

Article

Chaos-Driven Evolutionary Optimization of a Hybrid Machine Learning Model for Accurate Heart Disease Prediction

Zhiyi Song ^{1,†} , Jiaen Lin ^{1,†}, Aoqi Wang ^{1,†}, Yunlong Liu ^{1,*}, Menggen Li ¹, Guo Chen ², Zheng Xiang ¹ and Yuhang Zhang ¹

¹ School of Automation, Beijing Information Science and Technology University, Beijing 102206, China; 2023010772@bistu.edu.cn (Z.S.); 2023010786@bistu.edu.cn (J.L.); 2023010784@bistu.edu.cn (A.W.); 15737006525@163.com (M.L.); 2025010802@bistu.edu.cn (Z.X.); 2025010688@bistu.edu.cn (Y.Z.)

² College of Earth Science and Technology, Southwest Petroleum University, Chengdu 610500, China; 202331025207@stu.swpu.edu.cn

* Correspondence: liuyunlongbuaa@163.com

† These authors contributed equally to this work.

Abstract

In this paper, a hybrid machine learning approach is proposed for accurate prediction of heart disease. The features are categorized into numerical and categorical types, which are extracted independently by a deep neural network (DNN) and CatBoost, respectively, before being fused via logistic regression to produce the output; simultaneously, an improved chaotic evolutionary algorithm (CEO) is employed to optimise hyperparameters, and adaptive chaotic perturbations and a restart mechanism are introduced to enhance global optimisation capabilities. The integration of these components facilitates the successful development of a hybrid model, referred to as the CEO-Enhanced Hybrid CatBoost and Deep Learning Model (CEO-HCDLM), for effective heart disease classification. To assess model stability and generalizability, multi-seed validation, Gaussian noise injection, and cross-dataset validation using the Kaggle heart disease dataset were conducted. The results demonstrate the efficacy and practicality of this approach in supporting decision-making for cardiovascular risk assessment.

Keywords: heart disease prediction; machine learning; CatBoost; deep neural network; chaotic evolution optimization; clinical decision support

1. Introduction

Cardiovascular diseases continue to represent one of the leading causes of mortality worldwide, constituting a major public health challenge in the 21st century [1]. The growing burden of these conditions has elevated their significance on the global health agenda. According to estimates by the World Health Organization (WHO), approximately 17.7 million deaths in 2015 were attributed to cardiovascular diseases, accounting for 31% of all global deaths. Of these, an estimated 7.4 million deaths were due to coronary heart disease, with the majority of the disease burden concentrated in low- and middle-income countries [2]. Furthermore, the global economic cost associated with cardiovascular diseases has surpassed USD 100 billion annually, imposing substantial strain on healthcare systems worldwide [3].

Specific cardiac conditions, such as valvular heart disease, have a prevalence of 5–10% among individuals aged 65 to 74 years, increasing to 10–20% in those aged 75 years and older. The prevalence of these conditions is rising rapidly in high-income countries [4]. In China, a major barrier to the effective management and treatment of congenital heart



Academic Editors: George Obaido and Chidozie Williams Chukwu

Received: 1 June 2026

Revised: 17 July 2026

Accepted: 19 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

disease (CHD) in most medical centers is the lack of specialized cardiology teams. Furthermore, significant regional disparities exist in both the coverage and diagnostic accuracy of prenatal CHD screening [5,6]. China has the highest number of CHD patients worldwide, resulting in a considerable One Health burden, particularly in regions with lower sociodemographic indices [7,8].

In recent years, machine learning techniques have shown considerable promise in transforming cardiac risk prediction and diagnostic support. These approaches can automatically identify meaningful patterns from high-dimensional, nonlinear, and heterogeneous clinical datasets, frequently demonstrating higher predictive accuracy and better generalization performance compared to conventional statistical models [9,10]. Peng et al. developed an XGBoost-based model for cardiovascular disease risk assessment, achieving an AUC of 81%, thereby highlighting the practical utility of machine learning in clinical decision-making [11]. Moreover, Chandrasekhar et al. systematically evaluated the performance of six classical machine learning algorithms, including Random Forest, Logistic Regression, and Naive Bayes, on the Cleveland and IEEE Dataport datasets, offering valuable insights for model selection across diverse data environments [12].

However, the performance of machine learning models is highly dependent on the quality and structure of input features [13–15]. Insufficient training data represents a common bottleneck in supervised learning applications. Furthermore, real-world clinical datasets often encompass hundreds to thousands of variables across multiple dimensions, including laboratory indices, imaging features, medication history, and sociodemographic information, many of which may exhibit only weak associations with the prediction target or be redundant [16]. Such high-dimensional and sparse data structures are susceptible to the “curse of dimensionality”, which can lead to model overfitting, increased computational complexity, and diminished interpretability, ultimately constraining their practical utility in clinical settings [17–19].

Das et al. introduced five explainability techniques to interpret the predictions of four machine learning models for heart disease classification. Among them, the XGBoost model performed best, with a classification accuracy of 97.86% [20]. Arunachalam et al. utilized the X-boost, Adaboost and Random Subspace Classifier model to predict heart disease, and the Linear Support Vector Feature Measure (l-SVFM) was used to predict the features of cardiovascular disease. The final results were 96% accuracy and 97% precision [21].

Dileep et al. proposed a clustering-based bidirectional long short-term memory network (C-BiLSTM) for heart disease prediction. The method first applies the K-Means algorithm to cluster the UCI and a real-time heart disease dataset, with the clustering results subsequently serving as input to the C-BiLSTM model [22].

In a comprehensive comparative study, Ay et al. applied the cuckoo search, flower pollination algorithm, whale optimization algorithm, and Harris hawks optimization algorithm for feature selection on the Cleveland heart disease dataset and the Faisalabad Institute of Cardiology heart failure dataset sourced from the UCI Machine Learning Repository [23]. To enhance computational efficiency with high-dimensional data, Babu et al. introduced a quantum-enhanced machine learning approach that replaces traditional input layer neurons with qubits for efficient handling of complex tasks [24]. Ali et al. preprocessed a Kaggle dataset containing 1025 samples and 14 attributes using a ReplaceMissingValues filter and interquartile range for outlier detection [25].

By utilising a convolutional network with dense connections and an attention-based bidirectional long short-term memory (AMiLSTM) network, a hybrid deep learning (HDL) model known as DenseNet-ABiLSTM has been introduced by Saranya et al., achieving good results in the detection of arrhythmias [26].

With the continuous innovation and iterative development of various intelligent optimisation algorithms, Kumar et al. innovatively introduced two advanced optimisation strategies, the Hawks Optimizer (HO) and the Binary Krill Herd meta-heuristic optimiser (B-KHA), in their research [27,28]. Through the synergistic interaction of these algorithms and iterative optimisation, the accuracy of model solutions and search efficiency have been effectively enhanced, thereby enabling the precise identification of both global and local optimal solutions to the target problem, and providing a more efficient and reliable technical approach for solving complex optimisation problems.

