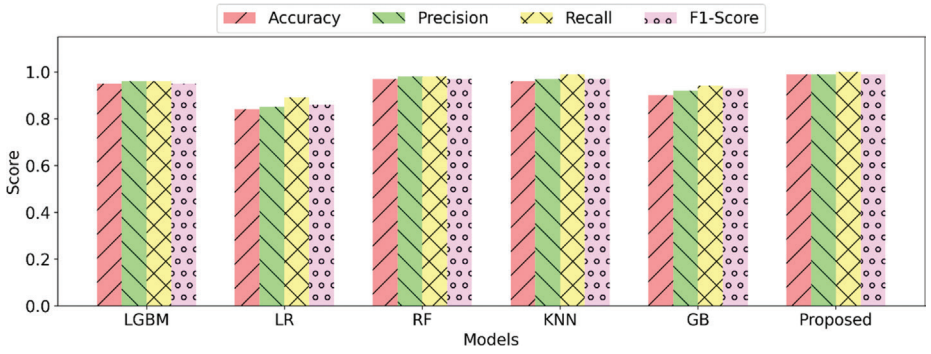


Figure 10 Multi-metric Classification Performance Comparison.

Our model achieved near-perfect performance, with all major metrics reaching 0.98, thus demonstrating highly balanced detection of both classes (Figure 10). RF and KNN followed, with strong but lower scores (0.96). The superior performance of our model arose from use of the optimized LightGBM classifier and LBO-driven feature refinement, which together enhanced the discriminative power and decreased misclassification in complex heart-disease patterns.

Our model achieved the highest AUC and specificity (both 0.98), reflecting excellent class separation and minimal false positives (Figure 11). RF and KNN followed, with values of 0.96, whereas LR had the weakest performance (0.83). Our model's superiority stemmed from LightGBM's leaf-wise boosting strategy, which created deeper, more discriminative splits and captured subtle clinical patterns missed by other models.

A comparative analysis (Table 3) demonstrated the clear superiority of our tuned LightGBM model, which achieved the highest overall performance with accuracy, precision, recall, F1 score, specificity, and AUC all at 0.98. This exceptional result arose from LightGBM's leaf-wise tree growth mechanism, which enabled deeper optimization of decision boundaries and efficient handling of complex, high-dimensional clinical data. In comparison, RF and KNN performed strongly, with balanced metrics of 0.96, benefiting from ensemble averaging and neighborhood-based decision rules but lacking the adaptive boosting capability of LightGBM. The baseline LGBM (0.94) exhibited solid but lower

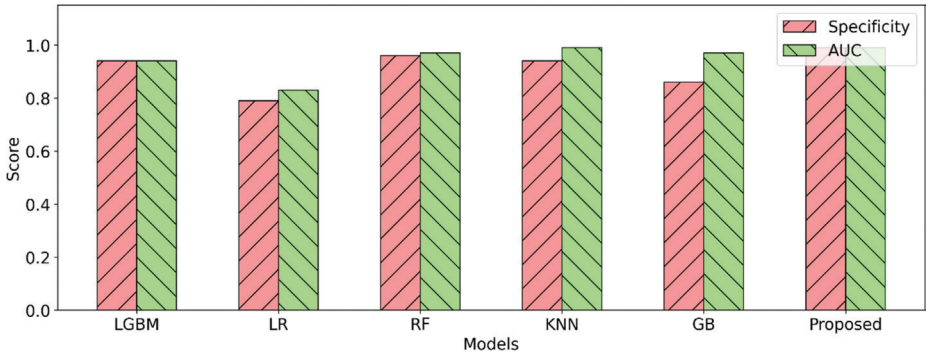


Figure 11 Specificity and AUC Comparison.

Table 3 Performance Comparison Table.

Methods	Accuracy	Precision	Recall	F1 score	Specificity	AUC
LGBM	0.94	0.94	0.94	0.94	0.94	0.94
LR	0.83	0.86	0.78	0.82	0.83	0.83
RF	0.96	0.96	0.96	0.96	0.96	0.96
KNN	0.96	0.95	0.96	0.95	0.96	0.96
GB	0.90	0.91	0.90	0.90	0.90	0.90
Developed model	0.98	0.98	0.98	0.98	0.98	0.98

performance, because of the absence of tuning optimizations, whereas GB achieved moderate accuracy (0.90) with slower convergence. LR lagged significantly (0.83 accuracy), because of its linear nature and limited ability to capture nonlinear feature interactions. Overall, our tuned LightGBM demonstrated superior accuracy, generalization, and robustness.

