International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 909-923
https://doi.org/10.70917/ijcisim-2026-3319
Article

# An Optimization Framework Using Flood Algorithm-Based Feature Selection for Heart Disease Prediction

**Jazya Moftah Ahmad Amshaher [1]* , Amna Elhawil [2]**

[1]Department Of Electrical And Computer Engineering School Of Applied Science And Engineering Libyan Academy For Graduate Studies Tripoli Libya .
[2]Department Of Computer Engineering Faculty Of Engineering University Of Tripoli Tripoli Libya .
E-mail: jazamoftah@su.edu.ly

**Abstract:** Heart disease is one of the common disease and has been ranked as the one most prevalent. Early detection is the important step to reduce mortality rates and treatment costs. This research proposes a technique for extracting the most relevant features using flood algorithm (FLA) to improve the performance of the diagnosis model. The feature selection model is evaluated using parameters derived from the confusion matrix and using six models:MLP, TabNet, Random Forest, SVM, Logistic Regression, and XGBoost, the models were testing to the heart disease dataset. The best results for this models were obtained using XGBoost with 7 selected features from 13features and the resulting performance for accuracy was 98.9%, specificity 98.8%, sensitivity 99.00%, area under the curve (AUC) 99.8%, and F1-Score 99%.
**Keywords:** Heart Disease, Feature Selection, Flood Algorithm.

## 1. Introduction

The disease of the heart ranks among the top diseases that cause mortality throughout the world. As per the statistics released by the World Health Organization, heart disease accounts for 30 percent of moralities in the world. This alarming statistic underscores the critical importance of early detection.Recent technology have significantly contributed to progress in the medical field, through the application of machine learning (ML) . Machine learning models analyze input data using statistical techniques to generate predictive outcomes[1].However, one of challenges in training such models is incorporating an excessive number of features is not always beneficial. Medical datasets often include both relevant and irrelevant features, where irrelevant ones may introduce noise, increase model complexity, and raise computational costs[2] without improving predictive performance.In this proposed research, we aim to employ the Flood algorithm optimizer (FLA), proposed by Ghasemi [3]. It is a meta-heuristic optimization algorithm. This algorithm will be applied, for the first time, for feature selection. The primary objective of this approach is to develop a predictive model characterize by accuracy. The integrate the Gini index to Flood algorithm to improve the search of the most relevant features. The impurity measures is a measure used to assess how effectively a feature distinguishes between different classes in a dataset. for the classification the : MLP, TabNet, Random Forest, SVM, Logistic Regression, and XGBoost algorithms were used. The main contributions of this work are summarized as follows:

1. multi-objective feature selection framework combining FLA and Gini index to guide feature selection toward high accuracy and class balance.
2. Integration of greedy search to automatically identify optimal hyper parameters.
3. A robust evaluation protocol using nested stratified cross-validation to avoid data leakage.
4. Analysis of feature importance in clinical context.

5. The method was evaluated using the MLP, TabNet and Comprehensive comparison with strong baseline models.
6. Providing a comprehensive evaluation using nested stratified cross-validation, confidence intervals, and statistical significance testing.

## 2. Literature review

Many medical datasets contain both relevant and irrelevant features. These irrelevant features negatively impact the target population, as excess features can lead to confusion and complicate target identification. Numerous studies have been published employing various techniques to improve the prediction of heart disease, aiming to provide more accurate diagnoses.

Zhu et al.[4]proposed feature selection approach to the Breast Cancer Diagnosis, which used Shapely additive explanation (SHAP) values for Recursive Feature Elimination (RFE), utilizing a Random Forest (RF) algorithm and solving the data imbalance through incorporated Borderline - smote. the proposed method was assessed using five machine learning models, K-Nearest Neighbor (KNN), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Light Gradient Boosting Machine (Light GBM), applied to the (WBCD) datasets. using the Particle Swarm Optimization (PSO) algorithm used for optimizing hyper parameters of five models . the Light GBM-PSO model was height performance and achieved accuracy of 90.0% in differentiating between benign and malignant cases.However Generalization across diverse datasets, as well as model complexity, leads to computational costs.

Rahman et al. enhanced feature interpret ability using cross-validation and SHAP analysis for five feature selection algorithms—Recursive Feature Elimination, Grey Wolf Optimizer, Particle Swarm Optimizer, Genetic Algorithm, and Boruta—using cross-validation and SHAP analysis to enhance feature interpret ability. using two classification algorithms: the light gradient boosting machine algorithm and the extreme gradient boosting algorithm (XGBoost). by the Boruta algorithm, outperformed other configurations with the Light GBM classifier, achieving an accuracy of 85.16%.But the limitation was the PIDD is small and demographically narrow[5].

All this studies[6]-[8] applied Relief as the feature selection method and applied on different dataset , but evaluated the selected features using different models such a Support Vector Machine (SVM), Neural Network.

Radwan et al. analyzed different machine learning (ML) models for predicting early blight and the late blight diseases of more than 4000 records of weather conditions. where improve the predictive performance of the models by identifying feature sets pertinent Feature selection methods such as the binary Grey-lag Goose Optimization (bGGO) ,that determine correlations with disease prevalence with several machine learning models were used : logistic regression, gradient boosting, multi layer perceptron (MLP), and support vector machine (SVM), as well as K-nearest neighbor (KNN) . Results demonstrated that the MLP model, with feature selection, achieved an accuracy of 98.3%,but limitation was-the dataset used temporal biases , may also be data imbalance issues [9].

Mojahid et al[10].discussion variable selection methods—specifically Recursive Feature Elimination (RFE), Principal Component Analysis (PCA) and Least Absolute Shrinkage and Selection Operator (LASSO), these methods evaluate with Support Vector Machines (SVM), Logistic Regression (LR), and extremer Gradient Boosting (XGBoost) for COVID-19,the best performance was LASSO with SVM achieve best performance of accuracy 0.7679 .They used a single dataset, which may limit the generalization of the results, and the analysis was limited to seven machine learning algorithms; not addressing deep learning techniques would enhance prediction accuracy.

Several studies[11]-[16] enhanced from classification performance combined filtering-based feature selection methods with bagging ensemble models . In these works, filter techniques such as Relief, Chi-Square, Information Gain, Based Feature Selection were first applied to rank the most relevant features . After that bagging-based models such as Random Forest ensembles were trained to improve minimize variance. The hybrid methods improvements in accuracy, computational efficiency.

