

ROHG-BXG: Robust Outlier Handling with GAN and BOA-Optimized XGBoost for Heart Disease Prediction

Shaik Sajeera Begum

School of Technology, Woxsen University, Hyderabad, Telangana, India.
E-mail: shaiksajeerabegum9@gmail.com
ORCID: <https://orcid.org/0009-0004-1913-9050>

Sanjay Dhanka

Centre for Internet of Things, MITS-DU, Gwalior, Madhya Pradesh, India.
Corresponding author: sanjaykumar506070@gmail.com
ORCID: <https://orcid.org/0000-0003-1604-9645>

Syed Javeed Pasha

School of Technology, Woxsen University, Hyderabad, Telangana, India.
E-mail: syedjaveedrs@gmail.com
ORCID: <https://orcid.org/0000-0002-4091-0960>

(Received on May 16, 2026; Revised on July 13, 2026 & August 20, 2026; Accepted on September 20, 2026)

Abstract

Heart disease (HD) is a major cause of death globally. Early and accurate prediction is crucial for effective clinical decision making. However, the available datasets in the healthcare domain are prone to data noise, structural outliers, and data imbalance, which adversely affect the performance of traditional machine learning (ML) models. To overcome these problems, the current study presents a novel approach called ROHG-BXG (Robust Outlier Handling with Generative Adversarial Network and Butterfly Optimization Algorithm-Optimized XGBoost) for HD prediction problem in the Personal Key Indicators of Heart Disease (PKIOHD) dataset and Framingham dataset. This proposed scheme includes a structural outlier detection algorithm (Isolation Forest), a loss-based on-the-fly outlier detection algorithm to eliminate unreliable samples (Outlier Reduction), a class balancing algorithm (Conditional Generative Adversarial Networks (CGAN)) and an optimization algorithm for the hyperparameters of the XGBoost algorithm (Butterfly Optimization Algorithm (BOA)). The experimental results achieved 92% and 93% accuracy for both datasets. The framework was also shown to have an excellent Brier Score of 0.0655 and 0.0391 and an excellent Expected Calibration Error (ECE) of 0.032 and 0.0217 for both datasets. The outcomes of this work show the proposed ROHG-BXG framework is a powerful and reliable tool to predict HD and can be utilized in intelligent clinical decision-support systems.

Keywords- Butterfly optimization algorithm, Generative adversarial network, Heart disease prediction, Isolation forest, On-the-fly outlier detection, Robust outlier handling, XGBoost.

1. Introduction

Cardiovascular diseases represent one of the topmost health issues and high percentages of the global mortality. Earlier prediction of the people who are at risk of HD will assist the medical personnel to take preventative measures and enhance patient care. ML methods have become more commonly used in medical diagnosis because they can be used to study and identify complicated relationships among lifestyle factors, demographic variables, and an increased risk of disease (World Health Organization, 2025). Several ML models have been extensively used to predict HD including Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Naive Bayes (NB), and Logistic Regression (LR). Even though these models have positive outcomes, they are often problematic when applied to real-life healthcare data. Medical datasets are often noisy, suffer errors during measurements, and include anomalous samples, which have adverse impacts on model training and prediction accuracy (Mishra and Mohapatra, 2023). Class

imbalance is also another significant issue in predicting HD. In most medical records, there are often a huge number of people who are healthy as compared to the number of patients who have HD. CGANs have been demonstrated to be effective in the generation of synthetic minority class samples (Pandey et al., 2025) and bio-inspired algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) have been applied to hyperparameter tuning. Nevertheless, the majority of the current literature addresses one of the facets of the issue and does not concurrently target noisy observations, outliers, and class imbalance. To address these constraints, this study suggests a hybrid ML architecture named ROHG-BXG that incorporates an effective outlier processing scheme, synthetic data creation, and metaheuristic optimization to enhance reliability of prediction of HD.

2. Literature Review

Survey and clinical records used as healthcare data sources are prone to noise and biased information as well as measurement error. These aspects are sources of anomalous data points that can shape the learning process of ML algorithms. The traditional methods of anomaly detection normally eliminate the detected outliers prior to training. Nonetheless, straightforward deletion policies can cause deletion of valuable information. Moreover, even after initial filtering, noise might still be present in the data. Besides outliers, the imbalance between classes plays an important role in the assessment of the ML algorithms (Begum et al., 2026). The dataset commonly has minority classes that are underrepresented with the HD patients thus classifiers are likely to learn non-significant patterns. Thus, there is a strong need to have a solid ML structure, which is sensitive to outliers, noisy data, and class imbalance to enhance predictive accuracy. The study has shown that ML frameworks can be utilized to properly analyze health data in order to make precise predictions and help medical decision-making. Nonetheless, the study has noted that data quality, noise, and lack of sufficient data are still significant threats to health care predictive modeling (Mim et al., 2025). The DEGAN-CS (Data-enhanced Generative Adversarial Network for Code Search) approach emphasizes the significance of improving the quality of the dataset by applying data augmentation techniques based on GANs (Hu et al., 2026). It shows the need to focus on the importance of the data-driven approach, which inspires the application of GAN-based techniques within the proposed framework.

The proposed framework by the authors (El-Sofany et al., 2024) includes the prediction of HD based on various methods of feature selection with ML algorithms to identify the significant risk factors influencing cardiovascular diseases. Their findings provided better prediction accuracy, in terms of making use of the best feature subsets. Nevertheless, the research primarily concerned the relevance of features and did not include issues regarding noisy healthcare data and outliers. Similarly, the authors (Majhi and Kashyap, 2024) proposed explainable artificial intelligence on HD prediction by integrating ensemble ML models with SHAP explanations to enhance the interpretability. Although the methodology increased transparency in model predictions, it did not examine challenges associated with the problem of class with unbalanced and data anomalies that are prevalent in real-world healthcare datasets. In this study (Zhou et al., 2024), examined a number of architectures such as convolutional neural networks (CNN), recurrent neural networks (RNN), and hybrid deep learning (DL) models. The authors emphasized that DL models have good predictive capabilities yet they usually need sizeable datasets and substantial computer power. Jiang et al. (2025) created a hybrid optimization-based DL framework that combines optimization algorithms and neural networks to detect HD in order to improve prediction capabilities. Although the strategy failed to incorporate effective preprocessing procedures to deal with aberrant samples, they discovered that their results had greater model accuracy.

Gupta et al. (2025) have compared various ML algorithms are LR, RF and SVM in the prediction of HD. The research had the conclusion that ensemble models are better in classification. Nevertheless, the authors have mentioned that preprocessing methods of noisy data is still a significant problem in healthcare

predictive modeling. In order to enhance cardiovascular disease classification models, Alwakid et al. (2025) looked into feature selection and forward selection. The study highlighted the significance of interpretability, bias mitigation, and transparency in healthcare ML systems. The framework addressed issues with noisy data and anomaly detection, it did not increase feature relevance. On the other hand, Alsabhan and Alfadhly (2025) compared different ML models in the prediction of HD based on clinical data. The findings showed that the ensemble models including RF and gradient boosting algorithms are effective compared to traditional classifiers. In spite of these advances, the study noted that dataset imbalance and data noise largely affect the reliability of the prediction. Pal et al. (2025) explored ensemble learning techniques for predicting cardiovascular illness and showed that a novel combination of classifiers improves prediction accuracy. However, the authors found out that in order to deal with imbalanced healthcare data and noisy data, unstable data processing techniques are needed.

Other researchers, such as Hassan et al. (2024), used ML models in the detection of coronary artery disease through clinical data and demonstrated good classification results. However, the studies primarily concerned comparison of algorithms and did not discuss the anomaly detection and reliability-conscious learning mechanisms. In this study proposed an optimistic Chaos Game Optimization (CGO-RNN) Recurrent Neural Network to predict HD at its onset. To reduce the computational complexity, the study used the Kernel Principal Component Analysis (KPCA) as a feature extraction method and dimensionality reduction. The proposed model had good prediction performance whose accuracy exceeded 98%. Nevertheless, the research was primarily concerned with optimization-based classification, and it lacked discussion on the issues associated with noisy observations, outliers, and class imbalance in medical records (Alam and Muqeem, 2024). Dubey et al. (2024) enhanced a hybrid deep learning system that combines the Crow Search Optimization Algorithm (CSOA) with deep neural networks (DNN) and CNN in predicting heart-related diseases. The Cardio dataset and NIH chest X-ray dataset experimental experiments were carried out, and in both experiments, CSOA was employed to optimize model performance. The findings were better classification performance attained at approximately 96% using Adam and Nadam optimizers. Nevertheless, the research primarily touched upon the optimization of deep learning and did not deal with such problems as noise, outliers, and imbalanced classes in health data. The dynamic method used by the proposed framework enhances prediction accuracy by selecting the classifiers that are more appropriate for that specific sample. In other words, intelligent model selection is critical in enhancing prediction accuracy with an ensemble model (Zhang et al., 2025). In this research work suggests a combination of MFS-DLPSO (Multi-feature selection Disjunctive particle swarm optimization) techniques and an XGBoost ML algorithm as the MFS-DLPSO-XGBoost model to predict cardiovascular diseases. This model resulted, for different metrics, in an accuracy of 74.7%, precision of 76.3%, recall of 71.4%, and F1-score of 73.6%. Furthermore, the model was able to reach an AUC of 80.8%, beating the XGBoost baseline model in a clear way (Cao et al., 2025). The authors developed an Adaptive Parallel Strategy Grey Wolf Optimizer using XGBoost (APSGWO-XGB) framework, which integrates feature selection along with the PSO-GWO based hyperparameter tuning technique to enhance the performance of heart disease prediction by achieving 94.2% accuracy on the Kaggle HD dataset. Nevertheless, the framework mainly concentrates on feature selection and optimization without considering any issues such as class imbalance and probability calibration (Elhadad et al., 2026).