Our model maintained consistently high performance in both test and validation phases, achieving an accuracy of 0.9850 and 0.9815, with AUC values of 0.9870 and 0.9850, respectively (Table 4). This robustness resulted from the use of the LBO algorithm, which efficiently balanced exploration and exploitation in identifying the most discriminative features while avoiding local optima. Consequently, the model sustained strong generalization and stability across datasets with minimal performance variation.

Statistical Significance Analysis

To ensure the reliability and statistical validity of the performance differences among classifiers, we performed DeLong's test and McNemar's test to evaluate whether the observed improvements in the AUC and classification accuracy of our model over baseline methods were statistically significant.

Table 4 Test and Validation Performance of our Model.

Metric	Test result	Validation result
Accuracy	0.9850	0.9825
Precision	0.9860	0.9838
Recall	0.9840	0.9812
F1 score	0.9850	0.9828
Specificity	0.9855	0.9830
AUC	0.9870	0.9848

Table 5 presents a statistical comparison between our model and other baseline classifiers with DeLong's test and McNemar's test. The 95% confidence interval for LGBM's AUC ranged from 0.975 to 0.992, thus indicating high discriminatory power. Lower P-values (<0.05) for SVM, LR, and KNN suggested statistically significant performance differences with respect to our model, thereby confirming our model's superiority. In contrast, RF and GB exhibited non-significant differences reflecting comparable but slightly lower predictive consistency.

Ablation Study

To evaluate the effectiveness of each component in our heart disease prediction framework, we conducted an ablation conducted by systematically

Table 5 Statistical Comparison of Classification Models with DeLong's and McNemar's Tests.

Model	95% Confidence interval (AUC)	P-value (vs. Developed model, DeLong test)	McNemar P-value (vs. developed model)
Developed model	[0.975-0.992]	-	-
SVM	[0.963-0.986]	0.041	0.032
LR	[0.961-0.984]	0.027	0.026
RF	[0.972-0.990]	0.360	0.110
KNN	[0.951-0.980]	0.015	0.008
GB	[0.973-0.991]	0.480	0.190

Table 6 Ablation Study of our Heart Disease Prediction Framework.

Configuration	Accuracy	Precision	Recall	F1 score	AUC
Baseline models					
Raw data (no preprocessing)	0.84	0.81	0.79	0.80	0.83
Preprocessing effect					
Only SMOTE	0.89	0.87	0.90	0.88	0.90
SMOTE + ENN	0.93	0.92	0.93	0.92	0.94
Feature selection effect					
SMOTE + ENN + all features (no LBO)	0.91	0.90	0.89	0.89	0.91
SMOTE + ENN + LBO	0.95	0.95	0.94	0.94	0.96
Classifier variation					
SMOTE + ENN + LBO + logistic regression	0.83	0.84	0.78	0.81	0.83
Final integrated model (SMOTE + ENN + LBO + LightGBM)	0.98	0.98	0.98	0.98	0.98

excluding or altering individual modules and analyzing the effects on model performance. The full pipeline included SMOTE + ENN preprocessing, LBO for feature selection, and the CBE classifier with a tuned LightGBM model. Key metrics such as accuracy, precision, recall, F1 score, and AUC were used to compare different configurations (Table 6).

Each component of the framework significantly contributed to overall model performance (Table 6). The raw baseline showed limited accuracy (0.84), because of class imbalance and noisy samples. Incorporation of SMOTE + ENN substantially improved balance and generalization, and LBO feature selection enhanced discriminative learning by removing redundant attributes. Replacing traditional classifiers with LightGBM within the CBE further increased the accuracy to 0.98, owing to its leaf-wise growth and optimal feature interaction handling. These results confirmed that each module played a critical and complementary role in achieving the final model's superior predictive performance.