Al-Alshaikh et al. integrated optimization algorithms into hybrid machine learning models and established a machine learning-based heart disease prediction method (ML-HDPM), which obtained a predictive accuracy of 95.5% and a precision rate of 94.8% in experiments [17]. Kumar et al. utilized Classical Genetic Algorithms (CGA) and Classical Particle Swarm Optimization (CPSO) to conduct hyperparameter tuning for traditional models. Meanwhile, quantum-inspired prediction models adopted Quantum Genetic Algorithms (QGAs) and Quantum Particle Swarm Optimization (QPSO) for parameter optimization, and the optimized parameters were finally imported into the Support Vector Machine (SVM) for prediction tasks [29]. Although these hybrid optimization methods can achieve competitive prediction accuracy, they cannot achieve satisfactory performance on complex heterogeneous datasets. Moreover, such models have limited generalization ability and are prone to unstable prediction results when facing unseen data, which affects their practical application effectiveness.

It is noteworthy that a novel metaheuristic algorithm, recently proposed by Dong et al. and termed CEO, has been introduced based on the chaotic dynamics observed in two-dimensional discrete memristive maps. By incorporating crossover and mutation operations from the differential evolution framework into a chaotic evolutionary mechanism, the algorithm effectively addresses the “zero-bias problem” commonly encountered in many emerging optimization algorithms [30].

In summary, existing studies have extensively investigated various machine learning and feature engineering techniques for heart disease prediction, achieving significant outcomes. However, these simple methods suffer distinct drawbacks: standalone tree models lack nonlinear fitting capacity for continuous clinical indicators, vanilla deep networks cannot natively process categorical medical features, and unoptimized structures easily fall into local optima, leading to poor cross-dataset robustness against noisy samples. To address these limitations, this paper aims to develop a cardiac risk identification model that achieves high predictive accuracy while maintaining strong robustness, even under constraints of limited data volume and feature availability.

The contributions of this paper are described as follows:

- The proposed CEO-HCDLM framework enables CatBoost to extract categorical features, utilizes DNN for processing numerical features, and integrates both through Logistic Regression for final prediction. Subsequently, the performance of the CEO-HCDLM model was systematically evaluated against established models, including XGBoost, LightGBM, and Logistic Regression, to demonstrate its superior predictive capability.
- A customized improved Chaotic Evolution Optimization (CEO) algorithm is developed for joint hyperparameter tuning of the CatBoost-DNN stacking hybrid model, which optimizes search efficiency, speeds up training convergence and improves overall prediction stability.
- A new adaptive chaotic perturbation and training stagnation restart module is integrated into the CEO framework, which dynamically modifies key hyperparameters and reactivates training to effectively avoid trapping into local optimal solutions.

2. Materials and Methods

Existing single machine learning and deep learning architectures cannot simultaneously adapt to mixed categorical-numerical clinical data and guarantee stable generalization on noisy, small-scale medical datasets. To overcome these limitations, we constructed the CEO-HCDLM, which effectively integrates CatBoost and DNN through Logistic Regression and employs an improved CEO algorithm for parameter tuning and model training. The overall workflow of the model is illustrated in Figure 1.

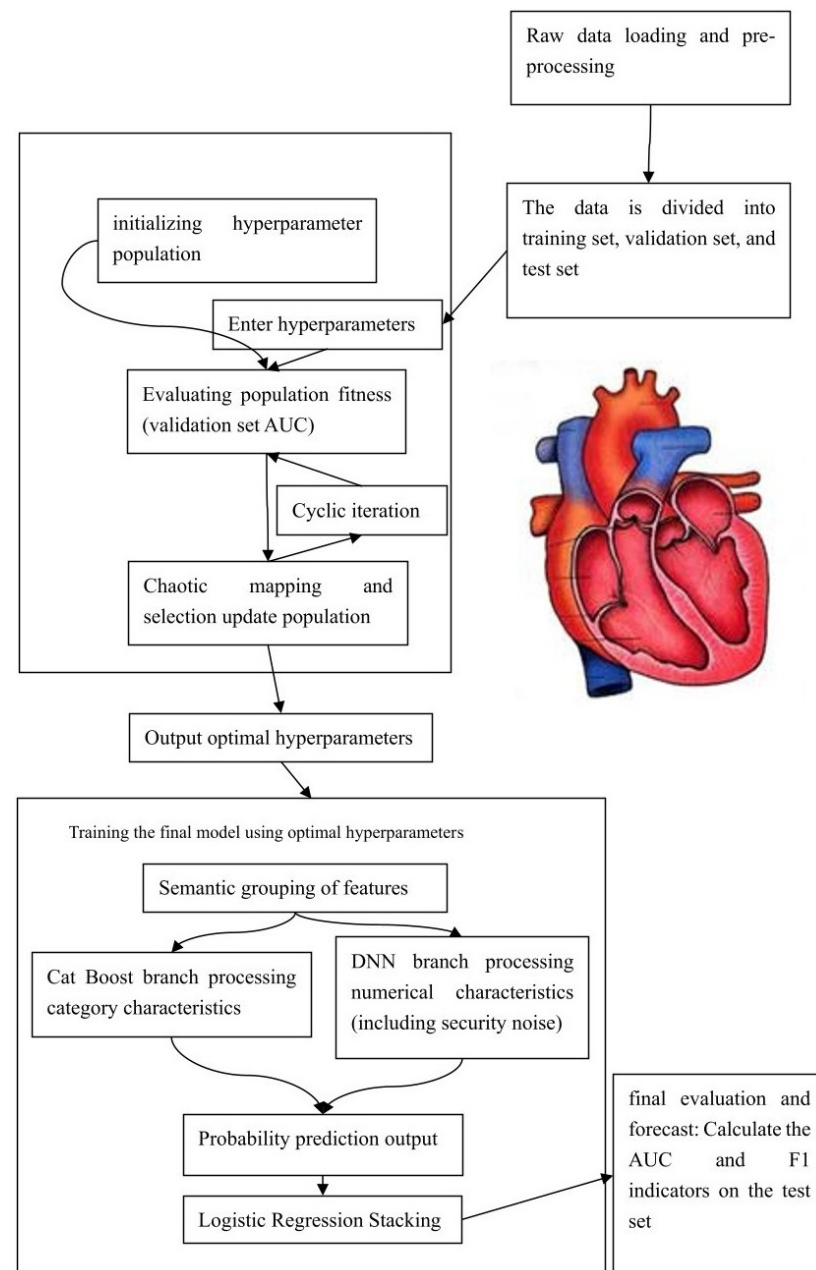


Figure 1. The overall workflow diagram.

2.1. Pre-Processing

Two publicly available datasets are employed in this paper: the Cleveland Heart Disease dataset and the Clinical Heart Disease dataset from the UCI Machine Learning Repository. To ensure experimental rigor and reproducibility, a standardized preprocessing protocol was applied. Samples were randomly partitioned into training, validation, and test sets in an 8:1:1 ratio. Missing values were imputed using the first valid category, and unknown categories were replaced with the first known category. To enhance the

performance and clinical interpretability of the cardiac risk prediction model, this study implemented systematic feature engineering on the raw dataset, expanding the feature dimension from basic indicators to 29 dimensions. Firstly, unordered categorical variables such as chest pain type, resting ECG type and ST-segment slope were subjected to one-hot encoding to eliminate spurious ordinal relationships in categorical variables whilst retaining discrete clinical attributes; secondly, the multivariate interaction features was introduced to quantify the synergistic effects of age and exertional angina, blood pressure and ST-segment changes, and cholesterol and ST-segment slope, thereby capturing the pathogenic mechanisms underlying the overlap of multiple risk factors in clinical practice.