Despite the fact that these studies prove the effectiveness of ML techniques in predicting cardiovascular disease, some significant constraints can be identified. To begin with, the majority of the current research is concerned with the comparison of algorithms or feature selection but not the occurrence of the noisy observations and the presence of anomalous samples in the healthcare data. Second, HD datasets are rarely studied with class imbalance, and more sophisticated generative models could impact the capability of the models to identify the minority cases correctly. Third, most of the existing literature lacks the use of

optimization methods to enhance hyperparameter tuning and model generalization. This could lead to a decline in predictive reliability of models used with real-world healthcare data with noise, outliers and imbalance. Hence, an effective ML system that can effectively work with outlier identification, data imbalance, and optimization of hyperparameters to enhance the predictive accuracy and dependability in HD prediction systems is necessary. To address the shortcomings of the existing literature, this work suggests the ROHG-BXG framework (Robust Outlier Handling with GAN and BOA-XGBoost) as the means of reliable prediction of HD. The proposed framework combines various elements to improve the standard level of data and model performance, unlike the traditional techniques using one preprocessing or classification method. Begum et al. (2026) developed a GAN-based XGBoost model in the previous research utilizing the PKIOHD and Framingham datasets. The previous model involved techniques for the GAN-based balancing of data, the elimination of outliers using the Z-score and IQR method, as well as optimization of hyperparameters through Optuna. The current ROHG-BXG framework builds on the previous study with some additions, such as the introduction of Isolation Forest for the detection of irregularities and a loss-based mechanism that minimizes the effect of unreliable data. Moreover, CGAN technology is employed for the conditional generation of data, and BOA is used for the hyperparameter optimization methods. In addition, this paper has evaluated the performance of probability calibration with the Brier Score and ECE methods and compared the costs of optimization techniques. Therefore, while both papers utilize the same datasets and the overall concept of the application of GAN-based balancing and XGBoost, which includes outlier handling, the current paper differs from the previous work in terms of the newly incorporated techniques, weighting mechanism, optimization strategy, and methods of evaluation.

The originality of the proposed framework is that dual outlier processing, reliability-sensitive data generation, and optimization-based classification are combined into one prediction framework. In particular, the model will initially use Isolation Forest to identify structural abnormalities in the data and filter abnormal values. A loss-based on-the-fly outlier weighting mechanism is then proposed to dynamically detect the unreliable training samples in the model learning. Moreover, CGAN is used to create simulated minority samples to reduce the imbalance of the classes. Lastly, the BOA would be employed in the optimization of the XGBoost classifier hyperparameters to improve the performance of the classification. The ROHG-BXG framework will promote a robust system, higher prediction reliability, and a more stable learning process through the incorporation of these elements in working with noisy and imbalanced healthcare data.

2.1 Contributions of the Research Study

Unlike the traditional ML models applied to predict HD presence, the proposed ROHG-BXG model is concerned with enhancing the accuracy of prediction in that it aims to eliminate several issues that the healthcare datasets in the problem are known to possess. Majority of the current studies focus mainly on the comparison of the classification algorithms or on the enhancement of the accuracy of the model without the explicit consideration of the noisy observations, outliers and imbalance between classes. The difference between the ROHG-BXG and these methods lies in the fact that it combines both a two-outlier treatment system, reliability-aware GAN-based data augmentation, and BOA-optimized XGBoost classification in a single pipeline. Also, in contrast to most of the past research, which solely employs measures of classification, the framework proposed in this research work integrates both discrimination and calibration performance, such as Brier Score and Expected Calibration Error (ECE) to determine the strength of the probabilities that have been forecasted. It is an integrated design that enhances the model robustness and gives stronger predictions to healthcare decision-support systems. Novel contributions of the proposed framework:

- In this study, a new hybrid framework called ROHG-BXG is introduced, which combines outlier detection at two stages, data balancing using CGAN, hyperparameter optimization via BOA, and final prediction through XGBoost all in one single workflow.
- A novel dual-stage outlier handling mechanism is introduced by combining Isolation Forest and on-the-fly loss-based outlier detection, enabling the framework to address both structural anomalies and unreliable training samples.
- This work uses a conditional data generation strategy based on CGAN to enhance the representation of the minority class after eliminating structural outliers. The method creates more realistic synthetic samples compared to the traditional oversampling techniques.
- The framework, which is proposed, does a detailed analysis based on both classification and calibration, offering a better measure of how well prediction can be done in clinical decision support systems.

3. Methodology

3.1 Data Description

The research employs the dataset available at the Kaggle notebook PKIOHD presented in **Table 1** below. The variables in the dataset divided into major categories. The Center for Disease Control (CDC) surveyed more than 400,000 people in the US. The health survey is based on dichotomous questions. The cleaning process led to the availability of 320,000 records that include 18 critical health factors; the majority of the data is unbalanced, and CGAN was used to balance the training data. The "Framingham" HD dataset has over 4,240 records with the inclusion of the target column, resulting in a total of 15 features. The dataset is used in predicting whether the patient is at 10-year risk of developing (CHD) coronary heart disease.

Imbalanced class analysis: The PKIOHD dataset is comprised of 263,179 instances from the majority class and only 24,636 instances from the minority class leading to an imbalanced ratio (IR) of 10.68:1. Similarly, in the Framingham dataset, the majority class comprises 2,852 instances and the minority class, 489 instances resulting in an imbalance ratio of about 5.83:1. The fact that there is a high degree of class imbalance in both these datasets may cause ML models to be skewed toward the majority class. This leads to lower predictive effectiveness on minority-class samples. That is why we made use of the CGAN technique to augment the minority-class samples and hence, bring better class balance that the ROHG-BXG model would train on after it is fed the balanced data detailed discussion about datasets are illustrated in the Result and Discussion (Section 6: **Tables 6** and **7**).

Table 1. PKIOHD and Framingham dataset description.

S. No.	(a) PKIOHD data features
Demographic factors	
1.	Age category: Classification into different age groups.
2.	Sex: Gender.
3.	Race: Racial background.
Behavioral risk factors	
4.	Smoking: History of smoking at least 100 cigarettes in a lifetime.
5.	Alcohol Drinking: Regular consumption of alcohol.
6.	Physical Activity: Frequency of engaging in physical activities outside work.
Clinical factors	
7.	BMI: calculated using height and weight.
8.	Diabetic: diabetes diagnosis history.
9.	Gen Health: Self-reported overall health status.
10.	Physical Health (PH): No. of days in the past month with poor physical health.
11.	Mental Health (MH): No. of days in the past month with poor mental health.
12.	Diff Walking: Difficulty walking or climbing stairs.
13.	Stroke: Diagnosed with a stroke (History).
14.	Sleep Time: Average no. of hours slept per day.

Table 1 continued...

15.	Asthma: diagnosed with asthma (History).
16.	Kidney Disease: Kidney Disease History.
17.	Skin Cancer: diagnosed with skin cancer (History).
S. No.	(b) Framingham data feature
Demographic factors	
1.	Age: years.
2.	Sex: gender
3.	Education: Educational attainment level of every person.
Behavioural risk factors	
4.	Current Smoker: Indicates whether the individual is a current smoker.
5.	Cigarettes Per Day: Average number of cigarettes smoked per day.
Clinical factors	
6.	Body Mass Index (BMI): calculated using height and weight.
7.	Blood Pressure Medication (BPMeds): Indicates whether the individual is taking medication for blood pressure
8.	Prevalent Hypertension: History of being diagnosed with hypertension.
9.	Diabetes: diabetes diagnosis (History) .
10.	Total Cholesterol: Total cholesterol level measured in mg/dL.
11.	Systolic Blood Pressure (SBP): Measurement in mmHg.
12.	Diastolic Blood Pressure: Measurement in mmHg.
13.	Heart Rate: Resting heart rate measured in beats per minute.
14.	Glucose: Blood glucose level measured in mg/dL.
15.	Prevalent Stroke: diagnosed with a stroke (History).

3.2 Data Preprocessing

The experiments were conducted on the both datasets Prior to training the models, the data were preprocessed, which consisted of categorical encoding and feature scaling. The data set has categorical and numerical attributes. Encoding was used to encode categorical variables. The Min-Max scaling was used to normalize numerical features so that all variables have the same range. Assume that the input dataset is in the form of Equation (1):

$$X = \{X_1, X_2, \dots, X_n\} \quad (1)$$

where, X_i is a representation of sample i . Once it has been preprocessed and scaled, it is transformed into the trained data of a model.