K-fold Cross-validation Analysis

To evaluate the robustness and consistency of our model, we used a five-fold CV procedure. The dataset was randomly divided into five equal subsets. During each iteration, four folds were used for training, and the remaining fold was used for testing. The overall performance was obtained by averaging the results across all folds to minimize bias from data partitioning.

Our model consistently performed well across all five folds, achieving a mean accuracy of 0.9850 and AUC of 0.9870 (Table 7). The results are reported as mean ± standard deviation across folds. The low standard deviations across all metrics indicated consistent and stable performance, and demonstrated the strong generalizability of our framework. The uniformity of results across folds demonstrated the model's strong generalization ability and stability. This consistency was attributed primarily to the efficient feature optimization by the LBO algorithm and the robust ensemble structure of the CardioBoost

model, thus ensuring reliable and reproducible predictions across varying data partitions.

Model Evaluation on an External Dataset

To evaluate the model's generalization on unseen data, we used the Heart Disease Dataset [35] as an external validation source. This assessment demonstrated our model's robustness and consistent predictive capability across independent clinical datasets.

Our model maintained outstanding performance on the Heart Disease Dataset, thus confirming its strong generalizability (Figure 12). The model achieved an accuracy of 0.9903, precision and specificity of 1.0000, and an AUC of 0.9996, indicating near-perfect classification. Although the recall (0.9810) was slightly lower, the high F1 score (0.9904) demonstrated a balanced precision-recall trade-off. This performance resulted from the LBO algorithm's adaptive exploration-exploitation strategy, which effectively identified the most discriminative clinical features, and the LightGBM's leaf-wise tree growth mechanism, which optimized complex decision boundaries for precise classification.

The model's exceptional discriminative ability was demonstrated by a ROC curve closely following the top-left boundary and an AUC of 0.9996 (Figure 13). This near-perfect result confirmed the model's superior balance between sensitivity and specificity in effectively distinguishing positive and negative cases across varying thresholds.

Discussion

The comprehensive evaluation, combining comparative analysis and ablation study, demonstrated the superior performance and clinical relevance of our CSEF for heart disease prediction. The tuned LightGBM-based model achieved near-perfect scores (0.98) across all metrics, accuracy, precision, recall, F1 score, specificity, and an AUC outperforming those of baseline classifiers (LR, GB, RF, KNN, and untuned LGBM) and recent studies in the literature. For example, a previously reported ensemble model [27] achieved an AUC of 0.951 on the PTB-ECG dataset and 0.950 on MIT-BIH, whereas another model [29] achieved an accuracy of 0.9815 with an extra tree-based ensemble. In comparison, our model not only improved discriminative

Table 7 Five-fold Cross-validation Results of our Model.

Fold	Accuracy	Precision	Recall	F1 score	Specificity	AUC
Fold 1	0.9846	0.9854	0.9837	0.9845	0.9848	0.9866
Fold 2	0.9852	0.9860	0.9841	0.9850	0.9855	0.9870
Fold 3	0.9848	0.9855	0.9839	0.9847	0.9850	0.9868
Fold 4	0.9854	0.9862	0.9844	0.9852	0.9857	0.9872
Fold 5	0.9850	0.9859	0.9840	0.9849	0.9853	0.9869
Mean	0.9850 ± 0.0003	0.9858 ± 0.0003	0.9840 ± 0.0003	0.9849 ± 0.0003	0.9853 ± 0.0003	0.9870 ± 0.0002

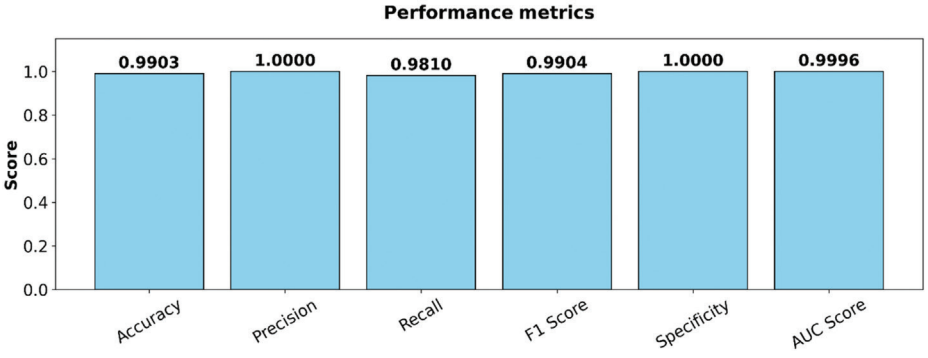


Figure 12 Performance Metrics of the Developed Model.

