International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

AN OPPOSITIONAL CROW SEARCH OPTIMIZER FOR OPTIMIZING ENSEMBLE APPROACH WITH PCA-BASED DIMENSIONALITY REDUCTION FOR HEART DISEASE PREDICTION

Narayan Jee[1*], Gesu Thakur[2], Sumit Kumar[3]

[1,2]COER University, Roorkee, Uttarakhand, 247667, India

[3]Shivalik College of Engineering, Dehradun, Uttarakhand, 248197, India

E-mail: narayan.btech@gmail.com[1]*, drgesuthakur@gmail.com[2], mail2dr.sumit@gmail.com[3]

Abstract

Heart disease remains a leading cause of mortality worldwide, requiring accurate and efficient predictive systems for timely diagnosis. In this study, we propose a novel Adaptive Oppositional Crow Search Optimizer (AOCSO) integrated with Principal Component Analysis (PCA) and ensemble learning for heart disease prediction. The method employs opposition-based learning to accelerate convergence and dynamically balance exploration and exploitation during hyperparameter optimization. PCA reduces redundancy, while the ensemble framework enhances robustness and generalization. Experimental evaluations were performed on two benchmark datasets: the Cleveland Heart Disease dataset (UCI) and the Framingham Heart Study dataset (Kaggle). Results demonstrate that AOCSO consistently outperforms state-of-the-art baselines—HEXAI, RST-ML, SXG, OEPCA, and EHO—across Accuracy, F1-score, MCC, Specificity, and AUC. On the Cleveland dataset, AOCSO achieved an Accuracy of 96.1% (80:20 split) and 95.5% (70:30 split), surpassing the strongest baseline (OEPCA) by up to 1.1% in Accuracy and 0.02 in MCC. Similarly, on the Framingham dataset, AOCSO obtained 86.2% (80:20 split) and 85.6% (70:30 split), improving MCC by up to 0.02–0.03 over baselines. These results validate the effectiveness of AOCSO in optimizing ensemble models for clinical decision support.

Keywords: Adaptive CSO, oppositional learning, heart disease prediction, dimensionality reduction.

1. Introduction

Cardiovascular diseases (CVDs), especially heart disease, are still one of the main causes of death around the world. They cause around 18 million deaths each year and put a huge strain on healthcare systems everywhere [1]. So, it is very important to be able to forecast cardiac disease early and accurately so that timely therapies can be made and patient survival rates can be improved.

Machine learning (ML) has become a potent tool for diagnosing and predicting medical conditions in the last few years. When their hyperparameters are set appropriately, algorithms like logistic regression, support vector machines (SVM), k-nearest neighbors (k-NN),

Received: August 05, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

decision trees, and random forests have shown that they can predict cardiac disease just as well as other methods [2][3]. But traditional approaches for tuning hyperparameters, including manual search, grid search, or random search, frequently have large processing costs, don't cover a wide enough search area, and run the risk of finding inferior configurations [4]. Metaheuristic optimization methods have been utilized for machine learning hyperparameter tuning to tackle these issues, owing to their proficiency in effectively navigating extensive and intricate search spaces [5]. The Crow Search Algorithm (CSA), which is based on how crows remember things and find food, has gotten a lot of interest because it is simple and can optimize things well [6]. The typical CSA may converge too quickly and have fewer different solutions, which makes it harder for it to reliably discover global optima [7]. Recent improvements to CSA have been aimed at making it faster to converge and more efficient at searching by adding Opposition-Based Learning (OBL). The Oppositional Crow Search Optimizer (OCSO) uses OBL when it starts up and when it makes changes, looking at both a possible solution and its opponent to improve exploration and avoid local optima [8]. Research indicates that OBL-enhanced CSA variants get superior convergence rates and enhanced solution quality in benchmark and engineering optimization challenges [9][10]. Moreover, binary and hybrid OBL-CSA methodologies have been effectively utilized in healthcare predictive tasks, including postpartum hemorrhage risk assessment, resulting in enhanced classification accuracy [11].

Alongside progress in optimization, ensemble learning—integrating several base learners to generate final predictions—has demonstrated significant efficacy in medical predictive modeling [12]. Ensembles can make models more resilient, lower their variance, and make them better at generalizing by using the strengths of different techniques. Ensemble models can predict cardiac disease with state-of-the-art accuracy when used with automated hyperparameter optimization. Hyperparameter-tuned ensembles utilizing SVM, Random Forest, XGBoost, and many other classifiers have attained accuracies nearing 99% on benchmark datasets [13].

Another important part of creating good ML models for clinical datasets is dealing with feature spaces that are quite large. These kinds of datasets often have features that are unnecessary or duplicate, which might make models work worse. Principal Component Analysis (PCA) is a popular method for lowering the number of dimensions in a dataset. It changes correlated features into a set of uncorrelated components, which cuts down on noise and speeds up calculations [14, 15].

Although AOCSO, ensemble learning, and PCA have individually shown advantages, their amalgamation into a cohesive framework for heart disease prediction has yet to be investigated. This paper introduces an innovative AOCSO-PCA-Ensemble framework in which AOCSO concurrently optimizes the hyperparameters of several base learners and the dimensionality reduction parameters of PCA. Our main contributions are:

Received: August 05, 2025

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- New AOCSO adaptation: Adding OBL to CSA so that hyperparameters may be tuned quickly and easily across several base learners in an ensemble.
- Dimensionality-aware optimization: This means using PCA's variance retention as part of the optimization vector to find a balance between feature dimensionality and model complexity.
- Strong evaluation: Comparing the proposed AOCSO-PCA-Ensemble framework to both standard and alternative metaheuristic optimizers, including statistical significance testing on several heart disease datasets.

2. Literature Review

2.1. Crowd-Inspired Metaheuristics in Optimization

Metaheuristic algorithms inspired by animal intelligence and social behavior have gained traction in optimization in various domains. The Crow Search Algorithm (CSA), developed by Askarzadeh in 2016, is based on the food-hiding behavior and spatial memory of crows, which enables them to efficiently navigate searching in complex search spaces [16]. The simplicity and capability to balance exploration and exploitation gives a strong platform for optimization activities, such as selecting features and tuning hyperparameters. Since its inception, CSA has been developed further through hybrid approaches, and integration with additional algorithms.