3.3 Isolation Forest for Outlier Detection

To enhance the quality of data, Isolation Forest was used to identify structural abnormalities of the data. Isolation Forest separates the anomalies by dividing the feature space randomly. Statistical aberration of a sample x is defined as in Equation (2):

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (2)$$

where, $h(x)$ represents the path length of sample x . $c(n)$ represents an average path length of an isolation tree. The samples that attain high scores on anomaly are treated as outliers and are eliminated in the training dataset (Chugh et al., 2025). The Isolation Forest algorithm is utilized prior to CGAN to eliminate only those structurally anomalous instances which could affect the learning process of the generator. The use of the algorithm ensures that the data representing the distribution of meaningful instances is preserved, while at the same time minimizing the impact of the noise instances.

3.4 On-the-Fly Loss-Based Outlier Detection

Healthcare data are full of noisy observations and the existence of anomalous samples which adversely impact on the model evaluation. In order to minimize the impact of this type of samples, a loss-based

adaptive weighting mechanism is presented. The first step is to train the initial LR model to predict the probability of prediction of every sample (Lam et al., 2026). The cross-entropy loss of the individual sample is given as in Equation (3). In the proposed on-the-fly loss-based outlier detection module, LR is used as an auxiliary model to estimate sample-wise prediction probabilities, then calculate cross-entropy loss. The loss values computed are converted to reliability weights and then fed into the training of the BOA optimized XGBoost classifier. LR is not used as the final prediction model it can be used only in the preprocessing stage to estimate the efficiency of the reliability. The reason for using LR as a supplementary reliability model is that it yields sample-based probabilistic outcomes, permitting the straightforward computation of cross-entropy loss. The purpose of LR in this phase is not delivering the final classification of the data but giving a straightforward and interpretable probability-based benchmark for estimating the reliability of the training samples. Thus, the samples that experience high prediction loss are deemed less trustworthy, lowering their reliability weights. The derived weights are then utilized in the training process of the XGBoost classifier optimized using BOA. The fact that LR allows for a linear probabilistic model does not mean that XGBoost's ability to learn and deal with nonlinear problem-solving is confined since LR is aimed solely at producing reliability estimates while the classification is performed by XGBoost. Hence, LR is focused on providing a viable computation of reliability while XGBoost establishes complex nonlinear correlations in the data presenting HD.

$$L_i = -\sum y_i \log(p_i) \quad (3)$$

where, y_i the true label and p_i is the predicted probability. Exponential moving average of the loss is expanded as in Equation (4):

$$\mu_t = (1 - \alpha)\mu_{t-1} + \alpha L_t \quad (4)$$

where, α is the smoothing factor. The deviation of each sample is determined using the z-score in the Equation (5) (Andrade, 2021).

$$z_i = \frac{L_i - \mu}{\sigma} \quad (5)$$

where, σ represents the standard deviation of losses. A reliability weight is then assigned to each sample using the error function in the Equation (6) (Lam et al., 2026).

$$w_i = 0.5 \left(1 - \operatorname{erf} \left(\frac{z_i - \tau}{\sqrt{2}} \right) \right) \quad (6)$$

Samples with the high deviations with lower weights, decrease their impact during training. As such, the reliability weights generated by the released auxiliary LR model define the way unreliable samples are weighted but not the final class prediction. The final classification will be performed by the optimized XGBoost classifier which ensures that the final model correctly captures the nonlinear dependencies among the clinical and demographic features.

3.5 CGAN for the Class Balance

In order to improve its resistance to the issue of noisy and unreliable samples, the suggested framework makes use of a CGAN. In contrast to the traditional Generative Adversarial Networks (GANs), CGAN is a data generation model that uses the information about the class labels as one more condition of the data generation process. This enables the generator to get to know the conditional data distribution so that the generated samples are related to certain class labels. A CGAN is composed of two competing neural networks, one of them being a generator G and a discriminator D . The aim of the generator is to generate conditioned synthetic samples on the label of the class. y , as the discriminator analyzes the authenticity of a sample in regard to whether it is generated or real. The adversarial learning method allows the generator

to acquire meaningful data distribution representations. The objective role of the CGAN may be specified as in Equation (7):

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} \left[\log \left(1 - D(G(z|y)) \right) \right] \quad (7)$$

In which x represents is actual data samples, z is random noise, and y is the conditional label data. This adversarial learning process helps CGAN to acquire the underlying conditional data distribution, and generate more organized and realistic samples, so that the proposed ROHG-BXG framework is more robust and ready to classify (Goodfellow et al., 2020).

3.6 BOA for the Tuning Hyperparameters of the XGBoost Model

BOA helps in optimizing the hyperparameters of the proposed model classifier to increase model performance. BOA is a swarm-based optimization method that has a swarm-based inspiration that is based on the foraging behavior of the butterfly, its parameters illustrated in **Table 2**. The intensity of fragrance of each butterfly is determined as in the Equation (8) (Arora and Singh, 2019).

$$f_i = cI^a \quad (8)$$

where, c is the sensory modality, I is the stimulus intensity, and a is a power exponent. The position update rule is given by in the Equation (9)

$$x_i^{t+1} = x_i^t + r^2(g^* - x_i^t)f_i \quad (9)$$

where, x_i^t represents the current solution, g^* is the global best solution, and r is a random number. BOA iteratively searches for optimal hyperparameter values such as learning rate, tree depth, and regularization parameters.

Table 2. BOA parameters.

Parameter	Value	Description
Population size	5	Number of butterflies
Iterations	8	Number of optimization cycles
Sensory modality (c)	0.01	Controls fragrance intensity
Power exponent (a)	0.1	Fragrance scaling factor
Switching probability (p)	0.8	Controls global/local search

Table 3. Model hyperparameters search space.

Hyperparameter	Description	Search range
n_estimators	No. of boosting trees	100 – 400
max_depth	Maximum depth of DT	3 – 10
learning_rate	Step size shrinkage	0.01 – 0.20
subsample	Training samples used per tree	0.70 – 1.00
colsample_bytree	Features used per tree	0.70 – 1.00
gamma	Minimum loss reduction required for splitting	0 – 3
min_child_weight	Minimum weight required in child nodes	1 – 8
reg_alpha	L1 term	0 – 5
reg_lambda	L2 term	0 – 5

XGBoost, which builds a set of DT through gradient boosting. XGBoost objective function is given as XGBoost is a type of ML model that is defined by the objective function is illustrated in the Equation (10).

$$L = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k) \quad (10)$$

where, l represents the training loss and Ω is the regularization term controlling model complexity. By combining BOA-based optimization with XGBoost classification, the proposed framework improves prediction performance while preventing overfitting and the Model hyperparameters are illustrated in the Table 3.

The BOA Algorithm was chosen due to its balanced capacity of exploration and exploitation, thus providing an opportunity for efficient global search as well as minimizing chances for premature convergence to local optimum. Moreover, BOA showed its efficiency in dealing with complex nonlinear optimization problems. The aim of the current study was to examine the efficiency of the ROHG-BXG framework through hyperparameter optimization based on the BOA algorithm. The comparative analysis with other optimization methods, including Classic Bayesian Optimization and Optuna, also evaluated.

4. Proposed ROHG-BXG Framework

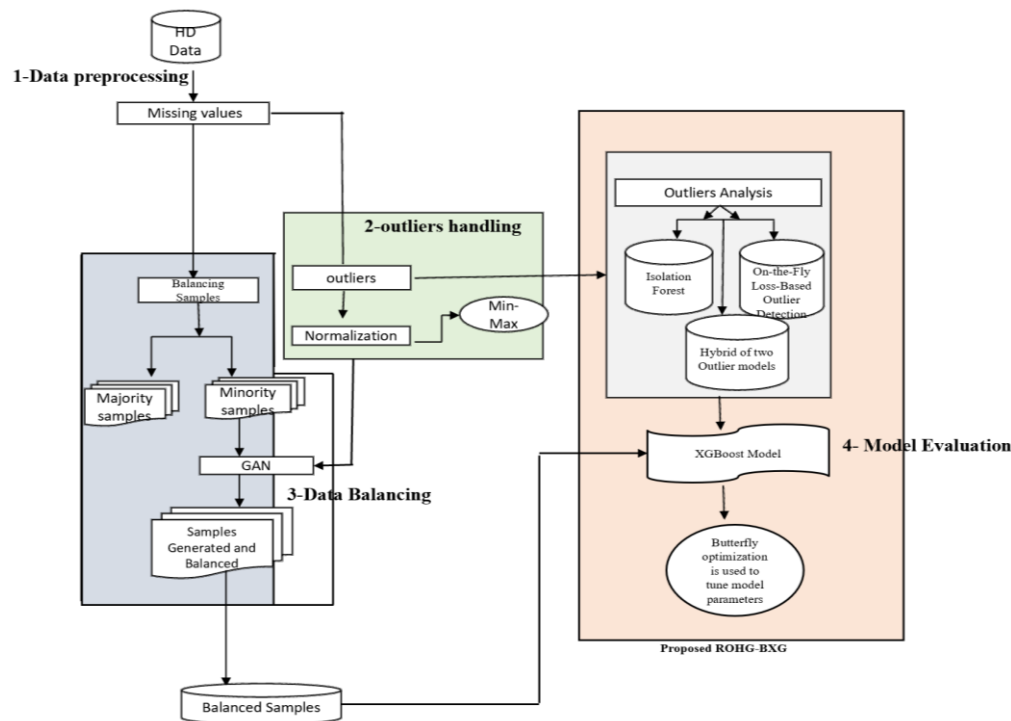


Figure 1. Overall framework of ROHG-BX.