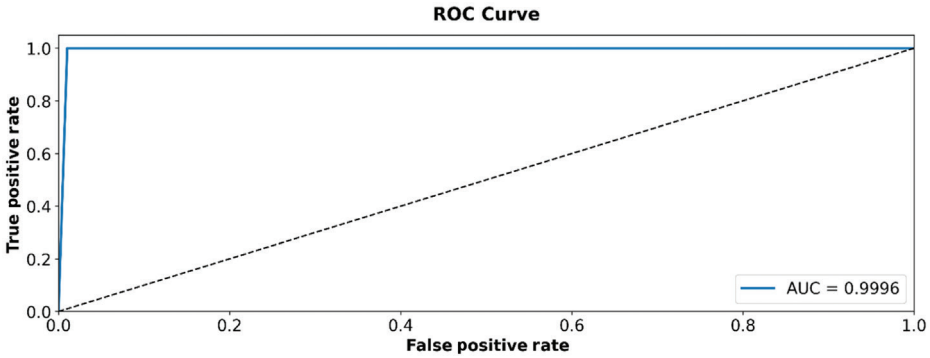


Figure 13 ROC Curve Analysis of the Developed Model on the Heart Disease Dataset.

performance but also decreased the false negative rate, thus indicating a substantial decrease in missed diagnoses and a clinically meaningful decrease in potential adverse cardiac events.

Furthermore, compared with recent deep learning approaches, such as the hybrid convolutional neural network–grasshopper optimization (CNN-GHO) model [36] and the CNN-flower pollination optimization/teaching-learning-based optimization-genetic algorithm (CNN-FPO/TLBO-GA) models [37], our CSEF demonstrated superior predictive performance, achieving a higher accuracy (0.98) and F1 score (0.98). Unlike CNN-based methods, which often require extensive computational resources and provide limited interpretability, CSEF offers clear feature importance via SHAP analysis, and consequently enables clinicians to understand the contributions of key risk factors, such as race, mental health, and physical activity, to predicted outcomes, thereby enhancing clinical relevance and deployability.

The ablation study highlighted the importance of each module. SMOTE + ENN effectively balanced classes and removed noisy samples, and LBO-based feature selection eliminated redundant features and minimized overfitting. LR, even with all modules, achieved lower performance (accuracy 0.83), because of its linear assumptions, thus underscoring the need for ensemble learning to capture complex cardiovascular patterns. From a practical perspective, our framework is suitable for early screening in community hospitals, and should support timely intervention and patient triage, as well as detailed risk assessment in tertiary care centers. Its inference time of 1.6 seconds and reasonable hardware requirements (Intel Core i5, 16 GB RAM) support its feasibility of integration into electronic medical record systems, to enable near real-time decision support. Overall, our model demonstrated not only statistical superiority over existing studies but also clear clinical relevance and deployability, thereby bridging the gap between predictive performance and real-world utility.

The CSEF has excellent predictive performance but nonetheless has several limitations, such as a high dependence on the availability of high-quality, well-annotated datasets across various healthcare institutions, which affects generalization performance. Additionally, although SHAP analysis enhances interpretability, ensemble, and boosting models, it exhibits “black-box” behavior preventing full clinical trust in automated decisions. In large-scale health datasets, issues including data privacy, transparency, and potential demographic or socioeconomic bias must be managed in real-world deployment. Future research must focus on integrating privacy-preserving learning mechanisms and bias-mitigation strategies validated on multi-institutional and real-time clinical datasets, to ensure wide reliability and regulatory acceptance.