Furthermore, leveraging CatBoost's native support for categorical features, an innovative semantic feature grouping approach was adopted to assign categorical features to CatBoost, while numerical features were directed to the enhanced DNN model. Additionally, each numerical feature was Z-score normalized to mitigate the influence of varying magnitudes on model training and to facilitate smoother gradient updates, as follows:

$$\sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(x_i^{(j)} - \frac{1}{n} \sum_{i=1}^n x_i^{(j)} \right)^2} \quad (1)$$

$$Z_i^{(j)} = \frac{x_i^{(j)} - u_j}{\sigma_j} \quad (2)$$

where $x_i^{(j)}$ is the value of the i th sample in the j th numerical feature column, u_j is the mean value of the j th numerical feature column, σ_j is the standard deviation of the j th numerical feature column, and $Z_i^{(j)}$ is the normalized value of the i th sample in the j th numerical feature column.

2.2. Hybrid CatBoost and Deep Learning Model

CatBoost is incorporated into the modeling pipeline to directly handle categorical variables, owing to its superior performance in managing categorical data. The specific implementation procedure is described as follows.

For the objective function, the log loss for the binary classification task at the t th tree round was used as follows:

$$\mathcal{L}_{\text{cat}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3)$$

$$p_i = \frac{1}{1 + e^{-F^{(t)}(x_i^{(c)})}} \quad (4)$$

where N is the total number of samples, y_i is the true label of the i th sample, $\mathcal{L}_{\text{cat}}$ is the logarithmic loss, p_i is the predicted probability that the i th sample belongs to the positive class, $F^{(t)}(x_i^{(c)})$ is the weighted sum of all trees at round t , and $x_i^{(c)}$ is the categorical discrete feature vector.

At round t , the negative gradient can then be computed as:

$$g_i^{(t)} = \frac{\partial \mathcal{L}_{\text{cat}}}{\partial F^{(t-1)}} = p_i^{(t-1)} - y_i \quad (5)$$

where $g_i^{(t)}$ is the negative gradient of the i th sample at round t , and $p_i^{(t-1)}$ is the predicted probability obtained from the previous round.

A symmetric tree is subsequently fitted in gradient space, denoted as $\Delta F^{(t)}(x^{(c)})$, so the model can be updated as follows:

$$F^{(t)}(x_c) = F^{(t-1)}(x_c) + \eta \Delta F^{(t)}(x^{(c)}) \quad (6)$$

where η is the learning rate.

Finally, the category probabilities for the validation and test sets can be obtained.

$$\hat{p}_i^{\text{CB}} = \sigma\left(F^I(x_i^{(c)})\right) \quad (7)$$

where I is the total number of iterations, and $\hat{p}_i^{\text{CB}}$ is the CatBoost predicted probability for the i th sample.

2.3. DNN

The DNN module incorporated two hidden layers in this paper. To improve gradient flow and strengthen nonlinear representation capacity, the SiLU activation function was utilized. The specific network architecture is described as follows:

$$h^{(l)} = \text{Dropout}\left(\text{SiLU}\left(\text{BN}\left(W^l h^{(l-1)} + b^{(l)}\right)\right)\right) \quad (8)$$

$$\text{SiLU}(x) = x \cdot \sigma(x) \quad (9)$$

$$\hat{p}_i^{\text{DNN}} = \sigma\left(W^T h_i^{(L)} + b^{(l)}\right) \quad (10)$$

where l is the number of layers, BN denotes 1D Batch Normalization, $h^{(l)}$ is the output of the l th layer, W^l is the weight matrix of the l th layer, and $b^{(l)}$ is the bias vector of the l th layer.

Binary cross-entropy was employed to penalize deviations between model predictions and the actual outcomes by quantifying uncertainty, thereby guiding the model to optimize its parameters and minimize prediction error.

$$\mathcal{L}_{\text{DNN}} = -\frac{1}{N} \sum_{i=1}^N \left[y_i \log \hat{p}_i^{\text{DNN}} + (1 - y_i) \log (1 - \hat{p}_i^{\text{DNN}}) \right] \quad (11)$$

2.4. Stacking

To achieve probabilistic splicing of two models, a two-dimensional feature vector was sampled for each sample as follows:

$$s_i = \begin{bmatrix} \hat{p}_i^{\text{DNN}} \\ \hat{p}_i^{\text{CB}} \end{bmatrix} \in \mathbb{R}^2 \quad (12)$$

Logistic regression was employed to perform linear mapping, followed by the application of the sigmoid function.

$$\eta_i = w^T s_i + b \quad (13)$$

$$\hat{p}_i^{\text{final}} = \sigma(\eta_i) = \frac{1}{1 + e^{-(w^T s_i + b)}} \quad (14)$$

Then, the categorical labels can be obtained as follows:

$$\hat{y} = 1\left(\hat{p}_i^{\text{final}} \geq \tau\right) \quad (15)$$

where $\tau = 0.5$ is the threshold value, and $1(\cdot)$ denotes the indicator function.

2.5. CEO

The CEO algorithm was employed to automate the tuning of seven key hyperparameters in the model. The optimization objective was defined as minimizing the negative AUC metric on the validation set, which equals maximizing the model's classification performance, as detailed below:

$$\text{fitness}(\theta) = -\text{AUC}_{\text{val}}(\theta) \quad (16)$$

where θ is the hyperparameter vector to be optimized.

A similar optimization process was employed for both CatBoost and DNN. The first step involves population initialization, which ensures that each parameter is randomly distributed within a reasonable range by generating an initial hyperparameter population using the following formula:

$$X_0 = lb + (ub - lb) \cdot U(0, 1)^{a \times b} \quad (17)$$

where X_0 is the initial population matrix, lb and ub are the lower-bound and upper-bound vectors, respectively, and $U(0, 1)^{a \times b}$ is a uniform random matrix with a rows and b columns.

For each population individual, the chaotic perturbation was updated. To compute the chaotic control parameter, the iteration starts by updating it via the Logistic map to ensure perturbation randomness.

$$c(t) = \text{logistic}_{\text{map}}(U(0, 1)) = 4 \cdot U(0, 1) \cdot (1 - U(0, 1)) \quad (18)$$

where $t \in [1, \max(\text{iter})]$ is the current iteration number, $c^{(t)} \in (0, 1)$ is the chaotic control parameter at iteration t , and $U(0, 1)$ denotes the uniform distribution on the interval $[0, 1]$.

The adaptive coefficient $\alpha(t)$ is calculated to control the step size of “approaching the optimal individual”, which decreases with the number of iterations:

$$\alpha(t) = 2 \cdot e^{-\left(\frac{4t}{\max(\text{iter})}\right)^2} \quad (19)$$

where $\alpha^{(t)}$ is the adaptive coefficient at iteration t , which decreases from 2 to 0 as t increases; $\max(\text{iter})$ is the maximum allowable number of iterations.