The suggested ROHG-BXG framework will enhance the credibility of the prediction of HD in the case of noisy and imbalanced healthcare data. Medical data tends to have anomalous cases and class distribution that might adversely interfere with ML model performance. In order to overcome these issues, the ROHG-BXG framework combines the strong preprocessing, generative data augmentation, and optimization-based classification into a single pipeline. The first phase is known as preprocessing and encoding of the dataset to be trained on the model. Isolation Forest is utilized to detect structural outliers and drop anomalous samples in order to enhance data quality. A loss-based on-the-fly outlier detector is then used to weight training samples adaptively, giving less weight to samples that are unreliable. A CGAN is used in order to alleviate the imbalance in the classes to create synthetic minority samples, which has the effect of enhancing

the classifier to identify HD cases. The BOA will be used to tune the hyperparameters of the XGBoost classifier to enhance predictive performance after balancing the dataset. Lastly, the XGBoost is optimized to predict HD. The general steps of the ROHG-BXG framework involve pre-processing, dual outlier management, GAN-based oversampling, BOA optimization, and classification, which can be used to provide powerful and dependable prediction in healthcare data. Also, the suggested framework assesses probability calibration and classification performance. Although the majority of the available literature is dedicated to the classification measures. Moreover, the calibration curves are studied to explore the consistency of the results of the predicted probabilities and the real outcomes so that the model would provide the credible risk estimates in predicting HD the ROHG-BXG model is depicted in the **Figure 1**. and illustrated in the **Table 4** (Algorithm 1).

Table 4. Proposed ROHG-BXG framework.

Algorithm 1: ROHG-BXG Framework for HD Prediction		
Require: Preprocessed HD dataset, normalized features, training–testing split (80:20)		
1.	Load the processed heart disease dataset containing patient features and class labels.	
2.	Dataset partition into training and testing.	
3.	Normalize the feature values to maintain consistent scaling.	
4.	Train a Logistic Regression model using the training data.	
5.	Obtain prediction probabilities for each training sample.	
6.	Compute the cross-entropy loss for each sample.	
7.	Estimate the mean and variance of the loss distribution using an exponential moving average.	
8.	Compute the Z-score for each sample based on the deviation from the loss distribution.	
9.	Assign reliability weights so that samples with large deviation have less influence during training.	
10.	Train a CGAN using minority class samples.	
11.	Generate additional samples conditioned on class labels to improve class balance.	
12.	Combine the original samples and generated samples to form an augmented dataset.	
13.	Initialize the Butterfly Optimization Algorithm to search for suitable parameters of the XGBoost classifier.	
14.	Evaluate candidate parameter sets using model performance measures.	
15.	Best parameter configuration that produces the better performance.	
16.	Train the XGBoost classifier using the optimized parameters and augmented dataset.	
17.	Apply the trained model to the testing dataset.	
18.	Evaluate the model using metrics	
Output: Predicted HD risk with probability values.		

5. ROHG-BXG Model Metrics

The proposed ROHG-BXG evaluated using several performance metrics, including classification and calibration measures illustrated in the **Table 5**.

Table 5. Metrics and their formulas.

S. No.	Metrics	Formula
1.	Accuracy	$\frac{TN+TP}{TP+TN+FP+FN}$
2.	Recall	$\frac{TP}{TP+FN}$
3.	Specificity	$\frac{TN}{TN+FP}$
4.	Precision	$\frac{TP}{TP+FP}$
5.	F1-Score	$\frac{2TP}{2TP+FP+FN}$
6.	NPV	$\frac{TN}{TN+FN}$
7.	PPV	$\frac{TP}{TP+FP}$

Besides these classification measures, there are ROC-AUC that are applied to assess the performance of discrimination. To determine the reliability of prediction, Brier Score and ECE are estimated, which give the information about the calibration of the predicted probabilities. Analysis is also made of the calibration curves to analyze the correspondence between the estimated occurrences and real events.

6. Results and Discussion

The dual outlier detection is a process that enhances the quality of the training data used as it deals with various forms of anomalies. Isolation Forest is used in the first step to detect and eliminate global structural outliers, which are samples that are substantially different than the general data distribution. Nevertheless, there are still some samples that can be normal statistically and have abnormal behavior when training the model. To overcome this, the second step uses an on the fly loss based detection scheme that considers the training loss of each sample and gives a low weight value to samples that have abnormally high loss values. This method enables the model to minimize the effect of hard or noisy samples without necessarily barring the possible useful information. Consequently, the dual stage structure enhances the stability of models, and improves the quality of data and the ability to learn with large scale healthcare datasets. The dataset partition and its outlier removal process is illustrated in the **Tables 6** and **7**.

Table 6. Summary of dataset partitioning and outlier detection results for PKIOHD data.

Stage	Dataset size
Original dataset	3,19,795
Before isolation forest	2,55,836
Outliers removed by IF	3838
Outliers removed by on-the-fly	7485
Balanced size	324513

Table 7. Summary of dataset partitioning and outlier detection results for Framingham data.

Processing step	Number of samples
Original Dataset Size	4240
Missing Values After Imputation	0
Samples Before Isolation Forest	3392
Outliers Removed by isolation Forest	51
Training Samples After Isolation Forest	3341
Outliers Removed by on the fly	87
Balanced Training Dataset (After CGAN)	5704

The BOA convergence behavior applied in the proposed ROHG-BXG is shown in **Figure 2**. The convergence curve illustrates how the optimization process is gradually being improved with a series of successive steps. Within the first few iterations (0-4), the algorithm has a relatively stable objective value of about 0.9121, which shows the exploration phase during which the candidate solutions are analyzed. The objective value starts accelerating with an observable peak at about iteration number 5 as the algorithm switches to exploitation as more preferable parameter configurations are found. Once this stage is reached, the curve plateaus with only slight oscillations, this means that the algorithm has settled down to an optimal solution. The last version exhibits a minor increase, which proves the stability and effectiveness of the optimizing process. This convergence effect shows that BOA can effectively explore the solution space and optimized the hyperparameters of the suggested ROHG-BXG framework and, thus, help to develop predictive capabilities and model stability in HD prediction.

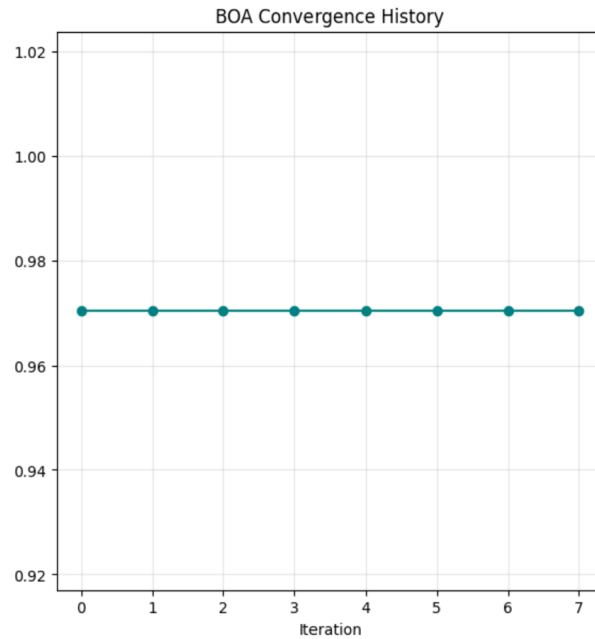


Figure 2. ROHG-BXG convergence.

Table 8 shows the relative performance of the various XGBoost settings with various outlier handling methods. The findings show that a strong outlier detection mechanism contributes to a highest stability and predictability of models. The baseline models that employed Isolation Forest and on-the-fly outlier detection both yielded similar results and reached an accuracy of 91% but the model proposed ROHG-BXG framework presents a better overall performance with 92% accuracy and F1-score with 89%. Moreover, the proposed model is calibrated to have a high probability with a Brier score of 0.0655 and ECE of 0.032 and a strong predictive reliability as indicated by PPV (89%), and NPV (93%) on PKIOHD data. These results affirm that the method of CGAN-based data balancing combined with the hybrid outlier removal enhances the suppleness and the generalization power of the XGBoost model, which makes the ROHG-BXG framework more effective in HD prediction than the base methodologies additionally evaluated on Framingham dataset also it is achieved better performance compared with other models.

Table 8. Comparative results of outlier handling methods in XGBoost models for the PKIOHD dataset.

Model configuration	Balancing technique	Outlier removal	Accuracy	F1-Score	Precision	Recall	AUC-ROC	Brier score	ECE	PPV	NPV
XGBoost with Isolation Forest	GAN	Isolation Forest	91%	89%	89%	92%	84%	0.0657	0.034	89	93
XGBoost with on the Fly	GAN	on the fly	91%	89%	89%	91%	83%	0.0688	0.039	89	92
ROHG-BXG (Proposed)	GAN	Isolation Forest+ On the fly	92%	89%	89%	91%	83%	0.0655	0.032	89	93

Table 9 presents comparisons of three hyperparameter optimization algorithms: the Bayesian Optimization, Optuna, and the ROHG-BXG algorithm. The proposed algorithm, ROHG-BXG, was the one which provided the best results for Accuracy (0.92%) and F1 score (0.91%), as well as the best Brier Score (0.0655). Bayesian Optimization is the one that gives the lowest ECE of (0.0108), but the overall

comparison results showed that the proposed ROHG-BXG algorithm managed to maintain its superior classification accuracy while being robust for the PKIOHD dataset.