Conclusion

In conclusion, the CSEF provides a robust and computationally effective solution to problems in heart disease prediction. By integrating advanced preprocessing techniques such as SMOTE and ENN, the framework effectively addresses class imbalance, missing values, and noise, while ensuring high data quality. Use of the LBO algorithm

for feature selection markedly decreases overfitting and enhances model performance. The CBE method, culminating in the LightGBM, achieves outstanding predictive accuracy, with metrics of 0.98 across accuracy, precision, recall, F1 score, specificity, and AUC. Our model outperformed existing models and therefore would provide healthcare providers with a reliable framework for timely interventions. This study advances the field of CVD prediction and paves the way to more efficient, scalable solutions in healthcare. Future research will explore the integration of deep learning for advanced feature extraction, real-time patient monitoring, multi-modal clinical data fusion, and extension of the framework to predict multiple cardiovascular and comorbid conditions, thus further advancing scalable and clinically meaningful predictive solutions.

Data Availability Statement

The dataset used in this study is publicly available from the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance System (BRFSS) at https://www.kaggle.com/datasets/luyezhang/heart-2020-cleaned.

Ethics Statement

No direct interactions with human or animal subjects were involved. Therefore, ethical approval and informed consent were not required.

Author Contributions

The authors confirm contribution to the article as follows: all authors reviewed the results and approved the final version of the manuscript.

Funding/Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors thank the Deanship of Sandip University, Nashik, for institutional support and encouragement in conducting this work.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Consent to Participation

This study did not involve human participants; therefore, informed consent to participate was not required.

Consent to Publication

All contributors agreed and provided consent to publish.

Competing Interests

None.

REFERENCES

1. Ahmad GN, Ullah S, Algethami A, Fatima H, Akhter SMH. Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection. IEEE Access 2022;10:23808–28.
2. Almazroi AA, Aldhahri EA, Bashir S, Ashfaq S. A clinical decision support system for heart disease prediction using deep learning. IEEE Access 2023;11:61646–59.
3. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023;118(17):3272–87.
4. Manikandan G, Pragadeesh B, Manojkumar V, Karthikeyan AL, Manikandan R, Gandomi AH. Classification models combined with Boruta feature selection for heart disease prediction. Inform Med Unlocked 2024;44:101442.
5. Poznyak AV, Sadykhov NK, Kartuesov AG, Borisov EE, Melnichenko AA, Grechko AV, et al. Hypertension as a risk factor for atherosclerosis: cardiovascular risk assessment. Front Cardiovasc Med 2022;9:959285.
6. Hossain MI, Maruf MH, Khan MAR, Prity FS, Fatema S, Ejaz MS, et al. Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. Iran J Comput Sci 2023;6(4):397–417.