For each individual population, new individuals were generated by combining the strategies of “learning from the best” and “chaotic perturbation” as follows:

$$X_i^{(t), \text{new}} = X_i^{(t)} + \alpha(t) \cdot \left[r_1 \cdot (X_{\text{best}}^{(t)} - X_i^{(t)}) \right] + [c(t) \cdot (ub - lb) + lb] \cdot r_2 \quad (20)$$

$$X_i^{(t), \text{new}} = \text{clip}(X_i^{(t), \text{new}}, lb, ub) \quad (21)$$

where $X_i^{(t)}$ is the i th individual in round t , $X_{\text{best}}^{(t)}$ is the optimal individual in round t of the population, $X_i^{(t), \text{new}}$ is the generated new individual in the t -th round of the population, lb and ub are the lower bound and upper bound, respectively, and $r_1, r_2 \in (0, 1)$ are independent uniform random numbers, $\text{clip}(\cdot)$ is a clipping function that ensures that the parameters of the new individual are not outside the range $[lb, ub]$.

Next, the fitness of $X_i^{(t), \text{new}}$, denoted as $\text{fitness}_{\text{new}}$, is calculated.

$$X_i^{(t)} = \begin{cases} X_i^{(t), \text{new}}, & \text{fitness}_{\text{new}} < \text{fitness}(X_i^{(t)}), \\ X_i^{(t)}, & \text{fitness}_{\text{new}} \geq \text{fitness}(X_i^{(t)}). \end{cases} \quad (22)$$

$$\text{fitness}\left(X_i^{(t)}\right)=\left\{\begin{array}{ll}\text{fitness}_{\text{new}}, & \text{fitness}_{\text{new}} < \text{fitness}\left(X_i^{(t)}\right), \\ \text{fitness}\left(X_i^{(t)}\right), & \text{fitness}_{\text{new}} \geq \text{fitness}\left(X_i^{(t)}\right).\end{array}\right. \quad (23)$$

$$\text{if } \text{fitness}\left(X_i^{(t)}\right) < \text{fitness}_{\text{best}}, \text{ then } \left\{\begin{array}{l} X_{\text{best}} \leftarrow X_i^{(t), \text{new}}, \\ \text{fitness}_{\text{best}} \leftarrow \text{fitness}\left(X_i^{(t)}\right).\end{array}\right. \quad (24)$$

If the new individual is superior to the current population in terms of fitness, it replaces the least fit individual; if it is optimal within the current population, it is integrated into the population accordingly.

Finally, the values of X_{best} and $\text{fitness}_{\text{best}}$ can be obtained when the number of iterations reaches the maximum allowable value.

$$(X_{\text{best}}, \text{fitness}_{\text{best}}), \quad \text{fitness}_{\text{best}} = -\text{AUC}_{\text{val}}(\theta) \quad (25)$$

2.6. Adaptive Chaos Perturbation and Restart Mechanisms

During the training of deep neural networks, models may experience convergence stagnation as a result of suboptimal hyperparameter configurations. To address this issue, this study proposes an innovative adaptive chaotic perturbation restart mechanism. Upon detection of training stagnation, this mechanism generates perturbation terms based on chaotic mapping, adaptively adjusts key hyperparameters, and subsequently restarts the training process. This approach enables the model to escape local optima, reduces the risk of overfitting, and significantly improves convergence efficiency.

In practical implementation, the Logistic map was utilized to generate chaotic sequences, which were subsequently employed to construct perturbation coefficients for dynamically adjusting network hyperparameters.

$$\delta = \text{logistic}_{\text{map}}(U(0,1)) = 4 \cdot U(0,1) \cdot (1 - U(0,1)) \quad (26)$$

where $\delta \in (0,1)$ is the perturbation factor for hyperparameter adjustment, and $U(0,1)$ denotes the continuous uniform distribution on the interval $[0,1]$.

The model can dynamically adjust its learning rate according to the training progress as follows:

$$\eta_{\text{dnn}}^{\text{new}} = \text{clip}\left(\eta_{\text{dnn}}^{\text{old}} \cdot (1 + \delta \cdot 0.2), 10^{-5}, 0.005\right) \quad (27)$$

where $\eta_{\text{dnn}}^{\text{old}}$ is the learning rate before chaotic perturbation, and $\eta_{\text{dnn}}^{\text{new}}$ is the adjusted learning rate after clipping.

The final adjustment of the dropout rate can be conducted as follows:

$$p_{\text{dropout}}^{\text{new}} = \text{clip}\left(p_{\text{dropout}}^{\text{old}} \cdot (1 + \delta \cdot 0.2), 0.3, 0.7\right) \quad (28)$$

where $p_{\text{dropout}}^{\text{old}}$ is the dropout rate before chaotic perturbation, and $p_{\text{dropout}}^{\text{new}}$ is the adjusted dropout rate after clipping.

After hyperparameter adjustment, if the validation loss exceeds the current optimal loss, the training process is deemed to have reached convergence stagnation, which triggers the restart mechanism; when the predefined maximum number of restarts is reached, training is terminated and the best model parameters obtained so far are retained.

As shown in Figure 2, when the loss on the training set is significantly lower than that on the validation set, the chaotic mechanism effectively resets and prevents the training from getting stuck in local optima.

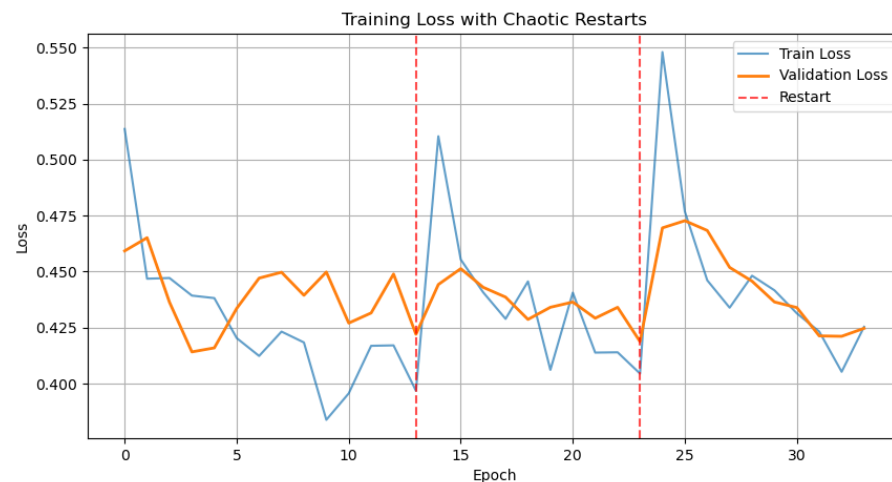


Figure 2. Training loss with chaotic restarts.

3. Results

The performance of the proposed strategy was assessed through a series of comprehensive simulation experiments. Evaluation metrics included precision, F1-score, and AUC, among others, to assess the predictive efficacy of the model.