Initially, to evaluate outlier removal techniques (Isolation Forest and On-the-Fly outlier detection), the model was evaluated using them separately one. They were to see the effect of each individually on the final result. A combination of the two approaches was decided based on the findings. After hybrid outlier removal combined Isolation Forest and On-the-Fly techniques in such a manner that the removal of noisy examples at the same time leaving good examples, the final model (ROHG-BXG) gave best overall results in the terms of prediction accuracy and model reliability then applied. Evaluated the efficacy of the BOA optimization algorithm by benchmarking it against the Optuna and Standard Bayesian Optimization algorithms on the PKIOHD and Framingham datasets. The new method turned out to outperform all others across a variety of performance measures, thus confirming that the method is efficient and applicable for HD prediction.

Table 9. Comparative performance of hyperparameter optimization techniques for PKIOHD dataset.

Hyperparameter optimization method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC AUC	Brier score	ECE
Bayesian Optimization	0.9161	0.89	0.91	0.88	0.83	0.0679	0.0108
Optuna	0.9163	0.89	0.91	0.88	0.83	0.0678	0.036
ROHG-BXG(Proposed)	0.92	0.89	0.89	0.91	0.83	0.0655	0.032

Table 10 highlights the performances of Optuna, Standard Bayesian Optimization, and the proposed model-based optimization technique as hyperparameter optimization methods in terms of several metrics of classification and discrimination. This proposed method showed the best results with an accuracy of 0.9557% and an F1-score of 0.9510%. Moreover, it had the lowest Brier Score (0.0391) and ECE (0.0217), making the technique to achieve higher performance when it came to both classification and probability calibration.

Table 10. Comparative performance of hyperparameter optimization techniques for Framingham dataset.

Metrics	Optuna	Standard Bayesian optimization	ROHG-BXG (Proposed)
Accuracy (%)	0.9527	0.9545	0.9557
Precision (%)	0.9813	0.9962	0.9850
Recall (%)	0.9228	0.9123	0.9193
F1 Score (%)	0.9512	0.9524	0.9510
Specificity (%)	0.9825	0.9965	0.9860
AUC	0.9831	0.9830	0.9833
Brier Score	0.0395	0.0390	0.0391
Calibration Slope	0.0840	0.0962	0.0960
ECE	0.0230	0.0203	0.0217

The comparative analysis of both data samples indicates that the suggested hyperparameter optimization based on the BOA leads to an overall well-balanced better performance when compared to Optuna and Standard Bayesian Optimization. While in some cases the optimization algorithms achieved higher results on certain metrics, the proposed model showed better results across multiple metrics. As a result, the proposed BOA algorithm not only improves classification performance but also enhances the credibility of the probabilities, making it a powerful and stable hyperparameter optimization method for HD prediction.

Figure 3 demonstrates the ROC and the calibration curve of the tested models and shows that the tested models may be characterized by a comparatively predictive nature of the baseline XGBoost settings and

the proposed ROHG-BXG framework. All the ROC curves show that all models have a high discriminative ability, and the value of the AUC exceeds 0.83, which means that all models are highly sensitive to predict heart disease. Although the baseline models based on Isolation Forest and an-impromptu outlier detection demonstrate competitive performances, the suggested ROHG-BXG model has a similar discriminative ability with the added advantage of greater robustness due to hybrid outlier treatment. The suggested ROHG-BXG model demonstrates the consistency of probability estimation which shows the predictive consistency even when the model operates on noisy or skewed data. Combining data balancing using GAN with two-stage outlier filtering, the presented framework minimizes the impact of anomalous samples and enhances the stability of models. In general, the visual analysis by ROHG-BXG model shows that there is a trade-off between the level of discrimination and the level of calibration, which is balanced and supports the ROHG-BXG model as effective to predict HD robustly. The proposed ROHG-BXG framework outperformed then the XGBoost model optimized by the Bayesian Optimization algorithm by Budholiya et al. (2022) which yielded accuracy of 91.8% on the Cleveland dataset with the added advantage of dual outlier detection and probability calibration. The proposed framework aims to enhance the prediction reliability by employing class balancing using CGAN and the hyperparameter optimization using BOA, rather than focusing on classification accuracy as in previous methods.

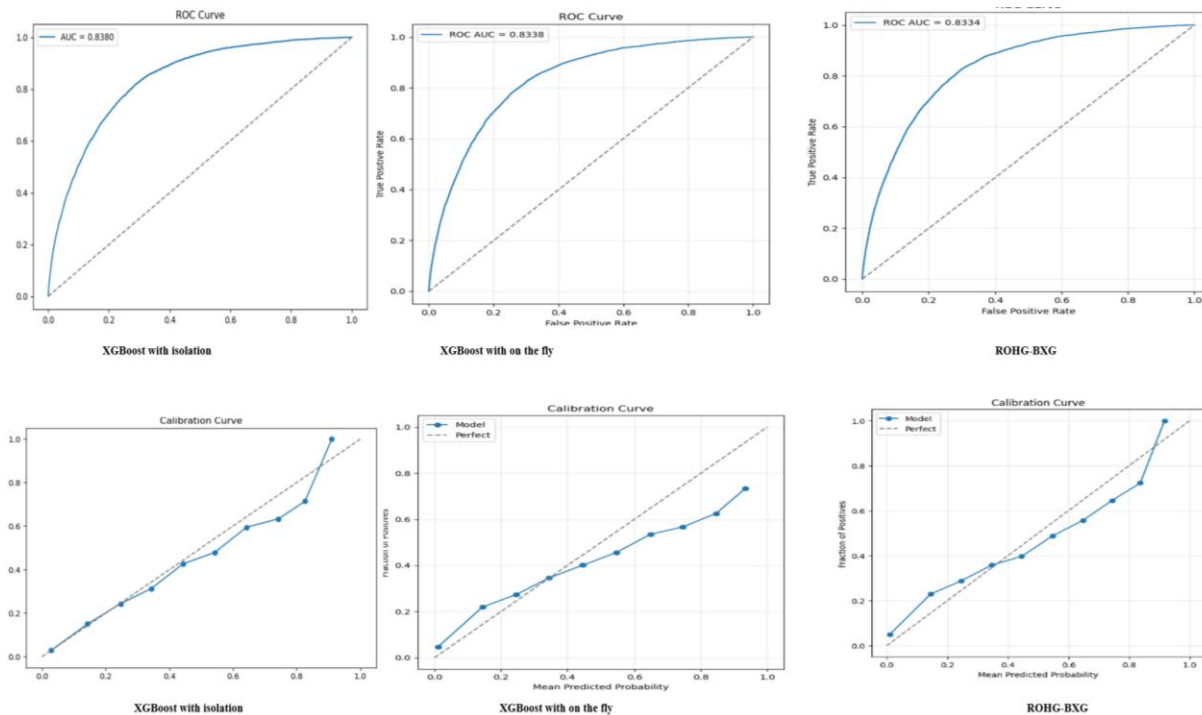


Figure 3. ROC curve's and calibration slope for the PKIOHD data.

Conducting the experiment, a computer system with an 11th Generation Intel® Core™ i5-11300H processor running at a frequency of 3.10 GHz and 16 GB RAM was used. The system had an integrated Intel® Iris® Xe Graphics and had a storage capacity of 477 GB. The experimental results show that the ROHG-BXG framework was able to complete the training period in 13 minutes on the PKIOHD dataset and in 9 minutes on the Framingham dataset. For prediction only 8 and 5 seconds were needed, respectively. Although training process requires additional computational efforts due to CGAN-based data generation and BOA-based hyperparameter optimization, these operations are done offline. After the classifier is

trained an optimized XGBoost classifier allows for quick predictions and shows the proposed framework is computationally efficient for practical HD prediction applications illustrated in **Table 11**, and the time comparison of baseline optimization techniques and the proposed are used illustrated in the **Table 12**.

Table 11. Execution time of the proposed ROHG-BXG framework.

Dataset	Training time	Prediction time
Proposed model on PKIOHD	13 min	8s
Proposed model on Framingham	9 min	5s

Table 12. Computational cost comparison on PKIOHD data.