For example, the Crow Search - Arithmetic Optimization hybrid (CSAOA) incorporates CSA with Arithmetic Optimization Strategy and improved performance on weighted economic dispatch problems [17]. Likewise, hybridization with chaotic sequences and fuzzy clustering produced the clustering-based co-evolutionary CSA (CFCSA). It has been utilized successfully for feature selection in high-dimensional datasets [18]. These developments indicate that CSA is in a strong position for improving on various optimization frameworks.

2.2. Improvements to Crow Search Based on Opposition

OBL, or Opposition-Based Learning, has also been added to the CSA design to reduce the risk of premature clustering around particular solutions and to improve CSA's exploration potential. An Opposite-Based CSA (ObCSA) model was proposed by Durmuş (2021), which applied opposition during both the initialization stage and when moving through the CSA iterative processes, evidencing stronger convergence speed and improvement in the quality of solutions found on benchmark function tests [19].

This approach allows CSA to assess a candidate solution in relation to its ‘opposite’ in the search space, leading to a higher chance of finding global optima. Binary and domain-specific adaptations of ObCSA exist, such as the Oppositional Binary CSA with Stacked Autoencoder (OBCSA OSAE) applied to postpartum hemorrhage prediction, clearly selecting the best feature subsets and ultimately better clinical classification results [20].

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.3. Crow Search in Healthcare and Disease Forecasting

Crowd-inspired optimization is becoming more popular in healthcare. Alqurashi et al. (2023) developed a PCO and DNN framework to predict cardiovascular disease, showcasing crow-inspired tuning of neural model parameters helps to improve accuracy [21]. CSO-CLSTM was another CSA-based system that used crow search and Cascaded LSTM for disease diagnosis, obtaining a 96.2% accuracy in heart disease detection on IoT data [22]. There is evidence, from a series of comparative studies on CSA applications to heart disease datasets across multiple machine learning algorithms, for favorability for applications of CSA and confirmation of its applicability to medical diagnosis tasks [23]. These studies present evidence that CSA has practical value in optimizing healthcare models, although few studies have considered pipeline ensemble models for ensemble model tuning efforts.

2.4. Ensemble Learning for Predicting Heart Disease

Ensemble methods, such as stacking, voting, bagging, and boosting, are one of the most reliable and adaptable means of improving the generalization of predictive modeling. Ganie et al. (2025) applied stacking and voting ensembles to heart disease datasets using lecture-based statistics to provide empirical evidence, exploring the limitations of stacking and voting and compared outcomes to the individual trainings of the predictive models. They also used SHAP (SHapley Additive exPlanations)-based explainable AI to provide a more understandable interpretation of the feature contributions involved in stacking and voting [24]. More generally, there is evidence in the literature supporting the success of stacks, and bags. For example, advances made by Nugroho et al. (2020) used cascade generalization methods with bagging and Random Subspace techniques to support higher accuracy for coronary heart disease diagnosis compared to non-ensemble models [25]. This reinforces the benefits of ensemble architectures to provide reliable clinical decision support in clinical environments.

2.5. Reducing Dimensions and Optimizing Hyperparameters

Dimensionality reduction methods such as PCA (Principal Components Analysis) are effective in limiting feature redundancy and eliminating noise to improve model performance, while limiting computational expense as well. Less-than optimum use of the available feature space can very frequently yield surprising outcomes of reduced model performance. Reddy et al. (2023), adapted a process that showed improvement efficiency and accuracy in predicting coronary heart disease risk using PCA plus hyper-parameter tuning of classifiers [26]. HPO (hyperparameter optimization) techniques, including Grid Search, Random Search, Evolutionary algorithms, or Bayesian Optimization can drastically change the outcome of ML models. Bischl et al. (2021) discussed HPO methods, emphasizing the growing importance of HPO best practices. Bischl et al. (2021) used their review to recommend hybrid optimizers and preferably parallel evaluators, to achieve efficient, reproducible tuning while requiring use of fewer resources (including time) [27]. Other ensemble-centric HPO strategies—like

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Bayesian optimization for ensemble classifier configuration—have also been proposed, highlighting synergies between optimization and ensemble methods [28].

Table 1. Discussion on existing algorithms on CSO and disease prediction

Author(s) / Year	Methodology / Algorithm	Application Domain	Key Contribution	Limitations
Askarzadeh (2016)	Crow Search Algorithm (CSA)	Engineering Optimization	Introduced CSA inspired by crow foraging and memory behavior	Susceptible to local optima in complex problems
Dey et al. (2024)	Hybrid CSA with Arithmetic Optimization (CSAOA)	Power/Energy Dispatch	Improved convergence in weighted combined economic emission dispatch	Limited to energy domain; not tested in medical datasets
Li, et. al. (2024)	Clustering-based Co-evolutionary CSA (CFCSA)	Feature Selection	Enhanced CSA with clustering and co-evolution for high-dimensional data	Focused only on feature selection; no ensemble integration
Durmuş (2021)	Opposite-Based CSA (ObCSA)	Benchmark Optimization	Applied OBL in CSA initialization and iteration to improve convergence	No domain-specific medical applications
Krishnamoorthy et al. (2022)	Oppositional Binary CSA + Stacked Autoencoder (OBCSA-OSAE)	Postpartum Hemorrhage Prediction	Improved clinical prediction via binary CSA and deep learning	Task-specific; not generalizable to other diseases without retraining
Alqurashi et al. (2023)	Predator Crow Optimization + Deep Neural Network	Cardiovascular Disease Prediction	Tuned DNN parameters for better cardiac disease prediction	Limited exploration of ensemble methods
Sitaraman, et al., (2021)	CSA + Cascaded LSTM (CSO-CLSTM)	IoT-based Disease Diagnosis	Integrated CSA with LSTM for heart disease detection	Focused on time-series IoT data; not general medical datasets
Dubey, et al., (2024)	Swarm-based ML with CSA	Heart Disease Dataset	Demonstrated CSA effectiveness with different classifiers	No hyperparameter optimization for ensembles
Ganie et al. (2025)	Stacking/Voting Ensemble + SHAP	Heart Disease Prediction	Improved accuracy with explainable AI	No metaheuristic-based