Ahmad et al. [17] presents comparison of feature selection methods and the impact of feature selection techniques such as Mutual Information (MI), Analysis of Variance (ANOVA), and Chi-Square on the performance of machine learning (ML) and deep learning (DL) models for heart disease prediction. The results showed that Mutual Information (MI)achieved the best accuracy, comparison other methods, especially with neural networks, an accuracy of 82.3%. However, the feature selection techniques and machine learning/ deep learning models were applied to a single dataset, this limits the generalization and validity of the results.

Kazerani et al. [18] proposed feature selection algorithm based on PSO algorithm of breast cancer diagnosis, The PSO was similarity algorithm with genetic algorithm,the proposed algorithm achieved the height accuracy of breast cancer diagnosis.

This researcher[19] - [20] based filter feature selection methods to enhance heart disease prediction models such Relief and Chi-Square techniques were applied to rank the most relevant features to model training. These approaches improved classification accuracy while reducing computational complexity. Moorthy et al. Employed A feature selection approach to improve heart disease prediction by integrated Random Forest- Feature Sensitivity and Feature Correlation technique [21]. The methodology was implemented using the Cleveland Heart Disease dataset and the accuracy was 81.16%. however, Feature sensitivity requires extensive analysis which may affect the efficiency of the model.

Elisseeff et al. discussed the difference between filtering and packaging methods in feature selection and found that when using a feature set obtained through back age methods, the model tends to select features identified using filtering methods, making the model more susceptible to the results of those statistical methods[22].

Several research [23]-[26]proposes a new approach for breast cancer diagnosis that combines a Mutual Information (MI)-based feature selection strategy with multiple machine learning classifiers such as K-Nearest Neighbor , Support Vector Machine , Random Forest, and Neural Networks.This proposed study improved accuracy of breast disease prediction by identifying essential features the best classification modelling techniques.

Gutiérrez et al.focused this study compare methodologies of feature selection to identify risk factors for early treatment,they used four feature selection algorithms: Univariate, Boruta, Galgo, and Elastic net, for classifying non-DKD and DKD patients used Random Forest[27]. The Boruta algorithm is a relevant features selection wrapper method based the RF classifier used of identifying crucial features for disease diagnosis, while GALGO employs genetic algorithms (GA) for selecting models based high fitness. Galgo with Random Forest achieved the best classification performance with an area under the curve of 0.80%. List of algorithms looked at for our study is display as in Table 1.

**Table 1.Related Studies**

| **Reference** | Dataset | Method | Models | Performance | Limitations |
|---|---|---|---|---|---|
| Zhu et al. [4] | Breast Cancer (WBCD) | SHAP + RFE + RF +Borderline-SMOTE | KNN, RF, LR, SVM, LightGBM + PSO tuning | Ligh tGBM-PSO (Accuracy = 90%) | High computational cost, model complexity |
| Rahman et al. [5] | PIDD | RFE,GWO, PSO, GA, Boruta + SHAP analysis | Light GBM, XGBoost | Boruta + LightGBM (Accuracy = 85.16% | Small dataset, limited demographic diversity |
| Studies [6–8] | Various datasets | ReliefF | SVM, Neural Networks | Varies across studies | Inconsistent evaluation across models |
| Radwan et al. [9] | Weather dataset (>4000 records) | bGGO feature selection | Logistic Regression, Gradient Boosting, MLP, SVM, KNN | MLP + FS (Accuracy = 98.3%) | data imbalance issues |
| Mojahid et al. [10] | COVID-19 dataset | RFE, PCA, LASSO | SVM, Logistic | LASSO + SVM | Single dataset, limited generalization, |

| | | | | | |
|---|---|---|---|---|---|
| | | | Regression, XGBoost | (Accuracy = 76.79%) | no deep learning methods |
| Studies [11–16] | Various datasets | ReliefF, Chi-square, Information Gain, Mutual Information | Random Forest, Decision Tree ensembles | Improved accuracy and efficiency | Limited robustness across different datasets |
| Ahmad et al. [17] | Heart Disease dataset | Mutual Information (MI), ANOVA, Chi-square | Machine Learning & Deep Learning (Neural Networks) | MI + Neural Network (Accuracy = 82.3%) | Single dataset, limited external validity |
| Kazerani et al. [18] | Breast Cancer | Particle Swarm Optimization | ML clasificar | High accuracy reported | computational cost |
| Studies [19–20] | Heart Disease datasets | ReliefF, Chi-square | SVM, Neural Networks | Improved classification accuracy | Limited methodological diversity |
| Moorthy et al. [21] | Cleveland Heart Disease dataset | RF-FSFC (Feature Sensitivity + Correlation) | Random Forest | Accuracy = 81.16% | High computational cost |
| Studies [23–26] | Breast Cancer datasets | Mutual Information (MI) | KNN, SVM, RF, Neural Networks | Improved predictive performance | Limited model comparison |
| Gutiérrez et al. [27] | DKD dataset | Boruta, GA (GALGO) | Random Forest | Accuracy = 0.80 | Limited evaluation scope, moderate performance |

## Background

### *TabNet Model*

TabNet is a deep learning architecture designed for tabular data, offering advantages over traditional machine learning models. TabNet introduces a sequential attention mechanism that enables the model to select the most relevant features at each decision step. It can also handle raw tabular data without extensive preprocessing. Additionally, TabNet is computationally efficient due to its use of shared feature transformers and decision steps, making it suitable for large datasets [28]. Moreover, TabNet has been successfully applied in medical applications, particularly in cardiovascular disease prediction, demonstrating high performance and interpret ability[29] .

### *Multilayer Perceptron (MLP)*

A multi-layer neural network (MLP) is a type of artificial neural network with a feed forward mechanism. An MLP consists of an input layer, one or more hidden layers, and an output layer. One of advantages of MLP is its ability to model nonlinear relationships, making it well-suited for disease prediction. Furthermore, it is flexible in handling high-dimensional data and can be combined with feature selection techniques to enhance performance[30].

## Research Methodology

This section briefly describes the experimental results obtained in the four phases:

1. Classification evaluation without feature selection .
2. Feature Selection using Flood Algorithm (FLA) .
3. Classification using the Gini Index for feature evaluation.
4. Classification using a greedy search approach.