Method	Computational cost (seconds)
Bayesian Optimization	120
Optuna	95
ROHG-BXG	180

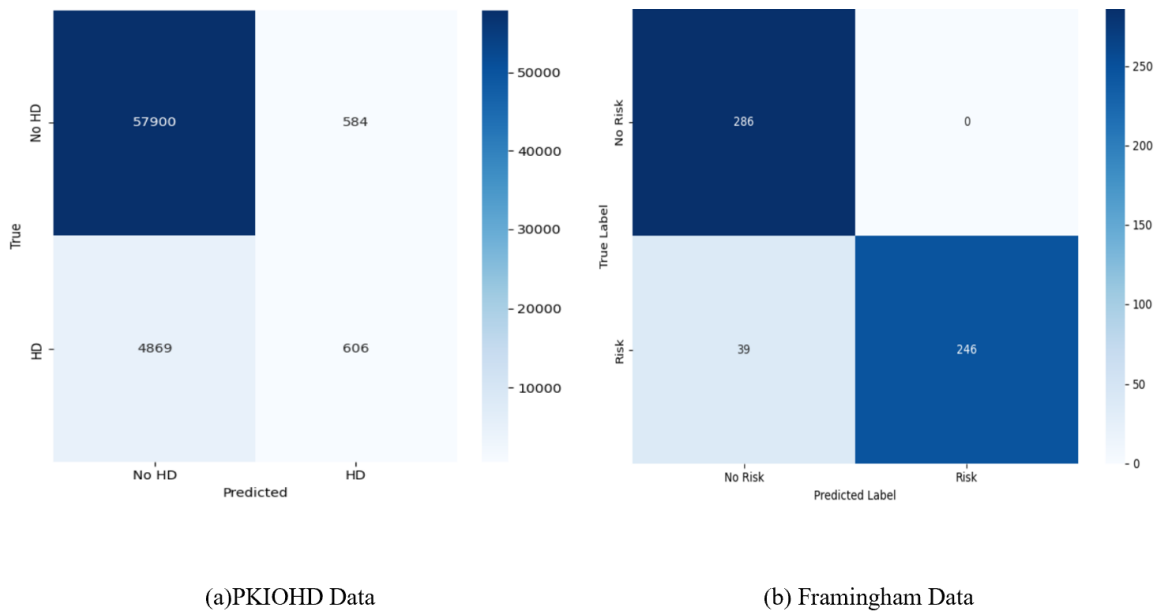


Figure 4. Confusion matrices of the proposed model.

Figure 4 depicts the confusion matrices of our ROHG-BXG model on the PKIOHD dataset and Framingham dataset. The confusion matrices indicate that the model can successfully distinguish and predict HD versus non-HD cases as the number of correct predictions is high with relatively few misclassifications. Classification accuracy of the PKIOHD dataset is very good, which can be seen from the confusion matrix, whereas the confusion matrix of the Framingham dataset also shows very good classification between the two classes. It is worth mentioning that such robustness and efficiency on HD prediction have been proven through different types of datasets by the developed ROHG-BXG model.

SHapley Additive exPlanations (SHAP) method is used to identify the main features that contribute to the predict HD. This demonstrates the capability of the model to pick up and represent clinically significant risk factors. In the PKIOHD dataset Sex, Age Category, General Health, Trouble Walking, and Diabetes

are the most important factors whereas in Framingham dataset the major predictors are Cigarettes per Day, Age, Heart Rate, Glucose, and Systolic Blood Pressure illustrated in the figure. The SHAP value above zero means a higher risk of HD, while a negative SHAP value would lead to a lower prediction level. These results indicate that the suggested ROHG-BXG model delivers reliable, transparent, and intelligible decision outcomes by clearly indicating the influence of each feature on the final decision, as illustrated in **Figure 5**.

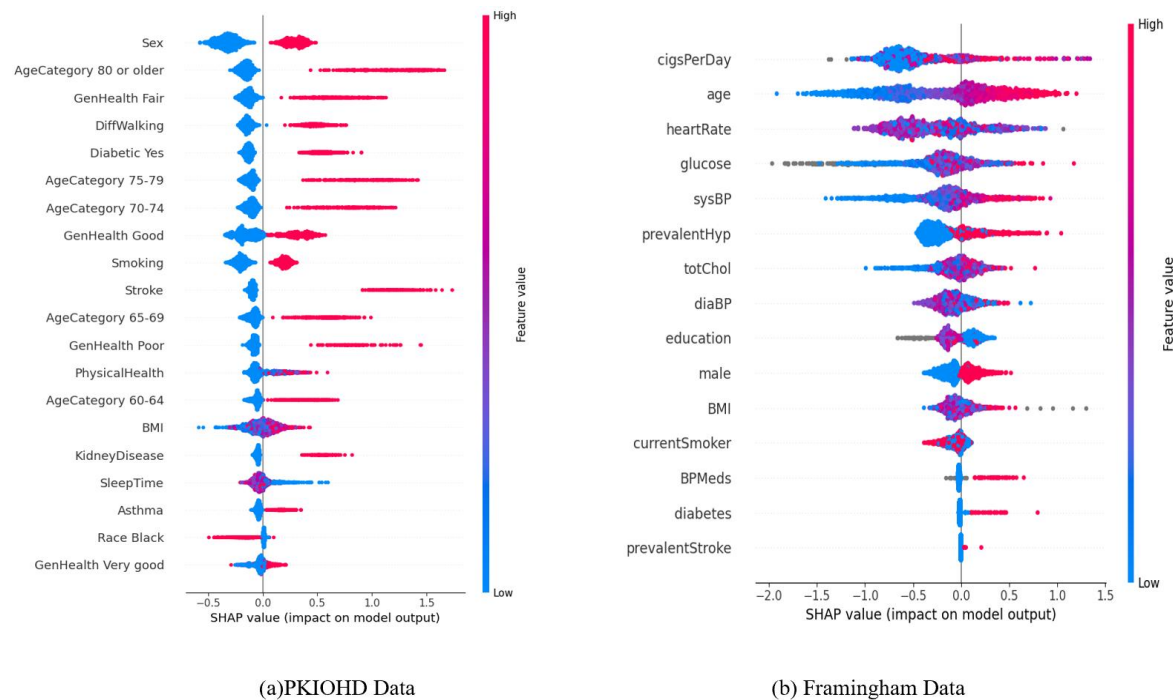


Figure 5. SHAP summary plots of the ROHG-BXG model.

Figure 6 illustrates the SHAP feature importance rankings derived from our proposed ROHG-BXG model applied to the PKIOHD and Framingham datasets. The mean absolute SHAP values, in this case, measure how much each feature contributed, on average, to the final result of predicting HD. The Sex feature and Age Category stand out as having the greatest impact in determining the outcome of a PKIOHD dataset prediction whereas, in the case of Framingham dataset, it has been found that Cigarettes Per Day along with Age make the biggest differences to prediction output. This result highlights that through model explanation techniques like SHAP, the proposed model can identify the key clinical parameters and demographic factors that have the greatest effect on prediction performance.

Table 13 conducts a comparative review of available methods of HD prediction and proposed model. Many studies provided insights ML and optimal methods to enhance the effectiveness. In this study (Chen et al., 2024) applied an Adversarial Generative Network (AGAN) with a fractional gradient descent and Swish activation to missing data imputation on both UCI and PKIOHD datasets, but did not provide the main metrics of evaluation (F1-score, calibration metrics, and predictive values). Budholiya et al. (2022) used Bayesian optimization with XGBoost on the Cleveland dataset and got an accuracy of 91.8%, but the evaluation metrics were not elaborated. In this study (Manikandan et al., 2024), LR with and without Boruta feature selection was compared on the Cleveland dataset, achieving an F1-score of 89.23 and an accuracy

of 88.52, but no sophisticated preprocessing techniques, including data balancing or outlier detection, were included. Similarly, Sharma and Agarwal (2024) explored the use of RF, SVM, and XGBoost models using Optuna-based hyperparameter optimization that demonstrated accuracy values of approximately 88% though it did not discuss the issue of class imbalance or robust anomaly treatment.

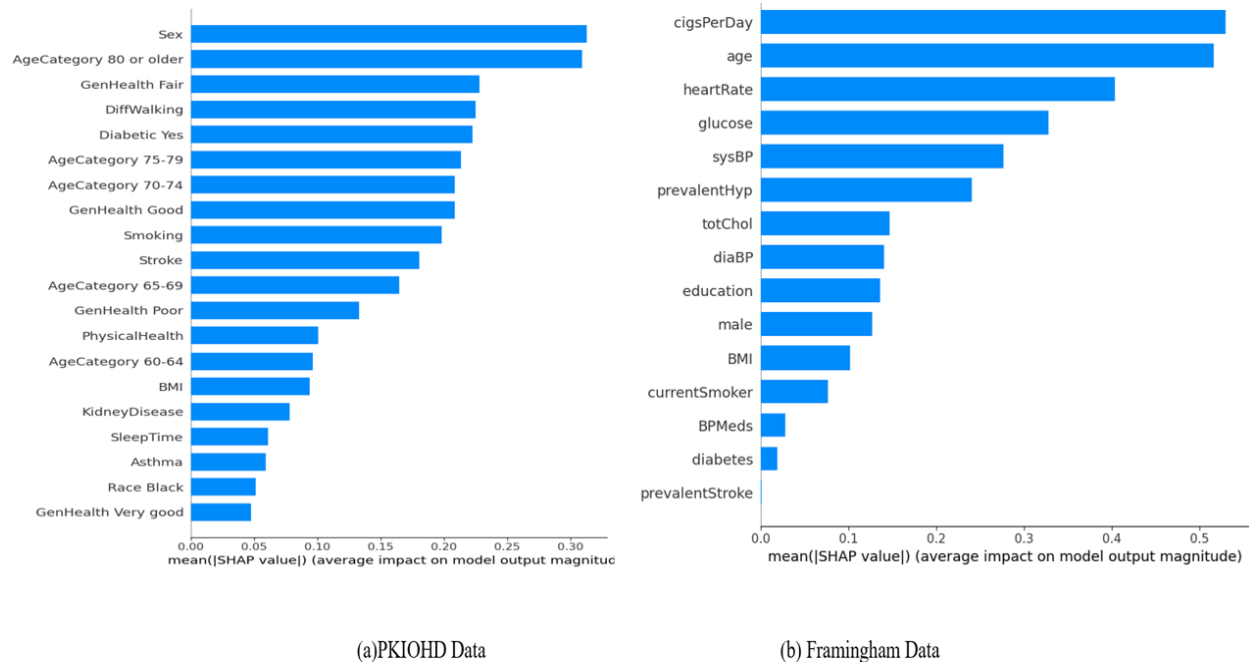


Figure 6. Global SHAP feature importance of the proposed ROHG-BXG model.