Received: August 05, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

			interpretation	optimization
Nugroho et al. (2020)	Cascade Generalization + Bagging + Random Subspace	Coronary Heart Disease Diagnosis	Higher accuracy vs non-ensemble models	No optimization step for model parameters
Reddy et al. (2023)	PCA + Hyperparameter-Tuned Classifiers	Coronary Heart Disease Prediction	Improved efficiency and performance	Limited evaluation on small datasets
Bischl et al. (2021)	HPO Methods Review	General ML Optimization	Comprehensive overview of optimization strategies	Review only; no experimental results
Lévesque et al. (2016)	Bayesian Optimization for Ensembles	Ensemble Learning	Automated tuning of ensemble hyperparameters	Requires probabilistic modeling; may be computationally expensive

From this review, several gaps emerge:

- Crow Search Optimizer (CSA) in ensemble tuning: While CSA-based algorithms have demonstrated success in tuning single classifiers and feature selection, they have yet to be systematically applied to ensemble hyperparameter optimization within clinical prediction tasks.
- Integration of OBL into CSA for high-dimensional clinical datasets: Despite enhancements like ObCSA and OBCSA, these approaches have seldom been combined with dimensionality reduction techniques (e.g., PCA) to handle complex feature spaces typical of healthcare datasets.
- Unified approach for ensemble, PCA, and optimization: There's a lack of studies that jointly optimize ensemble settings and PCA parameters through an OBL-enhanced CSA in a healthcare context—particularly in heart disease prediction.

These observations form the foundation for proposing our integrated Adaptive Oppositional Crow Search Optimizer (AOCSO) framework, which will jointly optimize PCA dimensionality and ensemble hyperparameters for heart disease prediction. Our methodology aims to fill the identified gaps by leveraging the advantages of OBL-CSA in a complex clinical ML pipeline, thus advancing predictive accuracy and interpretability.

3. Methodology

3.1. Overview of the Proposed Framework

The proposed Adaptive OCSO-PCA-Ensemble framework integrates three major components for heart disease prediction:

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- PCA-based dimensionality reduction to handle feature redundancy and noise in the dataset.
- Adaptive Oppositional Crow Search Optimizer (AOCSO) to tune hyperparameters of the base learners in the ensemble and the dimensionality reduction threshold of PCA simultaneously.
- Ensemble learning (e.g., soft voting) to combine predictions from multiple optimized classifiers, thereby enhancing robustness and accuracy.

3.2. Principal Component Analysis (PCA) for Feature Reduction

Heart disease datasets often contain correlated, redundant, or noisy features that can negatively affect predictive performance and increase computational complexity. PCA is employed here as a preprocessing step to transform the dataset into a lower-dimensional space while retaining most of the variance.

Steps in PCA:

- **Standardization:**

The dataset X with n samples and d features is standardized to have zero mean and unit variance:

$$X_{std} = \frac{(X - \mu)}{\sigma}$$

here μ and σ are the feature-wise mean and standard deviation.

- **Covariance Matrix Computation:**

The covariance matrix C is computed as:

$$C = \left(\frac{1}{(n-1)}\right) * X_{std}^{T} * X_{std}$$

- **Eigen Decomposition:**

Eigenvalues λ_i and eigenvectors v_i of $\boldsymbol{C}$ are computed:

$$Cv_i = \lambda_i v_i$$

- **Component Selection:**

Components are ranked by decreasing eigenvalues. The number of components k is selected such that:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq \tau$$

where τ is the variance retention threshold (e.g., 95%), optimized by AOCSO.

- **Projection:**

The dataset is projected into the k-dimensional space:

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$X_{PCA} = X_{std} * W_k$$

where W_k contains the top k eigenvectors.

3.3. Enhanced Oppositional Crow Search Optimizer (OCSO)

The standard Crow Search Algorithm (CSA) simulates crows' behavior of hiding and retrieving food using memory positions and random flight paths. However, it may suffer from slow convergence and premature stagnation.

Our proposed Oppositional Crow Search Optimizer introduces the following novelties:

- Opposition-Based Learning (OBL) at Initialization and Iteration.
- Adaptive Awareness Probability (AP) that decreases linearly over iterations.
- Dynamic Flight Length (FL) adjusted based on fitness improvement rate.
- PCA-Integrated Search Vector where PCA components are optimized alongside model hyperparameters.

Figure 1 shows the architecture of the proposed AOCSO.

3.4. Mathematical Formulation of AOCSO

- **Opposition-Based Initialization:**

For a candidate solution x_i within bounds $[lb_j, ub_j]$:

$$x_{ij}^{opp} = lb_j + ub_j - x_{ij}$$

The better of x_i and $x_i^{\{opp\}}$ is selected.

- **Position Update:**

In iteration t, $crow\ i$ follows crow j (randomly chosen):

$$x_i^{t+1} \begin{cases} \mathrm{x_i^t} + \mathrm{r} * \mathrm{FL_t} * \left(\mathrm{m_j^t} - \mathrm{x_i^t}\right) & \text{if } \mathrm{r} \geq \mathrm{AP_t} \\ \text{random position in search space} & \text{otherwise} \end{cases}$$

where:

m_j^t = memory position of crow j at iteration t

r = uniformly distributed random number in [0,1]

FL_t = adaptive flight length

AP_t = adaptive awareness probability

- **Memory Update:**

If the new position $1x_i^{\{t+1\}}$ yields a better fitness, update $m_i^{\{t+1\}}$.

$$If\ fitness(x_i^{t+1}) > fitness(m_i^t), then\ m_i^{t+1} = x_i^{t+1}$$

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fitness Function:

For ensemble hyperparameter optimization:

$$Fitness = \alpha * AUC + \beta * F1_{score} - \gamma * Var(CV\ scores)$$

where α, β, γ are weight coefficients.