More information about these stages will be provided in this section.

### *A. Classification Evaluation Without Feature Selection Phase*

In this phase, we constructed six prediction models using the training dataset for the MLP, TabNet, Logistic Regression, XG Boost, Random Forest and support vector machine classifiers. The models were constructed by using all the features without any process of feature selection.

### *B. Feature Selection using Flood Algorithm (FLA)*

FLA is introduced to be applied on optimization problems and engineering design problems. In this work FLA is used as feature selection technique, The parameters of (FLA) were configured to ensure the stability of the feature selection. The selection of FLA was based its strong capability in optimization in exploring the feature search space to identify the most subset of features. Unlike traditional sequential search approaches that evaluate a limited number of feature combinations, FLA explores multiple candidate solutions simultaneously, which improves the balance between exploration and exploitation .the Flood Algorithm was implemented with a population size of 200 and executed for 50 maximum iterations. where candidate solution represented a binary feature mask. To maintain an subset size, the number of selected features was between 3 features and 8 features. During the optimization process, candidate solutions were updated , which enhanced the diversity of the search process and reduced of premature convergence.

In addition, a probabilistic exploration mechanism was incorporated to generate new candidate solutions around the current best solution, thereby improving the algorithm's ability to escape local optima. The algorithm utilized dynamic control parameters that changed adaptive across iterations to achieve a balance between global exploration in early iterations and local exploitation in later iterations.Finally, to guarantee reprehensibility of the experimental results, the random seed was fixed at 42 throughout the entire optimization and training process.

### *C. Flood Algorithm (FLA) with Gini Index*

In this the experiments, the fitness evaluation process was using the Gini index within the flood algorithm, for improving the quality of feature selection a of the predictive models in predicting heart disease as in figure 1.The Gini index is a measure of statistical dispersion intended to represent [31].The Gini coefficient ranges from 0 to 1, where 0 represents perfect equality while 1 represents perfect inequality.The mathematical formulation of the Gini impurity is defined in[32] .

$$Gini_i = 1 - \sum_{k=1}^{c} p^2 k \qquad (1)$$

*C is the total number of classes.*

*pk represents the probability of class k.*

The goal is to maximize the accuracy of the model while minimizing imbalance between classes, the $(1 - Gini)$ formula was used, whereby lower purity values are converted to higher values in the objective function, thereby contributing to balancing performance between the two classes .The fitness in the Flood algorithm has been modified to be a combination of model accuracy and the Gini measure as follows:

$$Fitness_i = \alpha \times Accuracy_i + (1-\alpha) \times (1 - Gini_i) \qquad (2)$$

*where*

*$Accuracy_i$:Represent accuracy of model .*

*$Gini_i$: value the model's results in the generation.*

*α: The balance factor between accuracy and balance .*

The fitness function was formulated as a multi-objective optimization problem, taking two main goals simultaneously: (1) maximizing classification accuracy, and (2) improving class balance using a $Gini_i$. In this case, the parameter (α) was used to control the balance of the two objectives. Sensitivity analysis was done by changing the value of (α), where α = {0.2, 0.4, 0.6, 0.8}. The results of the experiment indicated the balance of the two objectives. Low values of (α) give greater emphasis to class balance, resulting in higher sensitivity but lower specificity and accuracy. Conversely, high values of (α) improve accuracy and specificity with a decrease in sensitivity. Intermediate values, especially (α = 0.6), achieved a better balance between sensitivity and specificity. The above findings suggest that (α) affects the learning bias of the model. Nonetheless, for this experiment, (α = 0.8) was used in order to ensure high accuracy with consideration of imbalanced data classes.

Furthermore, the proposed framework relies on feature selection rather than extraction. The optimization algorithm selects the most directly relevant subsets of the original features. A Gini index is integrated into the fitness function to guide the search toward feature subsets that exhibit high predictive performance and an improved class distribution.

**Steps Flood Algorithm as following :**

1. Initialization:Specify parameters such as the number of iterations, population size, and as spread rate.
2. Flood Spread Simulation:In this step each flood cell, spread the solution to discover new regions and this the solution is evaluated using an objective function.
3. Solution Update:This step include new regions through adding, removing, or replacing features .New solutions are accepted according to the evaluation metric.
4. Convergence Check:include reaching the specified number of iterations or change between iterations become a small
5. Result Extraction:The best solution obtained is selected.
6. Post-Processing:Evaluate the final solution on test data.

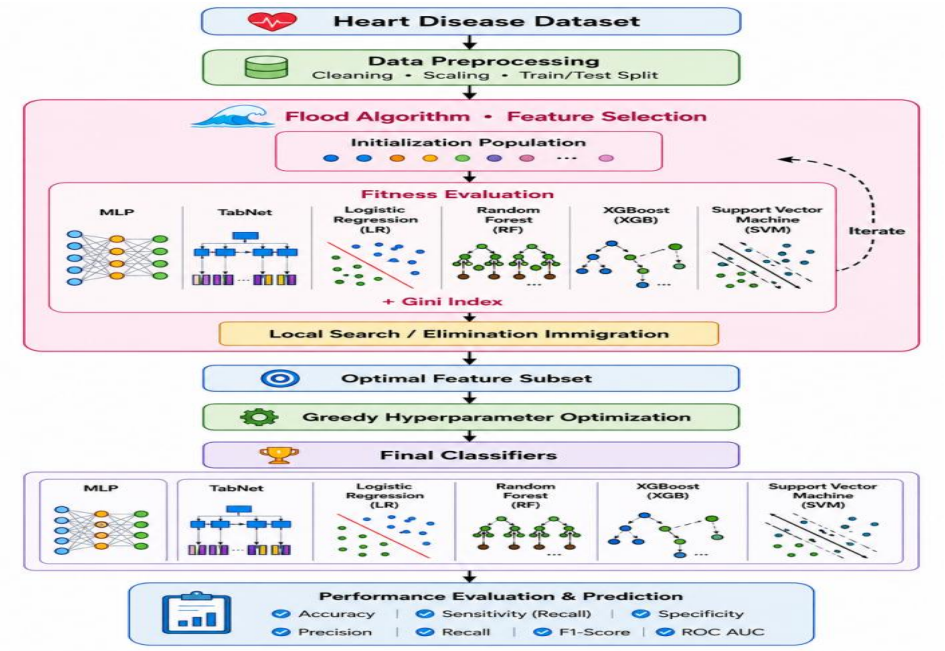


Fig.1:Proposed Flood algorithm

## D. *Optimizing models using Greedy Search.*

Optimizing hyper parameters is a crucial step in enhancing the prediction accuracy and generalization of models. In this study, a greedy search strategy was employed to optimize the hyper parameters of classification models. In this approach, the selection of hyper parameters takes place in such a way that it enhances the performance of the

model by testing various configurations of the parameters and using the configuration with the best classification accuracy.