Other methods like (Srinivas and Katarya, 2022) hyOPTXg that used Optuna optimization had F1-scores of up to 90.38 and (Deng et al., 2025) LightGBM with SMOTE-based balancing and SHAP analysis with 90.61% accuracy on the BRFS dataset. Also, Mandula and Kumar (2024) proposed an ensemble model that involves ANOVA and LASSO feature selection with adaptive boosted decision fusion, with a performance of 88-89. More recently, Sasirekha and Kanisha (2025) theorized the Adaptive Synthetic Ensemble Balancer (ASEB), which combined SMOTE and GAN-based balancing, except that it did not optimize hyperparameters but instead made modifications to the weights of the ensemble but less performance compared with the proposed model.

Conversely, the proposed model (2026) presents a holistic pipeline, which incorporates the CGAN-based data balancing scheme, dual outlier detection schemes, as well as BOA to tune the hyperparameters in the ROHG-BXG framework. In contrast to some of the past analyses that emphasized the single aspect of improvement like feature selection, balancing, or tuning, the suggested methodology considers the data imbalance issue, anomaly detection issue, and model optimization problem in one framework. Consequently, the model will attain 92% accuracy and 89% F1-score with the calibration values reported like Brier score (0.0655) and ECE (0.032) and are seldom reported in earlier research. Besides, the model exhibits high predictive reliability, having 93% PPV and 89% NPV on PKIOHD data, and also performed well on the Framingham dataset, with the highest accuracy of 0.9557 %, and with the lowest ECE then the proposed model performed well compare with existing model also explored with classification and calibration.

Table 13. Comparative analysis of proposed work with the existing work.

Year	Dataset	Data balancing	Outlier handling	Model parameter s tuning	Model	F1-Score	Accuracy (%)	Brier score	EC E	NP V	PP V
2024	HD & PKIOHD	X	X	Fractional GD + Swish Activation	AGAN is Used for Missing Data Imputation	X	X	X	X	X	X
2022	Cleveland HD	X	X	Bayesian Optimizati on	Optimized XGBoost	X	91.8	X	X	X	X
2024	Cleveland	X	X	X	LR with /without Boruta	89.23	88.52	X	X	X	X
2024	Kaggle HD	X	Standard	Optuna	RF, SVM, XGBoost	87 (all)	87.84, 88.07, 88.03	X	X	X	X
				Without Optuna	RF, SVM, XGBoost	86.32, 88, 86.16	87.05, 88.01, 86.87				
2022	Heart Failure HD UCI kaggle Dataset	X	X	Optuna	hyOPTXg	89.43	89.34	X	X	X	X
						90.38	88.5	X	X	X	X
2025	HD Dataset (BRFSS 2015; 253,682 Samples)	SMOTE	Standard	Grid Search, SHAP	LightGBM	X	90.61	X	X	X	X
	Framingham dataset					X	85.26	X	X	X	X
2024	HD Data	X	X	X	ALAN (ANOVA + LASSO) ET-ABDF (Ensemble Technique with Adaptive Boosted Decision Fusion)	89	88	X	X	X	X
2025	HD Dataset (BRFSS 2015; 253,682 Samples)	SMOTE+GAN	X	X	ASEB (Adaptive Synthetic Ensemble Balancer) only Modifies Ensemble and Class Weights; It Does Not Tune Hyperparameters.	85	85	X	X	X	X
Proposed Model (2026)	PKIOHD	GAN	Dual Outlier Handling	BOA	ROHG-BXG	89	92	0.0655	0.032	89	93

Generally, the proposed approach provides a better, more highly calibrated prediction model than the available ones because it employs sophisticated methods for data preprocessing, optimization, and probabilistic calibration assessment, which contribute to greater reliability in HD risk prediction. The most important novelty of the work is the introduction of a two-stage hybrid approach to outlier avoidance, aimed at calibrating the confidence of the last iteration to improve the model's dependability and its predictive accuracy. In contrast to the traditional methods, which use one preprocessing mechanism, the suggested framework will initially remove the global anomalies by applying Isolation Forest to remove the samples with structural inconsistencies to the training set. Then it implements an on-the-fly, loss-based detection scheme during model training, dynamically detecting samples exhibiting abnormal learning behavior and down-weighting them rather than simply discarding them. Such a twofold mechanism allows the model to handle hard outliers and training-time anomalous samples, enhancing learning stability. In addition, the framework includes calibration of model confidence at the end of the iterations so that the model predicted probabilities are more consistent with the actual predictive reliability of the model. The hybrid outlier mitigation, combined with adaptive training-time filtering and confidence calibration, represents one single pipeline, which makes predictions in large-scale datasets of HD more robust and credible. Overall, the comparative study shows that the proposed ROHG-BXG model is a more complete solution than current HD prediction models. In contrast to existing works that typically focused on feature selection, class balancing, or hyperparameter optimization separately, the proposed framework is able to combine all of them together into a single framework: dual outlier detection, data balancing using CGAN and hyperparameter optimization using BOA, and probability calibration. This combination enhances the prediction accuracy, robustness of the models, and reliability of probability estimation, making the proposed framework more applicable to practical healthcare applications.

Despite all the above-stated steps, such as dual-stage outlier detection, CGAN, and BOA, which are included in the design of the proposed algorithm, they are all applied to the model only in its offline version. In fact, once the model is trained, HD predictions can be performed using only the optimized XGBoost classifier, making the inference phase very efficient. Besides structural outliers, class imbalance, noisy observations, healthcare data sets may also raise such issues as overlapping class distributions, high-dimensional feature spaces, and small disjuncts. Overlapping classes can be defined as different classes having very similar attributes, resulting in the classification of one instance belonging to both classes. To avoid the curse of high-dimensional data, one of the common methods used to reduce the data size is to apply techniques that reduce the dimensionality of the feature space to the minimum needed to still capture most of the variation in the data and thereby improve the computation speed and model generalization capability. Small disjuncts refer to very small groups within a larger dataset that the ML models tend to ignore due to the insufficient information they can get out of those groups. The PKIOHD and Framingham datasets in the present study have only 17 features and 15 features, respectively, so the issue of high dimensionality is not that critical. The main focus of the present investigation is to propose a ROHG-BXG model capable of overcoming the prevalent data quality problems in the datasets, which are structural outliers, noisy observations, and class imbalance problems.

7. Conclusion and Future Work

In this research study, a hybrid ROHG-BXG model is used to predict HD with high reliability using healthcare data that can be noisy, contain outliers, be skewed, and have an uneven class distribution. The suggested framework incorporates dual outlier management, CGAN-generated synthetic data, and BOA-optimized XGBoost classification within a single pipeline. The Isolation Forest algorithm is applied to detect structural anomalies in the dataset, and a loss function based on the fly-weighting mechanism minimizes the contribution of untrustworthy training samples. Moreover, the CGAN model produces minority-class samples, which increases class balance and improves the model's learning. BOA is used to

optimize the XGBoost classifier by tuning hyperparameters to enhance prediction. The proposed framework is experimentally evaluated to demonstrate that it achieves high classification performance compared with several current ML methods. Moreover, probability calibration measures such as Brier Score and ECE are included, which are more informative about prediction reliability. Altogether, the ROHG-BXG framework should be seen as a strong and valid tool of predicting HD and could be used to assist clinical decision-making systems within health care setting. Future research may focus on advancing DL architectures and investigating transformer-based models to strengthen feature extraction and representation learning. Additionally, hybrid optimization strategies, such as combining arithmetic optimization with slime mould algorithms, could be explored to improve global search performance and tackle complex design optimization challenges more effectively. Advanced technologies in healthcare data collection can be explored in future studies to enhance the quality and variety of clinical data. Data from Internet of Things (IoT) based wearable devices, smartwatches, wireless body sensor networks, mobile health (mHealth) applications, electronic health records (EHRs) and remote patient monitoring (RPM) can give the real-time physiological and lifestyle data. The use of these technologies in the proposed ROHG-BXG framework could allow the prediction of HD in real-time, individual healthcare, and better clinical decision support.

Conflicts of Interest

There are no financial or personal connections that might be perceived as having influenced the findings or interpretations presented in this research.