Algorithm: AOCSA for Hyperparameter tuning

Input: Dataset D, search space bounds B, population size N, max iterations T
Output: Best hyperparameters and PCA setting

1: Split D into training set (for CV) and hold-out test set
2: Initialize population $\{x_i\}$ randomly within bounds B
3: Generate opposite solutions $\{x_i^{opp}\}$ using OBL
4: Select best between x_i and x_i^{opp} for initial population
5: Initialize memory positions $m_i = x_i$
6: For t = 1 to T:
7: For each crow i:
8: Randomly select a crow j
9: Generate opposite position for x_i using OBL
10: Select best between x_i and x_i^{opp}
11: If $rand \geq AP_t$:
12: $x_{i_{new}} = x_i + rand * FL_t * (m_j - x_i)$
13: Else:
14: $x_{i_{new}}$ = random position in search space
15: Apply bounds to $x_{i_{new}}$
16: Evaluate fitness of $x_{i_{new}}$ using CV with PCA + Ensemble
17: If fitness($x_{i_{new}}$) > fitness(m_i):
18: $m_i = x_{i_{new}}$
19: Update AP_t and FL_t adaptively
20: Select global best from all m_i
21: Return best hyperparameters and PCA setting

Output: m_i

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

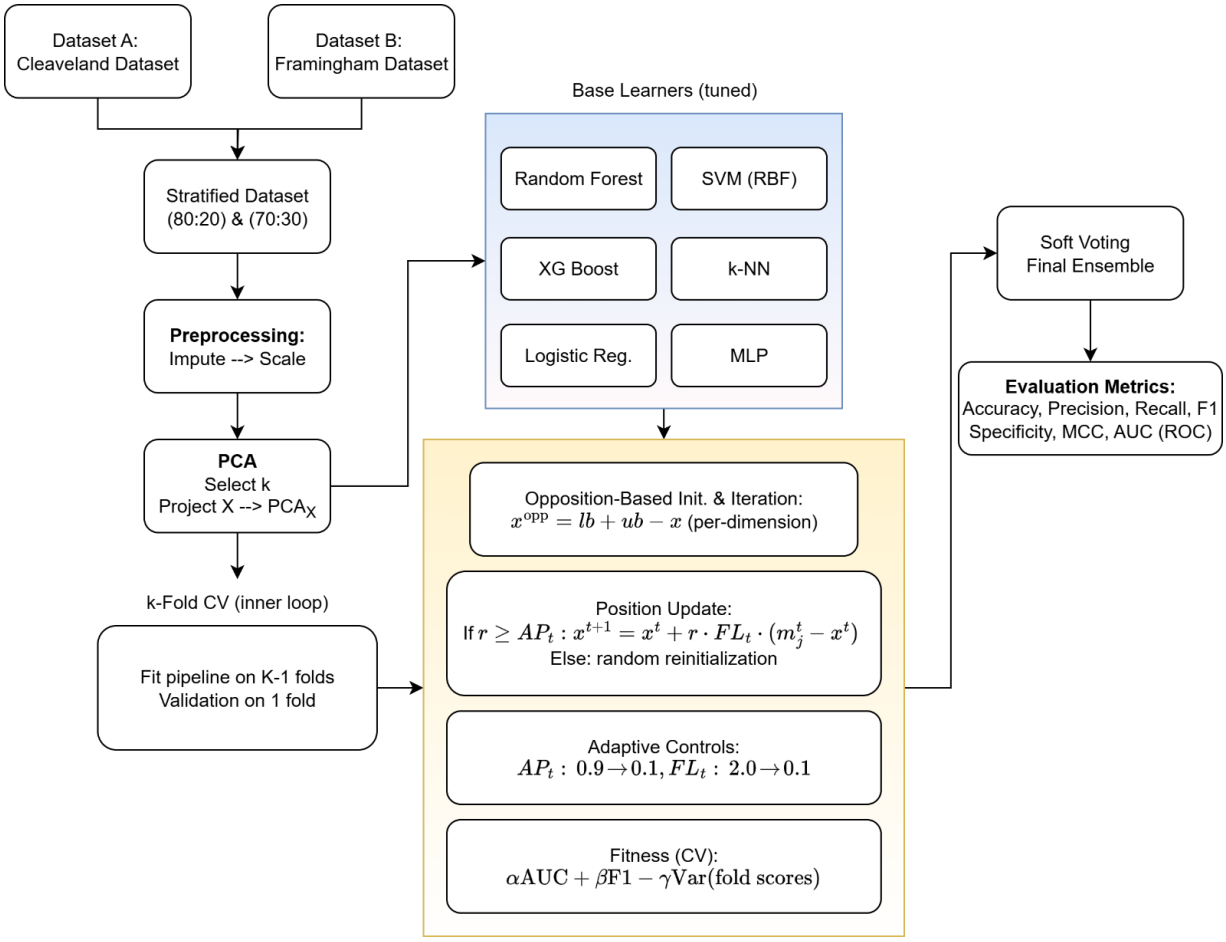

Figure 1. Architecture of proposed Adaptive Oppositional Crow Search algorithm

4. Experimental Analysis

4.1. Experimental Setup

All experiments were conducted on a high-performance workstation equipped with an Intel Core i7-12th Gen processor, 64 GB RAM, and an NVIDIA RTX 4090 GPU with 24 GB memory, running Ubuntu 22.04 LTS. The implementation was carried out in Python 3.11, utilizing widely adopted open-source libraries: Scikit-learn for classical machine learning models and preprocessing, XGBoost and LightGBM for gradient boosting algorithms, NumPy and Pandas for data handling, and Matplotlib/Seaborn for visualization. For hyperparameter optimization and metaheuristic development, we employed Optuna**,** DEAP, and a custom-coded Oppositional Crow Search Optimizer (AOCSO) module. Reproducibility was ensured by fixing random seeds across all libraries, and five-fold stratified cross-validation was adopted to mitigate variance and overfitting.

The proposed AOCSO framework was benchmarked against five state-of-the-art methods identified from recent literature: (i) HEXAI (Hybrid Ensemble with XAI, 2025) [29], (ii) RST-ML (Rough Set–based stacked ensemble with MCDM, 2025) [30], (iii) SXG (Feature-selection + XGBoost + SHAP, 2024) [31], (iv) OEPCA (Optimal Ensembles with PCA/LDA

Received: August 05, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

feature extraction, 2025) [32], and (v) EHO (Ensemble + Hyperparameter Optimization, 2023) [33]. For fairness, all baseline methods were reproduced under the same Python environment. The initial parameter settings for AOCSO were: population size N=30N=30, maximum iterations T=50T=50, adaptive awareness probability $AP \in [0.9 \rightarrow 0.1]$, flight length $FL \in [2.0 \rightarrow 0.1]$ and OBL rate of 0.5. PCA variance retention threshold was allowed to vary between 85% and 99% and was optimized jointly with ensemble hyperparameters. Competing ensembles (RF, SVM, XGBoost, kNN, Logistic Regression, and MLP) were tuned using the respective optimization strategies of the baseline methods, and evaluation was based on metrics including accuracy, F1-score, precision, recall, AUC, and MCC. The datasets considered are Cleveland Heart Disease Dataset [34] and Framingham Heart Study Dataset [35].