The greedy search algorithm is known for its computation efficiency and simplicity, especially when handling high-dimensional parameter space. This is due to the fact that the greedy approach does not check all possible combinations of parameters but concentrates on optimization at each iteration. This approach helps to improve the predictive abilities of the model through iterative optimization while keeping the computations feasible.

## 3. Results And Analysis

This section presents the experimental results obtained in the four phases. All evaluation metrics were obtained using a nested stratified 5-fold cross-validation strategy to ensure unbiased performance estimation. whereas the inner 3 folds were dedicated to feature selection using the (FLA).To evaluate the reliability of the proposed framework, 95% confidence intervals were calculated based on the results of the outer folds. In addition, paired statistical significance analysis was conducted using a paired t-test to examine whether the observed performance differences between the compared algorithms were statistically significant across the outer cross-validation folds.

### A. *Dataset Collection*

This study used the Heart Disease dataset is publicly available on Kaggle .The dataset contains 1025 patient records with 13 features and features collected over multiple time periods during .Fig. 2 illustrates the distribution of patients according to their diagnosis status, while the detailed description of the dataset features is presented in Table 2.

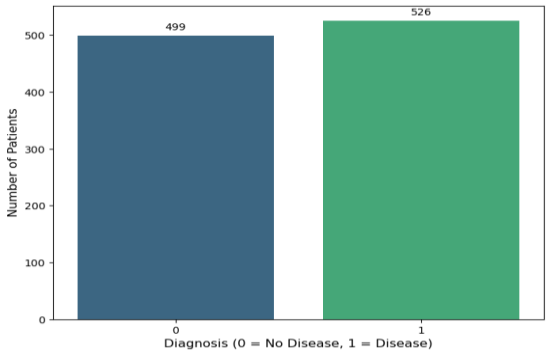


Fig 2. Distribution of patients based on diagnosis status

**Table 2.List of dataset features**

| NO | Feature | Description |
|---|---|---|
| 1 | age | Age in years |
| 2 | sex | (1 = male; 0 = female) |
| 3 | cp | Chest pain type |
| 4 | trestbps | Resting blood pressure |
| 5 | chol | Serum cholestoral |
| 6 | fbs | Fasting blood sugar |
| 7 | restecg | Resting electrocardio graphic |
| 8 | thalach | Maximum heart rate achieved |
| 9 | exang | exercise induced angina |
| 10 | oldpeak | ST depression |
| 11 | slope | Slope of the peak exercise |
| 12 | ca | Number of major vessels |
| 13 | thal | Results of nuclear stress test |
| 14 | target | Indicate to the presence of heart disease.the value 0 = no disease and 1= disease |

### B. *Data preprocessing*

The dataset was checked for any missing values, feature distribution, and class imbalance. The missing value problem was solved by using Simple Median method. Then, the process of feature scaling was performed using

Standard Divider method, in which features were scaled to have zero mean and one variance. The class imbalance problem was dealt with using Synthetic Minority Oversampling Technique (SMOTE). Nested cross-validation was utilized to avoid oversampling. The outer loop served to evaluate the model at the end, while the inner loop was responsible for selecting the features and optimizing the hyperparameters. To systematically evaluate the impact of each component of the proposed hybrid model, the experimental path was divided into four consecutive steps. The first step involved a classification phase without any feature selection. In the second step, feature selection was applied using a Flood algorithm. The third step involved applying a Gini index to the Flood algorithm's fitness function to achieve class balance. In the final step, the Flood algorithm, the Gini index, and a greedy hyper parameter optimization algorithm were combined.

### *C. Limitations of PCA*

As illustrated in Fig. 3, the first two principal components explain only 33% of the total variance, which reflects the high dimensional and noticeable overlap between classes Consequently, techniques such as PCA may have limited capability in achieving clear class separability detection, as shown in Fig. 3

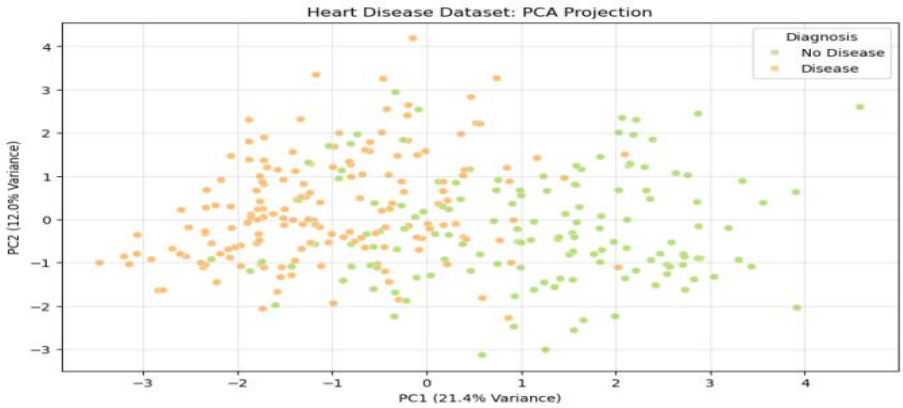


Fig 3. PCA: Visualizing cluster reparability

Furthermore, deep customization becomes a challenge when using complex learning, particularly large-scale model training, as well as advanced architectures like TabNet and other learning models. This issue was not addressed in the MLP model. L2(α) regulation was applied to control for model risk, while greedy searching was used to optimize the regulation coefficient and parameters apart. A TabNet model was created early on using patience equal to 5 during the training of children.Moreover, the dataset still lacks a separator of categories. In this competition, the sampling method (`stratify=y`) was applied when composing the dataset into two training sets and covering the original subscriptions for the categories.Finally, due to the different statistical variance criteria for the data set, was applied `Standard Scaler` before training all classification models .