3.1. Databases

The dataset on cardiac disease used in this study was obtained from the Kaggle platform and consists of patient records associated with cardiac diagnoses. The dataset comprises 918 samples, each defined by 11 features encompassing demographic information, vital signs, symptom presentation, and potential risk factors related to cardiac conditions. In the predictive modeling process, a total of 21 variables were included, with the presence of heart disease designated as the outcome variable to represent the patient's disease status.

The Cleveland Heart Disease dataset from the UCI Machine Learning Repository was also utilized. The original data were obtained from the Cleveland Clinic Foundation and comprised a total of 303 valid samples. Although the original dataset comprises 76 attributes, the publicly available subset includes only 14 key features. In this study, 13 attributes were selected as input features, and a binary classification variable was defined as the prediction target to indicate the presence or absence of heart disease in patients.

3.2. Experimental Setup for Evaluation

To ensure the reproducibility and stability of the results on both UCI Cleveland and Kaggle heart disease datasets, 11 repeated experimental runs with random seeds ranging from 40 to 50 were implemented for cross-dataset validation, and each dataset was stratified and split into training, validation and test sets in an identical 8:1:1 ratio. Bidirectional cross-dataset tests were conducted where the model trained and optimized on one dataset was directly evaluated on the other dataset. The CatBoost branch used unified base optimisation parameters across all experiments, whilst the DNN branch employed the AdamW optimisation algorithm, with the initial learning rate determined by chaotic evolutionary search and a fixed weight decay rate of 0.01. The batch size was set to 64, and the maximum number of training iterations was set to 100. The CosineAnnealingLR scheduler was employed, with the iteration period matching the number of training iterations. During backpropagation, gradient clipping with a maximum norm of 1.0 was applied, and an early stopping strategy with a patience level of 10 was implemented based on the validation set loss, whilst supporting up to two chaotic perturbation restarts for alleviating

local optimum trapping. All performance metrics were averaged over the 11 multi-seed trials to reduce stochastic fluctuations.

The experiments were conducted as follows.

- The entire dataset was used without employing any data dimensionality reduction or sampling methods.
- Categorical features were processed using numerical encoding techniques to efficiently handle variables with multiple discrete levels.
- The results for both datasets were obtained across a total of 11 datasets using random seeds ranging from 40 to 50.
- A specially designed chaotic evolution algorithm was applied to the processed data, which automated hyperparameter tuning and ensemble weight optimization, significantly enhancing model performance for the specific task.

Furthermore, Gaussian noise was introduced into 30% of the numerical features using a specifically designed stochastic function to assess the models' stability and generalization performance.

3.3. Simulation Results

In this section, we have included a CatBoost-DNN model without the CEO algorithm, a CatBoost-DNN model with the Successive Halving optimisation algorithm (SH-Hybrid), and a CatBoost-DNN model with the Tree-structured Parzen Estimator optimisation algorithm (TPE-Hybrid) as part of our ablation experiments. As shown in Figure 3, the CEO-HCDLM model exhibits satisfactory performance in predicting heart disease. This performance enhancement is attributable to the integration of several advanced methodologies, including categorical feature selection and transformation, data balancing, the synergistic application of CatBoost and DNN, and hyperparameter optimization through a chaotic evolutionary algorithm. Moreover, the performance of the CEO evolutionary algorithm surpasses that of the mainstream TPE optimization algorithm. We further conducted Wilcoxon rank-sum tests to statistically verify the significant performance gap between CEO and TPE optimizers, where * denotes statistical significance and ns indicates no significant difference. The test results demonstrate the evident advantage of CEO among evolutionary optimization algorithms.

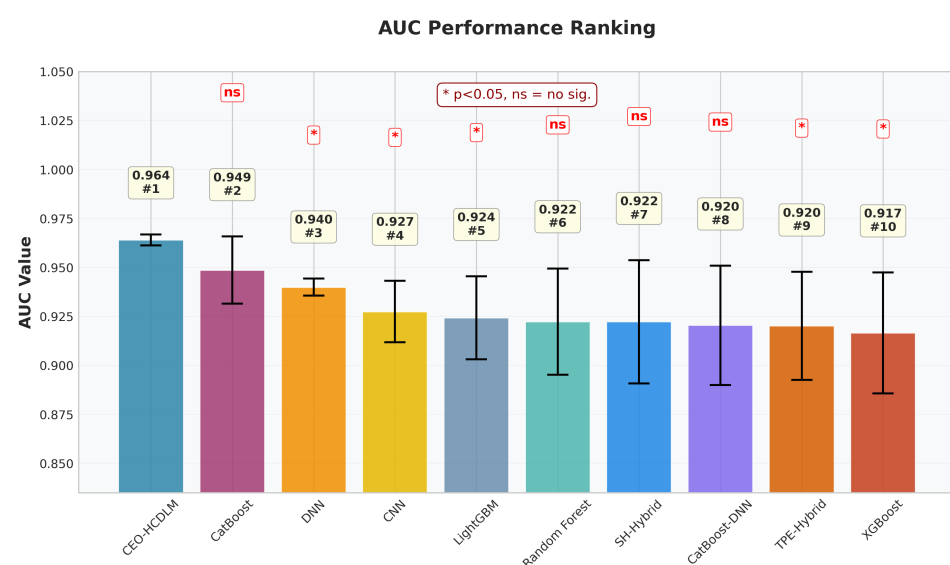


Figure 3. Average AUC analysis of predicting heart disease.

Figures 4 and 5 present a comparative analysis of precision and recall performance across various heart disease prediction methods, respectively, with CEO-HCDLM demonstrating superior performance. Simultaneously, it demonstrates that our model exhibits the most robust precision performance across various dataset partitions. This performance can be primarily attributed to an integrated chaotic evolution optimization framework, which combines early stopping, restart mechanisms, and a stochastic evolutionary selection approach to effectively avoid suboptimal hyperparameter configurations and mitigate the risk of converging to local optima.

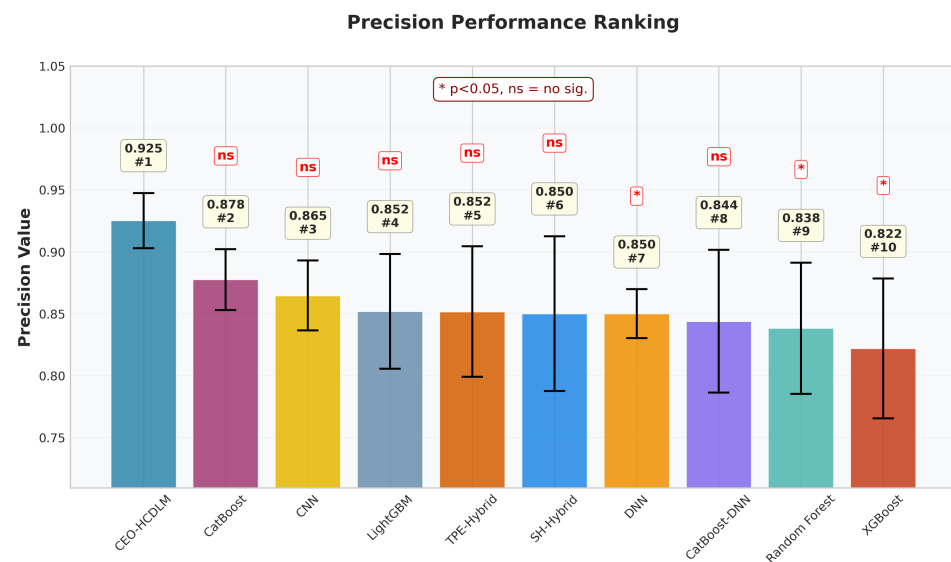


Figure 4. Average precision analysis of predicting heart disease.