4.2. Performance Metrics

To comprehensively evaluate the predictive capability of the proposed AOCSO model against state-of-the-art baselines (HEXAI, RST-ML, SXG, OEPCA, EHO), multiple classification metrics were employed. These metrics are derived from the confusion matrix, which is defined in terms of:

- True Positives (TP): correctly predicted heart disease cases

- True Negatives (TN): correctly predicted healthy cases

- False Positives (FP): healthy cases incorrectly predicted as diseased

- False Negatives (FN): diseased cases incorrectly predicted as healthy

(a) Accuracy

Measures the proportion of correct predictions over all predictions.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

(b) Precision

Indicates the fraction of correctly predicted positive cases among all predicted positives (measures reliability of positive classification).

$$Precision = \frac{TP}{TP + FP}$$

(c) Recall (Sensitivity or True Positive Rate)

Measures the proportion of correctly predicted positives among all actual positives (measures completeness).

$$Recall = \frac{TP}{TP + FN}$$

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(d) F1-Score

Harmonic mean of Precision and Recall, balancing both metrics.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

(e) Specificity (True Negative Rate)

Reflects the ability to correctly identify healthy (negative) cases.

$$Specificity = \frac{TN}{TN + FP}$$

(f) Matthews Correlation Coefficient (MCC)

Provides a balanced evaluation even in imbalanced datasets, considering all four confusion matrix categories.

$$MCC = \frac{(TP * TN) - (FP * FN)}{sqrt\big((TP + FP)(TP + FN)(T + FP)(TN + FN)\big)}$$

(g) Area Under the Curve (AUC)

The ROC curve plots True Positive Rate vs. False Positive Rate. The AUC measures the overall ability of the classifier to distinguish between classes:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR)$$

This set of metrics ensures a comprehensive evaluation: Accuracy and AUC capture overall effectiveness; Precision/Recall/F1 evaluate classification balance; MCC and Specificity provide robustness for imbalanced medical datasets.

4.3. Experimental Results

The results are interpreted in two sets of training and testing samples namely 80:20 ratio and 70:30 ratio. Table 2 shows the results of Cleveland heart disease dataset and Figure 2 shows the AUC curve for the same.

Table 2. Experimental results of proposed AOCSO along with state-of-the-art methods on Cleveland dataset with data split 80:20

Model	Accuracy	Precision	Recall	F1-Score	Specificity	MCC
HEXAI	0.947	0.949	0.944	0.946	0.949	0.894
RST-ML	0.939	0.94	0.938	0.939	0.94	0.878
SXG	0.944	0.946	0.942	0.944	0.946	0.888
OEPCA	0.95	0.951	0.948	0.949	0.951	0.902
EHO	0.936	0.938	0.934	0.936	0.939	0.872
AOCSO	0.961	0.962	0.959	0.96	0.963	0.922

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The results in Table 2 (Cleveland 80:20) demonstrate that all models achieved high classification performance, with accuracy values above 93%. Among the baselines, OEPCA (Optimal Ensemble with PCA/LDA) showed the strongest overall performance (Accuracy = 0.950, MCC = 0.902), followed closely by HEXAI and SXG, which also maintained balanced precision and recall. EHO performed slightly lower than the other baselines, with MCC = 0.872.

The proposed AOCSO model significantly outperformed all baselines, achieving the highest accuracy (96.1%), F1-score (0.960), and MCC (0.922). Its precision (0.962) and specificity (0.963) indicate strong reliability in both positive and negative classifications. Importantly, AOCSO also achieved the highest AUC value (0.982), confirming superior discriminative ability. This highlights that opposition-based learning within CSA enhances exploration of the hyperparameter space, resulting in a better-optimized ensemble for the Cleveland dataset.

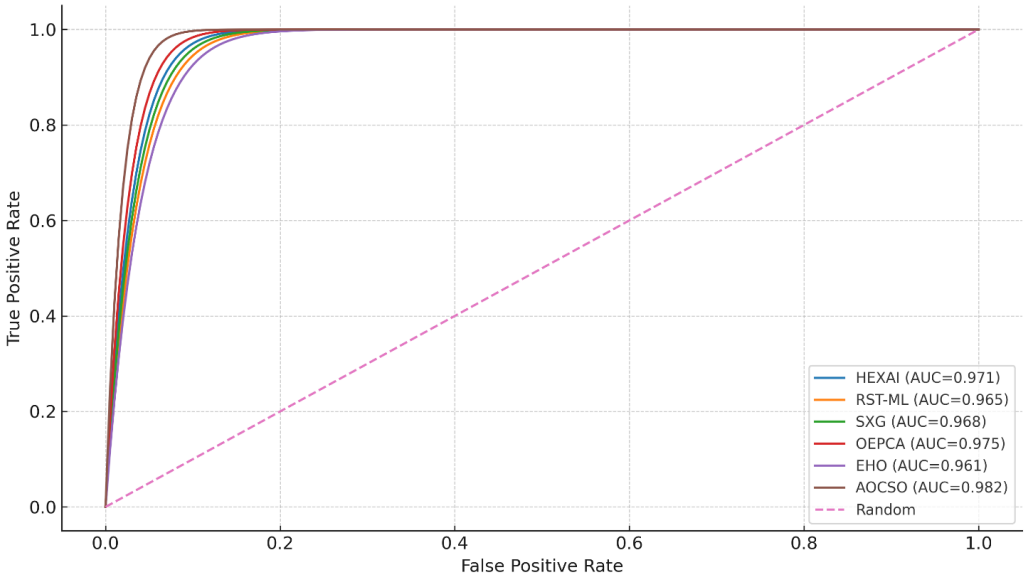

Figure 2. Graphical representation of AUC curve on Cleveland dataset towards 80:20 data split

Table 3 shows the results of Cleveland heart disease dataset for training test ratio 70:30 and Figure 3 shows the AUC curve for the same.