## 4. Results of Experiments

**Experiment (a): Classification without Feature Selection**

Six models were used for the training data set include MLP,TabNet, Logistic Regression, XG Boost, Random Forest and support vector machine without of feature selection,were assessed using all features of the dataset.Table 3 lists the experiment's outcomes for every class.

**Table 3.Performance of all models without feature selection**

| models | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **MLP** | 0.343 ± 0.470 (0.000–0.926) | 0.336± 0.460 (0.000-0.908) | 0.726±0.172 (0.513–0.939) |
| **TabNet** | 0.774 ± 0.155 (0.581–0.967) | 0.710 ± 0.166 (0.504–0.916) | 0.735±0.177 (0.515–0.955) |
| **LR** | 0.355 ± 0.486 (0.000–0.959) | 0.322 ± 0.441 (0.000–0.870) | 0.736±0.164 (0.532–0.940) |

| models | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **XGB** | 0.828 ± 0.011 (0.815–0.841) | 0.859 ± 0.014 (0.842–0.876) | 0.927±0.009 (0.916–0.938) |
| **RF** | **0.859 ± 0.032 (0.819–0.899)** | **0.865 ± 0.028 (0.830–0.900)** | **0.931±0.017 (0.910–0.952)** |
| **SVM** | 0.545 ± 0.348 (0.112–0.978) | 0.621 ± 0.361 (0.172–1.000) | 0.679±0.245 (0.375–0.983) |

The results indicated that using the RF classifier had the highest accuracy of 86% in predicting of disease. The Fig. 4 displayed show the learning curve of this model .

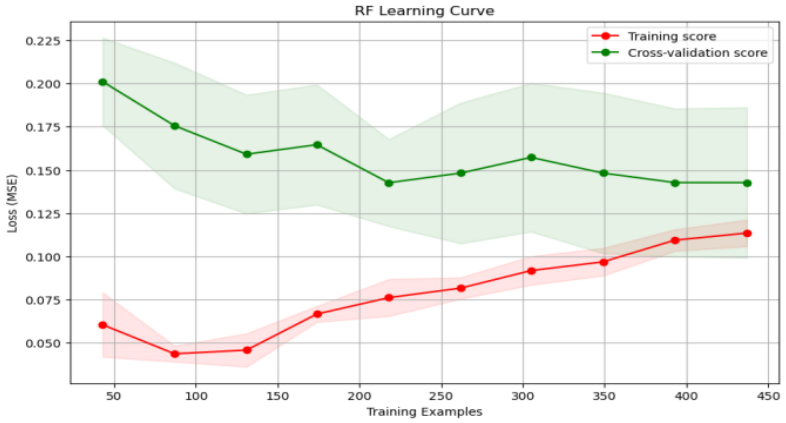


Fig .4: Learning curve of the Random Forest (RF) model

**Experiment (b): Classification with feature selection**

This experiment involves Flood algorithm based on requirement stated in Tabl 4.

**Table 4.FLA features selection requirement**

| Process | Requment |
|---|---|
| Attribute Evaluator | CFS Subset Evaluator (supervised, Class (nominal): 1 Outcome) |
| Search method | FLA search |
| Requirement to Obtains the best fit features | Start set:no atturbute<br>Population size:200<br>Number of iteration :50 |
| Random Seed | 42 |

The results indicated that the model that used SVM as classifier exhibited a high accuracy in predicting of heart disease was 90% and Performance of all models shown in Table 5.The results showed that out of 13 features, only eight features are the best fit feature-age, cp, restecg, exang, slope, ca, thal, sex: - as in Table 6.

**Table 5.Performance of all models with (FLA)**

| Model | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **MLP** | 0.825 ± 0.023 (0.797–0.853) | 0.852 ± 0.012 (0.837–0.867) | 0.917 ± 0.016 (0.897–0.937) |
| **TabNet** | 0.775 ± 0.023 (0.746–0.802) | 0.769 ± 0.024 (0.739–0.799) | 0.843 ± 0.016 (0.823–0.863) |

| Model | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **LR** | 0.831 ± 0.023 (0.802–0.860) | 0.863 ± 0.005 (0.856–0.870) | 0.910 ± 0.011 (0.896–0.924) |
| **XGB** | 0.836 ± 0.009 (0.824–0.848) | 0.872 ± 0.020 (0.847–0.897) | 0.930 ± 0.010 (0.918-0.942) |
| **RF** | 0.857 ± 0.012 (0.842–0.872) | 0.882 ± 0.009 (0.871–0.893) | .926 ± 0.007 (0.917-0.935) |
| **SVM** | **0.891 ± 0.025 (0.860–0.922)** | **0.909 ± 0.020 (0.884–0.934)** | **0.959 ± 0.013 (0.943–0.975)** |

**Table 6. Features selection requirement result of FLA**

| Model | #Feature | Name of selected feature |
|---|---|---|
| MLP | 8 | sex, cp, oldpeak, ca, thal, trestbps, thalach, slope |
| TabNet | 7 | sex, cp, thalach, exang,oldpeak, ca, thal |
| LR | 8 | sex,cp, trestbps, restecg, thalach, oldpeak ,ca, thal |
| XGB | 7 | sex, cp, trestbps, exang, ca, thal, slope |
| RF | 7 | cp, trestbps, exang, slope, ca, thal,fb |
| **SVM** | **8** | **age, cp, restecg, exang, slope, ca, thal, sex** |

**Experiment (c): Feature Selection Using FLA and Gini Index**

In this experiment, the Flood algorithm was combined with a Gini index to improve feature selection. A Gini index was incorporated into the Flood algorithm's fitness function to evaluate classification performance and class balance. The fitness function was designed to maximize accuracy while minimizing class heterogeneity by including the (1 − Gini) limit. Table 7 summarizes the performance of the evaluated classification models. The results indicate that incorporating a Gini index into the Flood algorithm's fitness function improved predictive compared to using the Flood algorithm alone ,the SVM outperformed the other models with an accuracy of 91.1% These results suggest that the Gini index guides the optimization process toward subsets of features that better separate the target classes and reduce the negative effects of class imbalance. Table 8 presents the optimal feature set with the highest fitness value. The difference between Flood-based feature selection and feature selection combined with a Gini index lies in the optimization objective. While Flood alone focuses on maximizing predictive performance, (FlA+Gini) considers both classification efficiency and category purity. The (FlA+Gini) method selects those features that help increase sensitivity and specificity, making it a more balanced and reliable system for diagnosis. This is important for its usage in the healthcare sector, where it is necessary to be able to identify both patients and healthy people correctly. Thus, according to all the experiments conducted, it was found out that the usage of a Gini index in the Flood algorithm significantly improves the quality of feature selection, making (FlA+Gini) a very promising method for disease prediction.