Acknowledgments

The authors thank the editor and anonymous reviewers for their insightful feedback, which enhanced the manuscript's quality. The authors affirm that this work was not funded.

AI Disclosure

The original manuscript was also reviewed in its entirety by the authors who carried out the required changes without AI tools.

Author Contributions

Shaik Sajeera Begum: Data Curation, Resources, Investigation, Formal Analysis, Conceptualization, Methodology, Project Administration, Prepared Original Draft, Review and Editing, Validation. **Syed Javeed Pasha and Sanjay Dhanka:** Conceptualization, Methodology, Supervision, Computational, Formal Analysis, Validation, Investigation, Project Administration, Review, and Editing.

Data Availability

In this research study, an open-source dataset has been used, which is freely available on Kaggle at the following link: https://www.kaggle.com/code/alawdisoft/personal-key-indicators-of-heart-diseases/notebook?select=heart_2020_cleaned.csv Kaggle and <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset?select=framingham.csv?select=framingham.csv>

Ethical Approval and Informed Consent for Data

There are no ethical issues related to this dataset.

References

- Alam, A., & Muqem, M. (2024). An optimal heart disease prediction using chaos game optimization-based recurrent neural model. *International Journal of Information Technology*, 16, 3359-3366. <https://doi.org/10.1007/s41870-023-01597-w>
- Alsabhan, W., & Alfadhly, A. (2025). Effectiveness of machine learning models in diagnosis of heart disease: a comparative study. *Scientific Reports*, 15, 24568. <https://doi.org/10.1038/s41598-025-09423-y>
- Alwakid, G., Ul Haq, F., Tariq, N., Humayun, M., Shaheen, M., & Alsadun, M. (2025). Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective. *BMC Cardiovascular Disorders*, 25(1), 123. <https://doi.org/10.1186/s12872-025-04550-w>

- Andrade, C. (2021). Z scores, standard scores, and composite test scores explained. *Indian Journal of Psychological Medicine*, 43(6), 555-557. <https://doi.org/10.1177/02537176211046525>
- Arora, S., & Singh, S. (2019). Butterfly optimization algorithm: a novel approach for global optimization. *Soft Computing*, 23(3), 715-734. <https://doi.org/10.1007/s00500-018-3102-4>
- Begum, S.S., Swamy, A., Dhanka, S., Abu-Shareha, A.A., Khan, M., & Husain, I. (2026). Generative adversarial networks and hyperparameter-optimized XGBoost for enhanced heart disease prediction. *Scientific Reports*, 16(1), 11326. <https://doi.org/10.1038/s41598-026-40322-y>
- Budholiya, K., Shrivastava, S.K., & Sharma, V. (2022). An optimized XGBoost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University Computer and Information Sciences*, 34(7), 4514-4523. <https://doi.org/10.1016/j.jksuci.2020.10.013>
- Cao, K., Liu, C., Yang, S., Zhang, Y., Li, L., Jung, H., & Zhang, S. (2025). Prediction of cardiovascular disease based on multiple feature selection and improved PSO-XGBoost model. *Scientific Reports*, 15(1), 12406. <https://doi.org/10.1038/s41598-025-96520-7>
- Chen, X., Zhang, N., Yang, X., Wang, C., Na, Q., Luan, T., Zhu, W., Zhang, C., & Yang, C. (2024). Cardiac disease diagnosis based on GAN in case of missing data. *PloS One*, 19(11), e0292480.
- Chugh, B., Malik, N., Gupta, D., & Alkahtani, B.S. (2025). A probabilistic approach driven credit card anomaly detection with CBLOF and isolation forest models. *Alexandria Engineering Journal*, 114, 231-242. <https://doi.org/10.1016/j.aej.2024.11.054>
- Deng, L., Lu, K., & Hu, H. (2025). An interpretable LightGBM model for predicting coronary heart disease: enhancing clinical decision-making with machine learning. *Plos One*, 20(9), e0330377. <https://doi.org/10.1371/journal.pone.0330377>
- Dubey, A.K., Sinhal, A.K., & Sharma, R. (2024). Heart disease classification through crow intelligence optimization-based deep learning approach. *International Journal of Information Technology*, 16(3), 1815-1830. <https://doi.org/10.1007/s41870-023-01445-x>
- Elhadad, A., Rayan, A., & Taloba, A.I. (2026). A novel APSGWO-XGBoost framework for heart disease detection: enhancing diagnostic accuracy and efficiency in healthcare. *Journal of Radiation Research and Applied Sciences*, 19(1), 102258. <https://doi.org/10.1016/j.jrras.2026.102258>
- El-Sofany, H., Bouallegue, B., & El-Latif, Y.M.A. (2024). A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Scientific Reports*, 14(1), 23277. <https://doi.org/10.1038/s41598-024-74656-2>
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144. <https://doi.org/10.1145/3422622>
- Gupta, A., Singh, A., Kumar, P., Raj, G., & Tomar, V. (2025). Analysis of machine learning techniques for prediction of cardiac diseases. *Procedia Computer Science*, 259, 1937-1946. <https://doi.org/10.1016/j.procs.2025.04.149>
- Hassan, M.M., Zaman, S., Rahman, M.M., Bairagi, A.K., El-Shafai, W., Rathore, R.S., & Gupta, D. (2024). Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. *Computers and Electrical Engineering*, 115, 109130. <https://doi.org/10.1016/j.compeleceng.2024.109130>
- Hu, H., Liu, J., & Fang, M. (2026). DEGAN-CS: an efficient code search model based on data-enhanced optimization of generative adversarial networks. *Pattern Recognition*, 172(Part D), 112709. <https://doi.org/10.1016/j.patcog.2025.112709>
- Jiang, W., Zheng, L., Tian, Y., & Kho, Y.H. (2025). A novel research on a heart disease prediction model based on hybrid intelligent optimization of deep networks. *Discover Networks*, 1(1), 6. <https://doi.org/10.1007/s44354-025-00008-9>

- Lam, T.C.W., O'Neill, N., Schran, C., & Schaaf, L.L. (2026). Cutting through the noise: on-the-fly outlier detection for robust training of machine learning interatomic potentials. *Machine Learning*. arXiv preprint.
- Majhi, B., & Kashyap, A. (2024). Explainable AI-driven machine learning for heart disease detection using ECG signal. *Applied Soft Computing*, 167(Part A), 112225. <https://doi.org/10.1016/j.asoc.2024.112225>
- Mandula, A., & Kumar, B.S.Y. (2024). Integrated feature selection and ensemble learning for heart disease detection: a 2-tier approach with ALAN and ET-ABDF machine learning model. *International Journal of Information Technology*, 16(7), 4489-4503. <https://doi.org/10.1007/s41870-024-02016-4>
- Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A.L., Manikandan, R., & Gandomi, A.H. (2024). Classification models combined with Boruta feature selection for heart disease prediction. *Informatics in Medicine Unlocked*, 44, 101442. <https://doi.org/10.1016/j.imu.2023.101442>
- Mim, T.J.H., Hassan, M.M., Paul, B.K., & Mollah, M.S.H. (2025). Machine learning approaches for cardiovascular disease prediction: a comparative study. *Biomedical Materials & Devices*, 1-15. (In press). <https://doi.org/10.1007/s44174-025-00564-2>
- Mishra, I., & Mohapatra, S. (2023). An enhanced approach for analyzing the performance of heart stroke prediction with machine learning techniques. *International Journal of Information Technology*, 15(6), 3257-3270. <https://doi.org/10.1007/s41870-023-01321-8>
- Pal, P., Singh, H.V., Grover, V., Manikandan, R., Karimi, R., & Khishe, M. (2025). Interactive cardiovascular disease prediction system using learning techniques: insights from extensive experiments. *Results in Control and Optimization*, 19, 100560. <https://doi.org/10.1016/j.rico.2025.100560>
- Pandey, A., Shivaji, B.A., Acharya, M., & Mohbey, K.K. (2025). Mitigating class imbalance in heart disease detection with machine learning. *Multimedia Tools and Applications*, 84(30), 36497-36522. <https://doi.org/10.1007/s11042-024-19705-8>
- Sasirekha, R., & Kanisha, B. (2025). Adaptive ensemble framework with synthetic sampling for tackling class imbalance problem. *Engineering Reports*, 7(4), e70109. <https://doi.org/10.1002/eng2.70109>
- Sharma, P., & Agarwal, S. (2024). Cardiovascular disease analysis using different machine learning techniques. In *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things* (pp. 1-6). IEEE. Namakkal, India. <https://doi.org/10.1109/AIMLA59606.2024.10531431>
- Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control*, 73, 103456. <https://doi.org/10.1016/j.bspc.2021.103456>
- World Health Organization (2025). Cardiovascular diseases (CVDs). [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- Zhang, Z.-L., Zhu, Y.-H., & Luo, X.-G. (2025). DES-AS: dynamic ensemble selection based on algorithm Shapley. *Pattern Recognition*, 157, 110899. <https://doi.org/10.1016/j.patcog.2024.110899>
- Zhou, C., Dai, P., Hou, A., Zhang, Z., Liu, L., Li, A., & Wang, F. (2024). A comprehensive review of deep learning-based models for heart disease prediction. *Artificial Intelligence Review*, 57(10), 263. <https://doi.org/10.1007/s10462-024-10899-9>