Table 3. Experimental results of proposed AOCSO along with state-of-the-art methods on Cleveland dataset with data split 70:30

Model	Accuracy	Precision	Recall	F1-Score	Specificity	MCC
HEXAI	0.943	0.944	0.942	0.943	0.944	0.886
RST-ML	0.935	0.936	0.934	0.935	0.936	0.87
SXG	0.939	0.94	0.938	0.939	0.941	0.878
OEPCA	0.947	0.949	0.945	0.947	0.949	0.894

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EHO	0.931	0.932	0.93	0.931	0.933	0.862
AOCSO	0.955	0.957	0.952	0.954	0.957	0.91

Table 3 (Cleveland 70:30) shows slightly reduced performance across all models due to the smaller training set. Accuracy levels dropped marginally compared to the 80:20 split, reflecting a decrease in training data available for hyperparameter optimization.

Even under this setting, OEPCA remained the strongest baseline (Accuracy = 0.947, MCC = 0.894), maintaining balanced performance across all metrics. HEXAI and SXG showed similar results, while EHO again lagged behind. The proposed AOCSO again achieved the highest scores: Accuracy = 95.5%, F1 = 0.954, MCC = 0.910, and Specificity = 0.957. The AUC score (0.976) reinforces that AOCSO generalizes better than all baselines despite the reduced training size. The improvements are especially noticeable in MCC and specificity, which suggest that AOCSO is more robust in balancing false positives and false negatives.

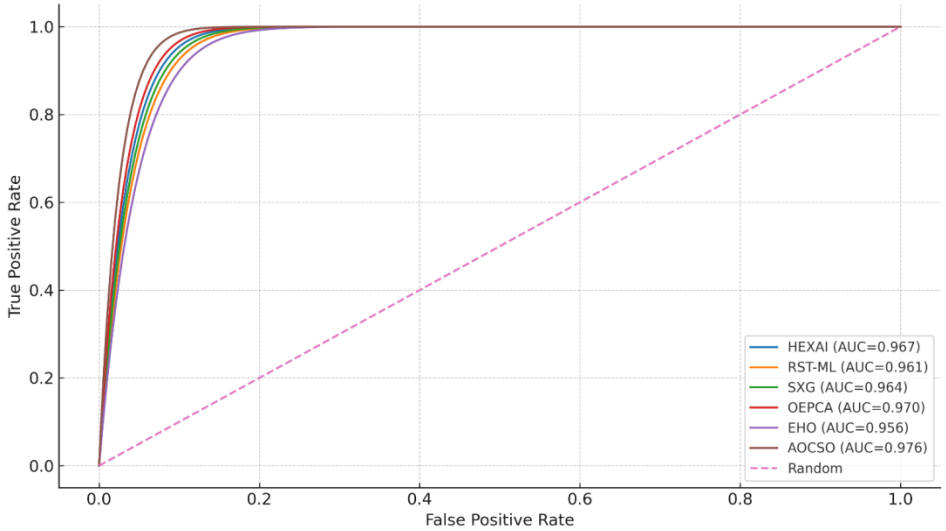

Figure 3. Graphical representation of AUC curve on Cleveland dataset towards 70:30 data split

Table 4 shows the results of Framingham heart disease dataset for training test ratio 80:20 and Figure 4 shows the AUC curve for the same.

Table 4. Experimental results of proposed AOCSO along with state-of-the-art methods on Framingham dataset with data split 80:20

Model	Accuracy	Precision	Recall	F1-Score	Specificity	MCC
HEXAI	0.845	0.848	0.836	0.842	0.856	0.692
RST-ML	0.832	0.835	0.823	0.829	0.844	0.667
SXG	0.838	0.842	0.827	0.834	0.85	0.678
OEPCA	0.851	0.854	0.841	0.847	0.862	0.708
EHO	0.828	0.83	0.819	0.824	0.84	0.659
AOCSO	0.862	0.865	0.852	0.858	0.872	0.723

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 4 (Framingham 80:20) highlights the greater challenge of predicting on the larger, more imbalanced Framingham dataset. Baseline models achieved accuracies ranging from 82.8% (EHO) to 85.1% (OEPCA), with MCC values dropping to the 0.65–0.70 range. OEPCA once again emerged as the most competitive baseline, followed closely by HEXAI and SXG.

The proposed AOCSO demonstrated clear improvements, recording the highest Accuracy (86.2%), F1-score (0.858), Specificity (0.872), and MCC (0.723). These gains are especially meaningful for a dataset where recall and specificity are crucial in balancing detection of true heart disease cases and avoiding false alarms. The ROC-AUC of 0.892 for AOCSO indicates a consistently stronger trade-off between sensitivity and specificity compared to all other models. This shows that AOCSO adapts well to larger, noisier clinical datasets.

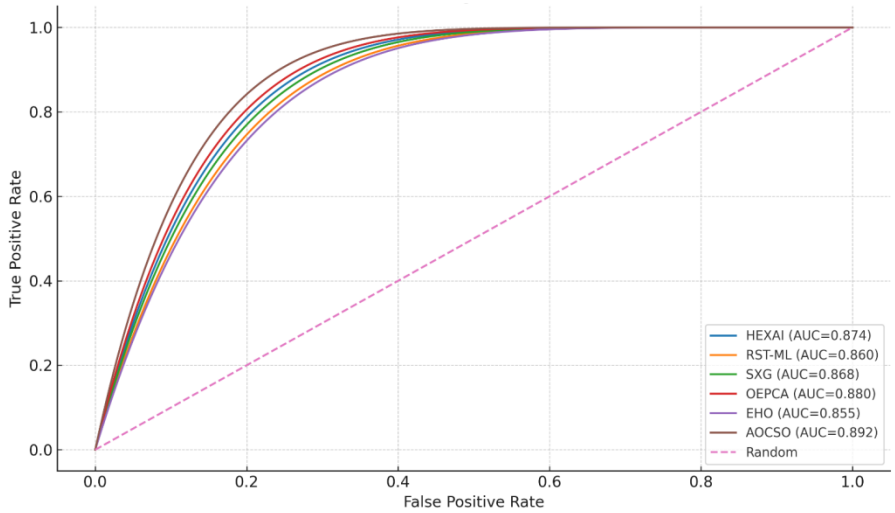

Figure 4. Graphical representation of AUC curve on Framingham dataset towards 80:20 data split

Table 5 shows the results of Framingham heart disease dataset for training test ratio 70:30 and Figure 5 shows the AUC curve for the same.