**Table 7.Feature selection testing results of FLA + Gini**

| models | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **MLP** | 0.826 ± 0.023 (0.798–0.854) | 0.846 ± 0.012 (0.830–0.862) | 0.908 ± 0.013 (0.892–0.924) |
| **TabNet** | 0.840±0.048 (0.781–0.899) | 0.819 ± 0.067 (0.736–0.902) | 0.874 ± 0.052 (0.809–0.939) |
| **LR** | 0.830 ± 0.027 (0.796–0.864) | 0.858 ± 0.012 (0.843–0.873) | 0.905 ± 0.017 (0.884–0.926) |

| models | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **XGB** | 0.834 ± 0.027 (0.801–0.867) | 0.858 ± 0.014 (0.840–0.876) | 0.928 ± 0.015 (0.909–0.947) |
| **RF** | 0.866 ± 0.014 (0.849–0.883) | 0.884 ± 0.004 (0.879–0.889) | 0.928 ± 0.007 (0.919–0.937) |
| **SVM** | **0.905 ± 0.019 (0.881–0.929)** | **0.914 ± 0.018 (0.892–0.936)** | **0.957 ± 0.014 (0.940–0.974)** |

**Table 8.Feature selection testing results of FLA + Gini**

| Models | #Features | Name of selected feature |
|---|---|---|
| **MLP** | 8 | sex, cp, thalach, oldpeak, ca, thal, trestbps, slope |
| **TabNet** | 7 | sex, cp, thalach, exang,oldpeak, ca, thal |
| **LR** | 8 | sex, cp, trestbps, restecg, thalach, oldpeak, ca, thal |
| **XGB** | 7 | sex, cp, trestbps, ca, thal, oldpeak, slope |
| **RF** | 8 | cp, trestbps, exang, slop, ca, thal, fbs |
| **SVM** | **8** | **age, cp, restecg, exang, slope, ca, thal, thalach** |

**Experiment (4): Classification with Feature Selection Using Greedy Algorithm and Gini Index**

The final stage of this work focuses on classification using the selected features obtained from the previous feature selection phase. To improve performance, a Greedy Search strategy was employed for hyper-parameter tuning of six models, namely MLP, TabNet, LR, RF, SVM, and XGBoost.Before presenting the results, a brief overview of the employed classification models and the adopted hyper-parameter optimization strategy is provided.

**a) Model Configuration and Hyper-Parameter Tuning**

The a nested stratified 5-fold cross-validation framework was adopted to ensure a reliable and unbiased evaluation.where the outer 5-fold cross-validation was used for final model evaluation, while an inner 3-fold cross-validation was employed for hyper-parameter tuning and model selection,within each inner fold, a Greedy Search approach was applied to identify the optimal hyper-parameter configuration. The algorithm evaluates parameter values alliterative and selects the value that achieved the highest performance while keeping the remaining parameters fixed. The search process continues until no improvement of performance.To prevent information leakage, all preprocessing were performed on the training data within each fold,including missing value imputation, feature scaling, and SMOTE oversampling, Validation and testing data remained unseen during parameter optimization.

**Multilayer Perceptron (MLP)**

The MLP was defined with fully connected layers with the Rectified Linear Unit activation function and Adam Optimizer. Early stopping and L2 regularization was done to avoid overfitting. Greedy Search was performed on the following hyperparameter :

- hidden_layer_sizes ∈ {(64), (128), (256), (128,64), (256,128), (512,256)}
- alpha ∈ {0.00001, 0.0001, 0.001, 0.01, 0.1}
- learning_rate_init ∈ {0.00005, 0.0001, 0.001, 0.005, 0.01}

**TabNet**

The classifier of the TabNet model was implemented with PyTorch-TabNet library. The initial configuration included Adam optimization, early stopping, and batch normalization mechanisms designed for tabular data. The following parameters were optimized using Greedy Search:

- n_d_options = [16, 64]
- n_a_options = [16, 64]
- n_steps_options =[3, 5]

The optimizer alliterative evaluated candidate values and selected those maximizing classification performance.In addition to MLP and TabNet, four classifiers were investigated, namely LR, RF, SVM, and XGBoost. For each classifier, Greedy Search was employed to optimize the most hyper-parameters based on validation

performance within the inner cross-validation folds. The optimization process aimed to identify parameters that maximize accuracy while maintaining model computational efficiency. Similar to MLP and TabNet, all tuning were performed on the training data to ensure a fair and unbiased evaluation.The initial hyper-parameter settings for all classifiers were selected according to commonly configurations reported in the literature.the Greedy Search strategy adapted these settings to the characteristics of the heart disease dataset. This approach identifies effective parameter combinations and eliminates the need for manual tuning.After optimization, all classifiers were evaluated using nested 5-fold cross-validation. The performance comparison of the classification models is presented in Table 9. Furthermore, the feature selection results indicated that only seven features (sex, cp, trestbps, ca, thal, oldpeak, slope) were sufficient to achieve the best performance by XGBoost model. The selected features are reported in Table 10 **.**

**Table 9.The performance of the models with greedy search**

| models | Precision (mean ± std, 95% CI) | F1-score(mean ± std, 95% CI) | AUC(mean ± std, 95% CI) |
|---|---|---|---|
| **MLP** | 0.949 ± 0.031 (0.910–0.988) | 0.954 ± 0.013 (0.938–0.970) | 0.990 ± 0.003 (0.986–0.994) |
| **TabNet** | 0.860 ± 0.069 (0.774–0.946) | 0.872 ± 0.048 (0.812–0.932) | 0.942 ± 0.035 (0.898–0.986) |
| **LR** | 0.829 ± 0.027 (0.795–0.863) | 0.862 ± 0.015 (0.844–0.880) | 0.906 ± 0.017 (0.885–0.927) |
| **XGB** | **0.989 ± 0.025 (0.958–1.000)** | **0.990 ± 0.023 (0.961–1.000)** | **0.998 ± 0.004 (0.993–1.000)** |
| **RF** | 0.976 ± 0.031 (0.938–1.000) | 0.978 ± 0.028 (0.943–1.000) | 0.998 ± 0.004 (0.993–1.000) |
| **SVM** | 0.963 ± 0.035 (0.919–1.000) | 0.969 ± 0.033 (0.929–1.000) | 0.987 ± 0.022 (0.960–1.000) |