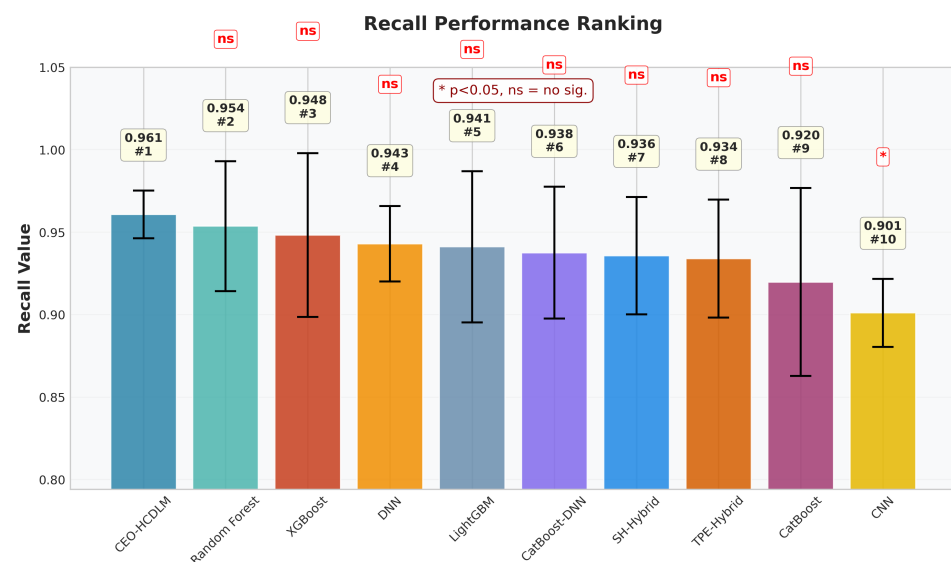


Figure 5. Average recall analysis of predicting heart disease.

As illustrated in Figures 6 and 7, the CEO-HCDLM model demonstrates better performance compared to other benchmark algorithms in terms of both F1-score and PR-AUC. A more balanced trade-off between precision and recall is achieved, enabling effective identification of discriminative patterns in cardiac disease data and demonstrating strong robustness even on imbalanced datasets with uneven sample distributions.

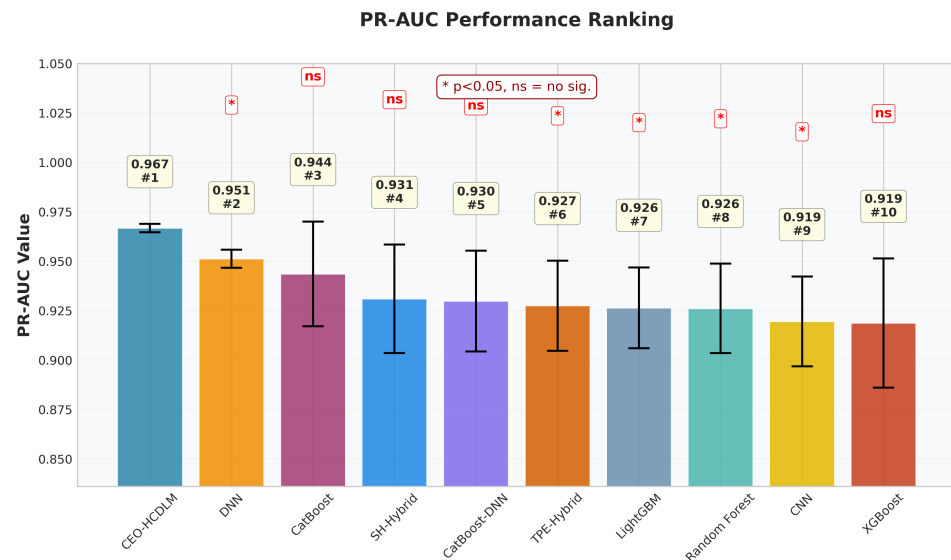


Figure 6. Average PR-AUC analysis of predicting heart disease.

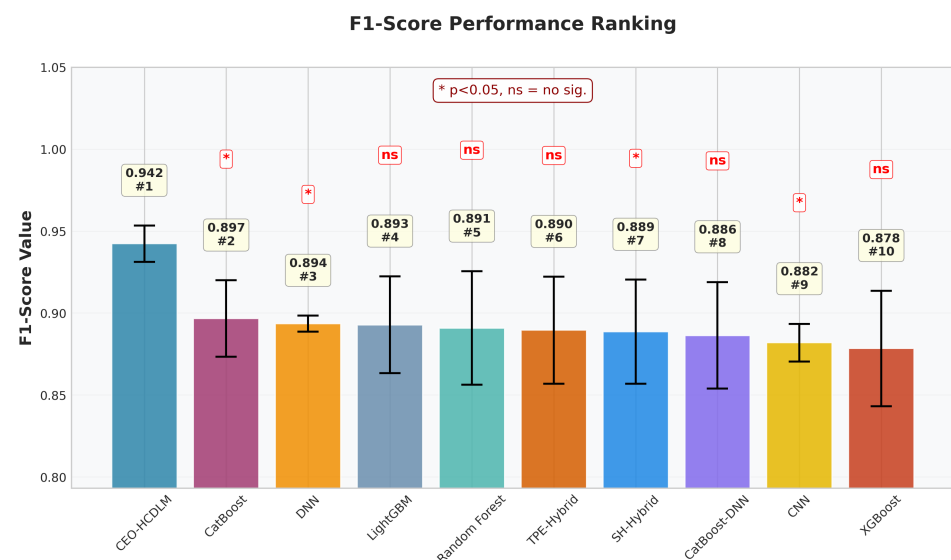


Figure 7. Average F1-score analysis of predicting heart disease.

Additionally, this study randomly selected 30% of the positive features from both numerical datasets, added Gaussian noise with a standard deviation σ , and truncated any negative values after noise injection to preserve medical interpretability. Specifically, the following procedures were applied to each of the selected feature columns:

$$x = (x_1, x_2, \dots, x_n)^T \quad (29)$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (30)$$

$$\tilde{x}_i = x_i + \epsilon_i \quad (31)$$

$$\hat{x}_i = \max(\tilde{x}_i, 0) \quad (32)$$

where x_i is the original value of the i th sample in the current feature column, ϵ_i is the noise value added to the i th sample and follows a Gaussian distribution with mean 0 and variance σ^2 , $\tilde{x}_i$ is the perturbed feature value, and $\hat{x}_i$ is the final obtained privacy-preserving feature value.

Following the introduction of Gaussian noise, the discrepancy between the model and the UCI dataset remains consistently within 1%, demonstrating the model's strong

robustness. The robustness of the model stems from its chaotic random restart mechanism, which enables continuous retraining and effectively prevents the model from converging to local optima. This enhances the model's generalization capability and improves its adaptability to real-world predictions on high-noise datasets that have not undergone rigorous preprocessing. The results are presented in Table 1.