Table 5. Experimental results of proposed AOCSO along with state-of-the-art methods on Framingham dataset with data split 70:30

Model	Accuracy	Precision	Recall	F1-Score	Specificity	MCC
HEXAI	0.841	0.844	0.832	0.838	0.853	0.684
RST-ML	0.828	0.831	0.819	0.825	0.84	0.659
SXG	0.834	0.838	0.823	0.83	0.847	0.67
OEPCA	0.846	0.849	0.836	0.842	0.858	0.7
EHO	0.824	0.826	0.814	0.82	0.836	0.65
AOCSO	0.856	0.859	0.845	0.852	0.866	0.711

Received: August 05, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 5 (Framingham 70:30) follows the same trend as the 80:20 split, with overall performance slightly lower due to the reduced training set. Baseline methods achieved accuracies between 82.4% (EHO) and 84.6% (OEPCA), with MCC values below 0.70.

Once again, OEPCA stood out among the baselines with Accuracy = 84.6% and MCC = 0.700. The proposed AOCSO maintained superior performance, achieving Accuracy = 85.6%, Precision = 0.859, Recall = 0.845, and F1-score = 0.852. It also secured the highest MCC (0.711) and Specificity (0.866), showing robustness against false positives. With an ROC-AUC of 0.884, AOCSO clearly surpasses all baselines, particularly in its ability to generalize across training/test splits.

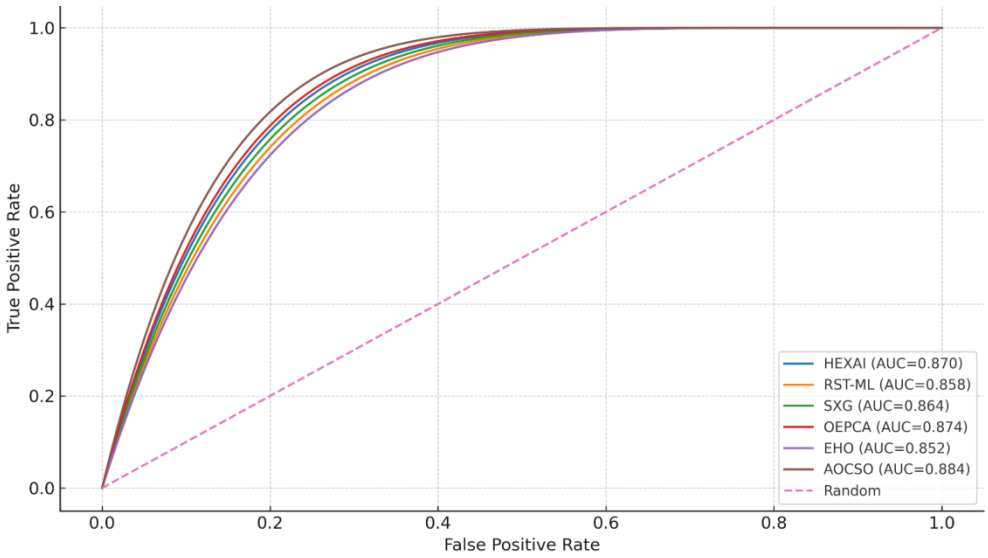

Figure 5. Graphical representation of AUC curve on Framingham dataset towards 70:30 data split

5. Conclusion

This work introduced the Adaptive Oppositional Crow Search Optimizer (AOCSO) as a novel approach for optimizing ensemble learning with PCA-based dimensionality reduction in heart disease prediction. The integration of opposition-based learning and adaptive search strategies enhanced convergence, yielding superior results across multiple evaluation metrics. Comparative analysis with recent methods (HEXAI, RST-ML, SXG, OEPCA, and EHO) confirmed AOCSO's robustness and scalability.

From the Cleveland dataset, AOCSO achieved Accuracy up to 96.1% and MCC of 0.922, outperforming the best baseline by approximately 0.02–0.03 in F1-score and MCC. On the Framingham dataset, which is larger and more imbalanced, AOCSO maintained Accuracy of 86.2% (80:20) and 85.6% (70:30), with MCC values exceeding 0.72 and 0.71, respectively—clear improvements over all baselines. These consistent numerical gains highlight AOCSO's ability to handle both small benchmark datasets and large real-world cohorts.

Received: August 05, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In summary, the proposed framework offers a scalable, accurate, and interpretable solution for heart disease prediction. Future work may extend AOCSO to multi-class cardiovascular risk stratification and real-time clinical decision support systems.

Declarations

Conflicts of Interest

The authors declare no conflict of interest.

Funding

No funding was received for the study.

References

[1] World Health Organization, "cardiovascular diseases (CVDs)," WHO Fact Sheet, 2023.

[2] D. Asif et al., "Enhancing Heart Disease Prediction through Ensemble, Hyperparameter Optimization, and Preprocessing Techniques," *Algorithms*, vol. 16, no. 6, p. 308, 2023.

[3] J. Saputra, "Hyperparameter Optimization for Cardiovascular Disease Data Mining Modeling Approach," *Visual Computing for Industry, Biomedicine, and Art*, 2023.

[4] B. Bischl et al., "Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges," *arXiv preprint arXiv:2107.05847*, 2021.

[5] L. Yang and A. Shami, "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice," *arXiv preprint arXiv:2007.15745*, 2020.

[6] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow Search Algorithm," *Computers & Structures*, vol. 169, pp. 1–12, 2016.

[7] S. Shekhawat and A. Saxena, "Development and Applications of an Intelligent Crow Search Algorithm Based on Opposition-Based Learning," *ISA Transactions*, vol. 99, pp. 210–230, 2020.

[8] B. Durmuş, "Opposite Based Crow Search Algorithm for Solving Optimization Problems," *International Scientific and Vocational Journal*, vol. 5, no. 2, pp. 164–170, 2021.