**Table 10.Feature selection after adding greedy search and FLA + Gini**.

| Model | # Features | Name of selected feature |
|---|---|---|
| **MLP** | 8 | sex, cp, thalach, oldpeak, ca, thal, trestbps, slope |
| **TabNet** | 7 | sex, cp, thalach, exang, oldpeak, ca, thal |
| **LR** | 8 | sex, cp, trestbps, restecg, thalach, oldpeak, ca, thal |
| **XGB** | **7** | **sex, cp, trestbps, ca, thal, oldpeak, slope** |
| **RF** | 7 | cp, trestbps, exang, slope, ca, thal, fbs |
| **SVM** | 8 | age, cp, restecg, exang, slope, ca, thal, thalach |

The experimental results demonstrated that the proposed framework achieved strong performance for heart disease , where the XGB classifier achieved 98.9%, the highest overall accuracy .The best hyper-parameters after greedy search with XGBoost {learning_rate: 0.1,subsample: 1.0,n_estimators: 300,max_depth: 5}

while computational efficiency, were differences observed across the classifiers. TabNet exhibited the highest computational cost, approximately 5013.88 minutes . This increased runtime can be attributed to its deep learning architecture, sequential attention mechanism, and the additional computational burden introduced by the nested cross-validation , feature selection , and greedy hyper-parameter strategy.Conversely, GXB achieved the lowest runtime, approximately 57.37 minutes. Its computational efficiency stems from XGBoost, which supports parallel tree construction and fast convergence, in addition to the reduced feature space obtained through the Gini and FLA-based feature selection process. Furthermore, the Greedy Search strategy efficiently identified suitable hyperparameter values without exploring the entire search space and optimization time while maintaining predictive performance.The reported 95% confidence intervals provide additional evidence regarding the reliability of the proposed framework. the narrow confidence intervals indicate stable performance across different data partitions and demonstrate strong model generalization. but wider confidence intervals suggest greater sensitivity to data variability. The XGB and RF models produced the narrowest confidence intervals, followed by MLP, indicating consistent predictive behavior across the cross-validation folds.To further validate the effectiveness of the proposed hybrid framework, statistical

significance analysis was performed using a paired t-test on the results obtained from the 5-fold cross-validation experiments. The statistical analysis confirmed that the observed performance improvements across the different stages of the framework were not due to random variation.Furthermore, the incorporation of the feature selection stage led to a substantial improvement in predictive performance compared with the baseline models trained using all available features. Although the integration of the Gini - based optimization strategy yielded additional performance gains, the greatest improvement was achieved during the greedy hyper-parameter optimization stage, as reported in Table 11. These findings demonstrate the effectiveness of the proposed framework in enhancing accuracy, improving model stability, and ensuring statistically reliable performance improvements.

**Table 11.Paired t-test between stages all algorithms.**

| | **S1 vs S2 (p-value)** | **S2 vs S3 (p-value)** | **S3 vs S4 (p-value)** | **S1 vs S4 (p-value)** |
|---|---|---|---|---|
| **MLP** | No(p=0.0751) | No (p =0.5823) | Yes (p =0.0001) | Yes (p = 0.0226) |
| **TabNet** | No (p = 0.2486) | No (p =0.1279) | Yes (p = 0.0382) | No (p = 0.0843) |
| **LR** | Yes (p = 0.0412) | No (p = 0.5275) | Yes (p = 0.0007) | Yes (p = 0.0087) |
| **XGB** | **No (p = 0.2114)** | **Yes (p = 0.0329)** | **Yes (p =0.0007)** | **Yes (p = 0.0001)** |
| **RF** | No (p = 0.3616) | No (p = 0.4263) | Yes (p = 0.0013) | Yes (p = 0.0047) |
| **SVM** | Yes (p = 0.0508) | No (p = 0.3046) | Yes (p = 0.0125) | Yes (p = 0.0358) |

**Comparative Evaluation: FLA vs. GA**

Building on our previous work [33], where a genetic algorithm (GA) was employed for feature selection across the same models, we noted that the Random Forest model achieved peak performance using a subset of eight optimal features {sex, cp, trestbps, exang, slope,ca, thal,oldpeak,}. As illustrated in Fig.5, comparing the Flood Algorithm (FLA) against the GA clearly demonstrates that the FLA yields superior performance.

Fig .5: Comparison of performance between GA and FLA Algorithm**s**

## 5. CONCLUSION

This study proposed a hybrid framework that integrates the Flood Algorithm (FLA) for feature selection, a Gini index-based fitness function for improving class discrimination, and Greedy Search for hyper-parameter optimization for heart disease prediction ,the proposed framework simultaneously considers feature relevance, class purity, and classifier optimization,unlike conventional feature selection approaches that optimize only accuracy, resulting in a more reliable diagnostic system.The experimental was conducted using six models MLP, TabNet, Logistic Regression, Random Forest, Support Vector Machine, and XGBoost under a nested stratified 5-fold cross-validation protocol. The results demonstrated that each stage contributed to improving predictive performance, while the final stage produced the greatest performance gains.The XGBoost model achieved the best performance and maintaining the lowest computational time . RF and MLP also demonstrated height predictive capability and stable confidence intervals, whereas TabNet exhibited comparatively higher computational cost .The proposed framework successfully reduced the feature space while preserving the most clinically relevant features. where only seven to eight features were sufficient to achieve predictive performance across different classifiers.Furthermore, the statistical analysis confirmed the effectiveness of the proposed framework. The paired t-test demonstrated statistically significant improvements in the final optimization stage for almost all classifiers, indicating that the observed performance gains were not caused by random variation. The narrow 95% confidence intervals obtained by XGBoost, RF, and MLP further verified the stability and generalization capability of the proposed methodology. The proposed methodology improves predictive accuracy, enhances model stability, reduces feature dimensional, and maintains computational efficiency, making it a promising decision-support tool for intelligent clinical diagnosis.Future work will focus on validating the proposed framework using larger datasets incorporating explainable artificial intelligence techniques to improve model interpret ability, and extending the proposed methodology to other medical diagnosis applications .