7. Naija A, Yalcin HC. Evaluation of cadmium and mercury on cardiovascular and neurological systems: effects on humans and fish. Toxicol Rep 2023;10:498–508.
8. Swathy M, Saruladha K. A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. ICT Express 2022;8(1):109–16.
9. Narkhede M, Pardeshi A, Bhagat R, Dharme G. Review on emerging therapeutic strategies for managing cardiovascular disease. Curr Cardiol Rev 2024;20(4):86–100.
10. Karki A, Manandhar L. Understanding ischemic stroke: symptoms, risks, and the importance of timely intervention. Am J Patient Health Info 2024;1(2).
11. Bonham PA, Droste LR, González A, Kelechi TJ, Ratliff CR. 2024 Guideline for management of wounds in patients with lower extremity arterial disease: an executive summary. J Wound Ostomy Continence Nurs 2024;51(5):357–70.
12. Qadri AM, Raza A, Munir K, Almutairi MS. Effective feature engineering technique for heart disease prediction with machine learning. IEEE Access 2023;11:56214–24.
13. Dhakal B, Pokharel B. Coronary artery disease (CAD) demystified-causes, symptoms & treatment. Am J Patient Health Info 2024;1(1).
14. Upadhyay RK. High cholesterol disorders, myocardial infarction and its therapeutics. World J Cardiovasc Dis 2023;13(8):433–69.
15. Wang Y, Li G, Yang L, Luo R, Guo G. Development of innovative biomaterials and devices for the treatment of cardiovascular diseases. Adv Mater 2022;34(46):e2201971.
16. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. Algorithms 2023;16(2):88.
17. Ramesh TR, Lilhore UK, Poongodi M, Simaiya S, Kaur A, Hamdi M. Predictive analysis of heart diseases with machine learning approaches. Malays J Comput Sci 2022;2022(1):132–48.
18. Khan A, Qureshi M, Daniyal M, Tawiah K. A novel study on machine learning algorithm-based cardiovascular disease prediction. Health Soc Care Community 2023;2023(1):1406060.
19. Guazzi M, Wilhelm M, Halle M, Van Craenenbroeck E, Kemps H, de Boer RA, et al. Exercise testing in heart failure with preserved ejection fraction: an appraisal through diagnosis, pathophysiology and therapy–A clinical consensus statement of the Heart Failure Association and European Association of Preventive Cardiology of the European Society of Cardiology. Eur J Heart Fail 2022;24(8):1327–45.
20. Rashid Y, Bhat JI. Topological to deep learning era for identifying influencers in online social networks: a systematic review. Multimed Tools Appl 2024;83(5):14671–714.
21. Arumugam K, Naved M, Shinde PP, Leiva-Chauca O, Huaman-Osorio A, Gonzales-Yanac T. Multiple disease prediction using Machine learning algorithms. Mater Today Proc 2023;80:3682–5.
22. Ahsan MM, Siddique Z. Machine learning-based heart disease diagnosis: a systematic literature review. Artif Intell Med 2022;128:102289.
23. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biol 2025;79:103470.
24. Baghdadi NA, Farghaly Abdelaliem SM, Malki A, Gad I, Ewis A, Atlam E. Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data 2023;10(1):144.
25. Naser MA, Majeed AA, Alsabah M, Al-Shaikhli TR, Kaky KM. A review of machine learning's role in cardiovascular disease prediction: recent advances and future challenges. Algorithms 2024;17(2):78.
26. El-Sofany HF. Predicting heart diseases using machine learning and different data classification techniques. IEEE Access 2024;12:106146–60.
27. Rath A, Mishra D, Panda G. Imbalanced ECG signal-based heart disease classification using ensemble machine learning technique. Front Big Data 2022;5:1021518.
28. Kumar A, Singh KU, Kumar M. A clinical data analysis based diagnostic system for heart disease prediction using ensemble method. Big Data Min Anal 2023;6(4):513–25.
29. Asif D, Bibi M, Arif MS, Mukheimer A. Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization. Algorithms 2023;16(6):308.
30. Ghasemieh A, Lloyed A, Bahrami P, Vajar P, Kashef R. A novel machine learning model with stacking ensemble learner for predicting emergency readmission of heart-disease patients. Decis Anal J 2023;7:100242.
31. Nissa N, Jamwal S, Neshat M. A technical comparative heart disease prediction framework using boosting ensemble techniques. Computation 2024;12(1):15.
32. Mondal S, Maity R, Omo Y, Ghosh S, Nag A. An efficient computational risk prediction model of heart diseases based on dual-stage stacked machine learning approaches. IEEE Access 2024;12:7255–70.
33. Sultan SQ, Javaid N, Alrajeh N, Aslam M. Machine learning-based stacking ensemble model for prediction of heart disease with explainable AI and K-fold cross-validation: a symmetric approach. Symmetry 2025;17(2):185.
34. Ahmed M, Sulaiman MH, Hassan MM, Bhuiyan T. Predicting the classification of heart failure patients using optimized machine learning algorithms. IEEE Access 2025:13.
35. Chaurasia V, Chaurasia A. Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique. Biomed Mater Devices 2023;1(2):932–41.
36. Polepaka S, Ram Kumar RP, Palakurthy D, Manasa V, Saritha A, Dixit S, et al. Optimized convolutional neural network using grasshopper optimization technique for enhanced heart disease prediction. Cogent Eng 2024;11(1):2423847.
37. Ram Kumar RP, Raju S, Annapoorna E, Hajari M, Hareesa K, Vatin NI, et al. Enhanced heart disease prediction through hybrid CNN-TLBO-GA optimization: a comparative study with conventional CNN and optimized CNN using FPO algorithm. Cogent Eng 2024;11(1):2384657.