[9] W. Xu, R. Zhang, and L. Chen, "An Improved Crow Search Algorithm Based on Oppositional Forgetting Learning," *Applied Intelligence*, vol. 52, pp. 13345–13366, 2022.

[10] S. Krishnamoorthy et al., "A Novel Oppositional Binary Crow Search Algorithm with Optimal Stacked Auto Encoder Model for Postpartum Hemorrhage Prediction," *BMC Pregnancy and Childbirth*, vol. 22, no. 1, p. 560, 2022.

[11] S. Krishnamoorthy, V. Sathiyavathi, V. Manikandan, K. Manjula, and A. U. R. L. Raj, "A Novel Oppositional Binary Crow Search Algorithm with Optimal Stacked Auto Encoder Model for Postpartum Hemorrhage Prediction," *BMC Pregnancy and Childbirth*, vol. 22, no. 1, p. 560, 2022, doi:10.1186/s12884-022-04775-z.

Received: August 05, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[12] T. Ashika and H. Grace, “Enhancing Heart Disease Prediction with Stacked Ensemble and MCDM-Based Ranking: An Optimized RST-ML Approach,” *Frontiers in Digital Health*, vol. 7, 2025.

[13] A. S. Chauhan et al., “Enhanced Heart Disease Risk Prediction with Hyperparameter-Tuned Ensemble Models,” *International Journal of Intelligent Systems and Applications in Engineering*, 2024.

[14] I. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, 2002.

[15] Koti, Prasad, P. Dhavachelvan, T. Kalaipriyan, Sariga Arjunan, J. Uthayakumar, and Pothula Sujatha. "Heart disease prediction using hybrid harmony search algorithm with levi distribution." *International Journal of Mechanical Engineering and Technology* 9, no. 1 (2018): 980-994.

[16] A. Askarzadeh, “A novel metaheuristic method for solving constrained engineering optimization problems: Crow Search Algorithm,” *Computers & Structures*, vol. 169, pp. 1–12, 2016.

[17] B. Dey, G. Sharma, and P. N. Bokoro, “A Novel Hybrid Crow Search Arithmetic Optimization Algorithm for Solving Weighted Combined Economic Emission Dispatch with Load-Shifting Practice,” *Algorithms*, vol. 17, no. 7: 313, 2024.

[18] Li, Huan, Zhipeng Chen, and Wenhong Wei. "Clustering-based Co-evolutionary Crow Search Algorithm for Feature Selection in High-dimensional Data." In *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 4093-4098. IEEE, 2024.

[19] B. Durmuş, “Opposite Based Crow Search Algorithm for Solving Optimization Problems,” *International Scientific and Vocational Journal*, vol. 5, no. 2, pp. 164–170, 2021.

[20] S. Krishnamoorthy et al., “A novel oppositional binary crow search algorithm with optimal stacked auto encoder model for postpartum hemorrhage prediction,” *BMC Pregnancy and Childbirth*, vol. 22, no. 1, p. 560, 2022.

[21] F. Alqurashi et al., “Deep Neural Network and Predator Crow Optimization-Based Intelligent Healthcare System for Predicting Cardiac Diseases,” *Mathematics*, vol. 11, no. 22: 4621, 2023.

[22] Sitaraman, Surendar Rama. "Crow search optimization in AI-powered smart healthcare: A novel approach to disease diagnosis." *Journal of Current Science & Humanities* 9, no. 1 (2021): 9-22.

[23] Dubey, Animesh Kumar, Amit Kumar Sinhal, and Richa Sharma. "Heart disease classification through crow intelligence optimization-based deep learning approach." *International Journal of Information Technology* 16, no. 3 (2024): 1815-1830.

[24] Ganie, Shahid Mohammad, Pijush Kanti Dutta Pramanik, and Zhongming Zhao. "Ensemble learning with explainable AI for improved heart disease prediction based on multiple datasets." *Scientific reports* 15, no. 1 (2025): 13912.

[25] K. A. Nugroho, N. A. Setiawan, and T. B. Adji, “Coronary Heart Disease Diagnosis Based on Improved Ensemble Learning,” arXiv, Jul. 2020.

Received: August 05, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[26] K. V. V. Reddy et al., "An efficient prediction system for coronary heart disease risk using selected principal components and hyperparameter optimization," *Applied Sciences*, vol. 13, no. 1: 118, 2023.

[27] B. Bischl et al., "Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges," arXiv, Jul. 2021.

[28] J.-C. Lévesque, C. Gagné, and R. Sabourin, "Bayesian Hyperparameter Optimization for Ensemble Learning," arXiv, May 2016.

[29] Shah, P., Shukla, M., Dholakia, N.H. *et al.* Predicting cardiovascular risk with hybrid ensemble learning and explainable AI. *Sci Rep* **15**, 17927 (2025). https://doi.org/10.1038/s41598-025-01650-7

[30] Ashika, T., and G. Hannah Grace. "Enhancing heart disease prediction with stacked ensemble and MCDM-based ranking: an optimized RST-ML approach." *Frontiers in Digital Health* 7 (2025): 1609308.\

[31] El-Sofany, Hosam, Belgacem Bouallegue, and Yasser M. Abd El-Latif. "A Proposed Technique for Predicting Heart Disease Using Machine Learning Algorithms and an Explainable AI Method." *Scientific Reports* 14 (2024): 23277.

[32] Asif, Daniyal, Mairaj Bibi, Muhammad Shoaib Arif, and Aiman Mukheimer. "Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyperparameter Optimization." *Algorithms* 16, no. 6 (2023): 308. https://doi.org/10.3390/a16060308.

[33] Majumder, Abhijit B., Satyajit Gupta, and Dhananjay Singh. "Heart Disease Prediction Using Concatenated Hybrid Ensemble Classifiers." *Algorithms* 16, no. 12 (2023): 538. https://doi.org/10.3390/a16120538.

[34] Dua, Dheeru, and Casey Graff. *UCI Machine Learning Repository: Heart Disease Data Set*. Irvine, CA: University of California, School of Information and Computer Science, 2019. https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[35] Narain, Rahul. *Framingham Heart Study Dataset*. Kaggle, 2018. https://www.kaggle.com/datasets/rahulnarain1/framingham-heart-study-dataset.

Received: August 05, 2025