**References**

1. K. P. Murphy, Probabilistic Machine Learning: An Introduction. Cambridge, MA, USA: MIT Press, 2022.
2. Y. Wu et al., “Partial multi-label feature selection with feature noise,” Pattern Recognition, vol. 162, 2025.
3. M. Ghasemi, K. Golalipour, and M. Zare, “Flood algorithm (FLA): An efficient inspired meta-heuristic for engineering optimization,” The Journal of Super computing, 2024.
4. J. Zhu et al., “An integrated approach of feature selection and machine learning for early detection of breast cancer,” Scientific Reports, vol. 15, no. 1, p. 13015, 2025.
5. F. Rahman, S. Hossain, J.-J. Tiang, and A.-A. Nahid, “Diabetes prediction using feature selection algorithms and boosting-based machine learning classifiers,” Diagnostics, vol. 15, no. 20, p. 2622, 2025.
6. K. Liu, Q. Chen, and G.-H. Huang, “An efficient feature selection algorithm for gene families using NMF and Relief F,” Genes, vol. 14, no. 2, 2023.
7. Y. Ma, Y. Xiao, and W. Zhang, “Relief F feature selection and a SSA-IRF model for continuous blood pressure estimation from ECG and PPG signals,” International Journal of Biology and Life Sciences, 2025.
8. A. Desiani, R. Hidayat, and I. Wibowo, “The comparison of Relief and C4.5 for feature selection on heart disease classification using back propagation,” Indonesian Journal of Computing and Cybernetics Systems, vol. 17, no. 2, 2023.
9. M. Radwan, A. A. Alhussan, A. Ibrahim, and S. M. Tawfeek, “Potato leaf disease classification using optimized machine learning models and feature selection techniques,” Potato Research, vol. 68, no. 2, pp. 897–921, 2024.
10. H. Z. Mojahid et al., “Examining the impact of feature selection techniques on machine and deep learning models for the prediction of COVID-19,” Malaysian Journal of Computing, vol. 10, no. 1, pp. 2135–2158, 2025.
11. U. Ahmed et al., “Hybrid bagging and boosting with SHAPE-based feature selection for enhanced predictive modeling in intrusion detection systems,” Scientific Reports, vol. 14, Art. no. 30532, 2024.
12. H. Zouhri, A. Idri, and A. Ratnani, “Evaluating the impact of filter-based feature selection in intrusion detection systems,” International Journal of Information Security, vol. 23, no. 2, 2023.
13. M. A. Hossain and M. S. Islam, “A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection,” Scientific Reports, vol. 13, Art. no. 21207, 2023.
14. H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar, “Feature selection strategies: A comparative analysis of SHAPE-value and importance-based methods,” Journal of Big Data, vol. 11, Art. no. 44, 2024.
15. M. Fatima, O. Rehman, I. M. H. Rahman, A. Ajmal, and S. J. Park, “Towards ensemble feature selection for lightweight intrusion detection in resource-constrained IoT devices,” Future Internet, vol. 16, no. 10, 2024.
16. H. Zouhri, A. Idri, and A. Ratnani, “Evaluating the impact of filter-based feature selection in intrusion detection systems with random forest,” International Journal of Information Security, 2023.
17. B. Ahmad, J. Chen, and H. Chen, “Feature selection strategies for optimized heart disease diagnosis using ML and DL models,” arXiv p reprint, 2025.
18. R. Kazerani, “Improving breast cancer diagnosis accuracy by particle swarm optimization feature selection,” International Journal of Computational Intelligence Systems, vol. 7, 2024.
19. Z. Noroozi, A. Orooji, and L. Erfannia, “Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction,” Scientific Reports, 2023.
20. I. B. S. Mahendra et al., “Implementation of feature selection chi-square to improve the accuracy of the classification model using the random forest algorithm on coronary artery disease,” Journal of Applied Intelligent Systems, 2023.
21. S. Moorthy, “A novel feature selection approach with integrated feature sensitivity and feature correlation for improved prediction of heart disease,” Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 1–15, 2022.
22. I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” Journal of Machine Learning Research, vol. 3, pp. 1157–1182.
23. O. Roubhi, A. Taha, and S. Benkraouda, “Joint mutual information-based feature selection for speech emotion recognition,” Engineering, Technology & Applied Science Research, vol. 17, no. 1, 2025.
24. R. Rayan, M. Khalid, and F. Al-Hamadi, “Modified mutual information feature selection for COVID-19 prediction,” Journal of Medical Systems, vol. 48, no. 7, 2024.
25. O. Oladimeji, A. Adept, and H. Bello, “Mutual information-based radiomic feature selection for breast cancer diagnosis,” Radiography, vol. 30, pp. 215–226, 2024.
26. M. Ahmad, S. Khan, and R. Ali, “Mutual information for heart disease prediction: Comparative study across different models,” arXiv preprint arXiv:2503.16577, 2025.
27. V. Gutiérrez, C. Tejada, and J. Tejada, “Evaluating feature selection methods for accurate diagnosis of diabetic kidney disease,” Bio medicines, 2024.
28. M. H. Nikzad et al., “A novel systematically optimized TabNet model for prediction tasks,” Array, 2024.
29. F. Zhou et al., “ACDIM: A cardiovascular disease risk prediction model based on TabNet and deep learning,” Electronics, vol. 13, no. 24, 2024.
30. M. T. Hossain, M. S. Hossain, and K. F. Al Amin, “Heart disease prediction using optimized multilayer perceptron neural network,” IEEE Access, vol. 8, pp. 13456–13467, 2020.

31. D. A. Wagner, “The Gini learning index: A new framework to measure learning inequality across contexts,” International Journal of Educational Development, 2025.
32. T. Hastie, R. Tibshirani, and J. Friedman, “The Elements of Statistical Learning: Data Mining,” Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009.
33. J. Amshaher and A. Elhawil, “Feature selection using optimization algorithms in heart disease detection”, Academy Journal for Basic and Applied Sciences, vol. 8, no. 1, pp. 1–11, May 2026, doi: 10.5281/zenodo.20345015.