Table 1. Dataset change analysis of predicting heart disease.

Assessment Criteria	Kaggle Dataset with Gaussian Noise	UCI Dataset with Gaussian Noise	Kaggle Dataset Without Gaussian Noise	UCI Dataset Without Gaussian Noise
AUC	0.9637	0.9767	0.9640	0.9798
F1	0.9431	0.9423	0.9424	0.9457
Recall	0.9590	0.9572	0.9608	0.9679
Precision	0.9279	0.9318	0.9252	0.9273
PR-AUC	0.9670	0.9811	0.9667	0.9838

The performance of the proposed CEO-HCDLM in heart disease prediction was systematically evaluated through comprehensive simulation experiments. Satisfactory predictive performance was consistently demonstrated across multiple evaluation metrics, particularly on the Kaggle heart disease dataset, where the model achieved high scores with an average AUC of 96.40%, precision of 92.52%, F1-score of 94.24%, and PR-AUC of 96.67%.

Collectively, these results validate the strong and consistent performance across different benchmark datasets, demonstrating its generalizability to diverse data distributions and the robust performance under noisy conditions, highlighting its ability to withstand real-world data perturbation. The minor performance variations observed between datasets and noise conditions further underscore the model's stability and reliability, distinguishing it from less robust approaches that often suffer from substantial performance degradation when exposed to noisy or heterogeneous data. These findings collectively support the effectiveness and practical utility of the proposed model for heart disease prediction.

4. Discussion

Experimental results show the proposed method performs well across multiple metrics, maintaining high precision in both training and testing. The method demonstrates concurrent improvements in false positive rate and recall, suggesting its ability to effectively reduce false alarms while preserving high predictive precision. Additionally, the balanced F1-score provides further evidence of an optimal trade-off between precision and recall.

The proposed CEO-HCDLM addresses inherent limitations of standalone CatBoost, DNN and simple logistic regression, each of which fails to jointly handle heterogeneous heart disease data. The tree models lack strong nonlinear modeling capacity for continuous physiological indicators, basic deep networks cannot natively process unordered categorical variables, and unoptimized stacking architectures frequently converge to local optima with volatile performance across datasets. The split-branch design separates categorical and numerical features for targeted extraction via CatBoost and DNN respectively, and integrates their predictive probabilities through logistic regression, with enhanced chaotic evolution optimization automatically searching optimal hyperparameters to stabilize training convergence. Furthermore, logistic regression provides clear linear weights for the two model branches, making the predictive model easier to interpret in clinical analysis and reducing the risk of overfitting caused by non-linearity in small-scale healthcare datasets. Though the multi-module structure introduces mild computational complexity, the two-branch scheme retains interpretable categorical risk signals while capturing complex nonlinear relationships of numerical biomarkers, and the embedded adaptive chaotic restart mechanism

maintains consistent AUC, F1 and PR-AUC under Gaussian noise and cross-dataset tests even on small datasets.

Most mainstream hybrid learning models currently employ simple serial feature concatenation and ensemble methods with fixed weights, resulting in significant fluctuations in generalisation performance across different medical datasets. The dynamic dual-branch fusion architecture developed in this study effectively addresses the issue of feature redundancy commonly found in traditional hybrid models. However, this study still has certain shortcomings in the data processing stage. The publicly available clinical datasets used in this study contain inherent noise and missing samples; existing pre-processing methods struggle to fully eliminate issues such as data distribution bias and sample imbalance. Furthermore, certain inefficient features relating to heterogeneous indicators were not sufficiently filtered out, which to some extent limited further improvements in the model's predictive stability and generalisation ability.

5. Conclusions

This study investigated the effectiveness of the CEO-HCDLM algorithm in predicting heart disease. In the preprocessing stage, missing value imputation and normalization were applied to enhance the performance of the classification model. To effectively extract diverse feature types, categorical features were fed into CatBoost, numerical features were processed by a deep neural network (DNN), and finally, a Logistic Regression layer combined with a Sigmoid activation function was employed as a stacking mechanism to integrate the outputs efficiently. The improved CEO algorithm is employed to adapt hyperparameters to diverse datasets by leveraging a specific adaptive approach that identifies and exploits the underlying parameter evolution patterns, thereby accelerating the model's overall convergence rate without compromising its global optimality.

Although the proposed hybrid method achieves stable predictive performance on public heart disease datasets, all evaluations are restricted to open-source benchmark data, lacking independent external clinical cohort verification from real medical institutions. Future research should prioritize extending the applicability of this method to a wider range of cardiovascular conditions and assessing its practical feasibility in real-world clinical settings. In addition, the method proposed in this paper does not explore more complex nonlinear fusion strategies, so future research can focus on the deep interaction between heterogeneous features. Overall, this study underscores the significant transformative potential of advanced machine learning methods in the field of cardiovascular disease prediction and management.

Author Contributions: Z.S.: Modeling, Investigation, Supervision, Validation, Writing, Visualization. J.L.: Modeling, Investigation, Supervision, Validation, Writing, Visualization. A.W.: Investigation, Writing, Visualization, Supervision, Validation. Y.L.: Investigation, Validation, Writing, Funding. M.L.: Investigation, Writing, Visualization. G.C.: Investigation, Writing, Visualization. Z.X.: Investigation, Writing, Visualization. Y.Z.: Investigation, Writing, Visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61903043 and Grant 61971048 and the National College Students' Innovation and Entrepreneurship Training Program under Grant 202611232361.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All datasets analyzed in this work are publicly accessible open benchmark datasets. The utilized heart disease datasets can be accessed via the following links: <https://archive.ics.uci.edu/dataset/45/heart+disease> (accessed on 18 September 2025) and <https://www.kaggle.com/datasets/amirmahdiabbootalebi/heart-disease> (accessed on 18 September 2025). There are no access restrictions on these datasets, and other researchers are welcome to reuse the data for subsequent analysis and verification.

Acknowledgments: The authors gratefully acknowledge the computational support provided by the DeepSeek-V2.5 large model, which was utilized for experimental construction, data calculation, result statistics and auxiliary analysis throughout the research process. The research framework, experimental scheme design and final discussion of conclusions were independently determined and verified by the authors, and all experimental raw data were manually checked and validated by the authors to guarantee the reliability and authenticity of the research results.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rehman, M.U.; Naseem, S.; Butt, A.U.R.; Mahmood, T.; Khan, A.R.; Khan, I.; Khan, J.; Jung, Y. Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. *Sci. Rep.* **2025**, *15*, 13361. [\[CrossRef\]](#) [\[PubMed\]](#)
2. World Health Organization Regional Office for Africa. Cardiovascular Diseases. Available online: <https://www.afro.who.int/health-topics/cardiovascular-diseases> (accessed on 18 September 2025).
3. Zhou, C.; Dai, P.; Hou, A.; Zhang, Z.; Liu, L.; Li, A.; Wang, F. A comprehensive review of deep learning-based models for heart disease prediction. *Artif. Intell. Rev.* **2024**, *57*, 263. [\[CrossRef\]](#)
4. Praz, F.; Beyersdorf, F.; Haugaa, K.; Prendergast, B.; Hubert, A.; Rosenhek, R.; Lengyel, M.; Maisano, F.; Montalescot, G.; Vahanian, A.; et al. Valvular heart disease: From mechanisms to management. *Lancet* **2024**, *403*, 1576–1589. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Zhang, Y.; Wang, J.; Zhao, J.; Huang, G.; Liu, K.; Pan, W.; Sun, L.; Li, J.; Xu, W.; He, C.; et al. Current status and challenges in prenatal and neonatal screening, diagnosis, and management of congenital heart disease in China. *Lancet Child Adolesc. Health* **2023**, *7*, 479–489. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ma, K.; He, Q.; Dou, Z.; Hou, X.; Li, X.; Zhao, J.; Rao, C.; Feng, Z.; Sun, K.; Chen, X.; et al. Current treatment outcomes of congenital heart disease and future perspectives. *Lancet Child Adolesc. Health* **2023**, *7*, 490–501. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Martin, S.S.; Aday, A.W.; Allen, N.B.; Almarzooq, Z.I.; Anderson, C.A.M.; Arora, P.; Avery, C.L.; Baker-Smith, C.M.; Bansal, N.; Beaton, A.Z.; et al. 2025 heart disease and stroke statistics: A report of US and global data from the American Heart Association. *Circulation* **2025**, *151*, e41–e660. [\[CrossRef\]](#) [\[PubMed\]](#)
8. El-Chouli, M.; Meddis, A.; Christensen, D.M.; Gerds, T.A.; Sehested, T.; Malmborg, M.; Phelps, M.; Bang, C.N.; Ahlehoff, O.; Torp-Pedersen, C.; et al. Lifetime risk of comorbidity in patients with simple congenital heart disease: A Danish nationwide study. *Eur. Heart J.* **2023**, *44*, 741–748. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Rani, P.; Kumar, R.; Jain, A.; Lamba, R.; Sachdeva, R.K.; Kumar, K.; Kumar, M. An extensive review of machine learning and deep learning techniques on heart disease classification and prediction. *Arch. Comput. Methods Eng.* **2024**, *31*, 3331–3349. [\[CrossRef\]](#)
10. El-Sofany, H.; Bouallegue, B.; Abd El-Latif, Y.M. A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Sci. Rep.* **2024**, *14*, 23277. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Peng, M.; Hou, F.; Cheng, Z.; Shen, T.; Liu, K.; Zhao, C.; Zheng, W. Prediction of cardiovascular disease risk based on major contributing features. *Sci. Rep.* **2023**, *13*, 4778. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Chandrasekhar, N.; Peddakrishna, S. Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes* **2023**, *11*, 1210. [\[CrossRef\]](#)
13. Laurer, M.; van Atteveldt, W.; Casas, A.; Welbers, K. Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI. *Political Anal.* **2024**, *32*, 84–100. [\[CrossRef\]](#)
14. Malakouti, S.M. Heart disease classification based on ECG using machine learning models. *Biomed. Signal Process. Control* **2023**, *84*, 104796. [\[CrossRef\]](#)
15. Noroozi, Z.; Orooji, A.; Erfannia, L. Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Sci. Rep.* **2023**, *13*, 22588. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Rajagopal, R.; Ranganathan, V. Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification. *Biomed. Signal Process. Control* **2017**, *34*, 1–8. [\[CrossRef\]](#)
17. Al-Alshaikh, H.A.; P, P.; Poonia, R.C.; Saudagar, A.K.J.; Yadav, M.; AlSagri, H.S.; AlSanad, A.A. Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Sci. Rep.* **2024**, *14*, 7819. [\[CrossRef\]](#) [\[PubMed\]](#)

18. Mir, A.; Ur Rehman, A.; Ali, T.M.; Javaid, S.; Almufareh, M.F.; Humayun, M.; Ateeq, M.; Qamar, A.M.; Khan, M.S.; Ahmad, J.; et al. A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques. *ESC Heart Fail.* **2024**, *11*, 3742–3756. [[CrossRef](#)] [[PubMed](#)]
19. El-Sofany, H.F. Predicting heart diseases using machine learning and different data classification techniques. *IEEE Access* **2024**, *12*, 106146–106160. [[CrossRef](#)]
20. Das, S.; Sultana, M.; Bhattacharya, S.; Sengupta, D.; De, D. XAI-reduct: Accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI. *J. Supercomput.* **2023**, *79*, 18167–18197. [[CrossRef](#)] [[PubMed](#)]
21. Arunachalam, S.K.; Rekha, R. A novel approach for cardiovascular disease prediction using machine learning algorithms. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e7027. [[CrossRef](#)]
22. Dileep, P.; Rao, K.N.; Bodapati, P.; Gokuruboyina, S.; Peddi, R.; Grover, A.; Sheetal, A. An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm. *Neural Comput. Appl.* **2023**, *35*, 7253–7266. [[CrossRef](#)]
23. Ay, Ş.; Ekinci, E.; Garip, Z. A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. *J. Supercomput.* **2023**, *79*, 11797–11826. [[CrossRef](#)] [[PubMed](#)]
24. Babu, S.V.; Ramya, P.; Gracewell, J. Revolutionizing heart disease prediction with quantum-enhanced machine learning. *Sci. Rep.* **2024**, *14*, 7453. [[CrossRef](#)] [[PubMed](#)]
25. Ali, M.M.; Paul, B.K.; Ahmed, K.; Bui, F.M.; Quinn, J.M.W.; Moni, M.A. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Comput. Biol. Med.* **2021**, *136*, 104672. [[CrossRef](#)] [[PubMed](#)]
26. Saranya, K.; Karthikeyan, U.; Kumar, A.S.; Salau, A.O.; Tin, T.T. DenseNet-ABiLSTM: Revolutionizing multiclass arrhythmia detection and classification using hybrid deep learning approach leveraging PPG signals. *Int. J. Comput. Intell. Syst.* **2025**, *18*, 33. [[CrossRef](#)]
27. Kumar, A.S.; Rekha, R. An improved hawks optimizer based learning algorithms for cardiovascular disease prediction. *Biomed. Signal Process. Control* **2023**, *81*, 104442. [[CrossRef](#)]
28. Kumar, A.S.; Rekha, R. A dense network approach with Gaussian optimizer for cardiovascular disease prediction. *New Gener. Comput.* **2023**, *41*, 859–878. [[CrossRef](#)]
29. Kumar, A.; Dhanka, S.; Sharma, A.; Bansal, R.; Fahlevi, M.; Rabby, F.; Aljuaid, M. A hybrid framework for heart disease prediction using classical and quantum-inspired machine learning techniques. *Sci. Rep.* **2025**, *15*, 25040. [[CrossRef](#)] [[PubMed](#)]
30. Dong, Y.; Zhang, S.; Zhang, H.; Zhou, X.; Jiang, J. Chaotic evolution optimization: A novel metaheuristic algorithm inspired by chaotic dynamics. *Chaos Solitons Fractals* **2025**, *192*, 116049. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.