Feature selection using optimization algorithms in heart disease detection

Jazya Amshaher
Dept. of electrical and computer engineering
Libyan academy for graduate studies
Tripoli, Libya
jazamoftah@su.edu.ly
Prof.Amna Elhawil
Dep.of computer engineering
faculty of engineering
university of tripoli
Tripoli, Libya
a.elhawil@uot.edu.ly

***Abstract*—Heart disease continues to be among the leading causes of mortality worldwide, necessitating its early detection for improving patient prognosis. Although conventional statistics-based algorithms and feature selection approaches have been applied for heart disease prediction purposes, the complex nature of medical data may not necessarily be accounted for using such techniques. Removing unnecessary features improves how well machine learning algorithms perform. In this study, we present a step-by-step feature selection methodology leveraging Genetic Algorithms (GA) and the Gini index. Utilizing greedy search for parameter tuning and and a comprehensive suite of confusion matrix metrics for evaluation, we compared six architectures: MLP, TabNet, Random Forest, SVM, Logistic Regression, and XGBoost are applied to the heart disease dataset. Our results demonstrate that the Random Forest (RF) model delivers superior diagnostics. By reducing the feature space from 13 to 8, the RF model attained exceptional predictive power, charting 98.4% accuracy, 98.9% sensitivity, 98% specificity, and a 98.5% F1-score.**

Keywords—Heart Disease, Feature Selection, Genetic Algorithm

I. INTRODUCTION

Heart diseases continue to be one of the major sources of mortality globally, resulting in about 17.90 million mortalities each year [1], [2]. Early detection of this condition is, hence, critical in increasing patient survival rates. ML algorithms have extensively been utilized in heart disease prediction owing to their ability to work on complicated medical data sets [3]. Medical data sets are characterized by numerous unnecessary attributes which lower the precision levels of the algorithms while increasing their computational complexity. Many researchers, in past investigations, were concerned only with the maximization of accuracy.This paper introduces a framework for Multi-Objective Optimization that combines Genetic Algorithm (GA) based feature selection, class balancing using the Gini index, and greedy hyper parameters tuning. The Gini index is included in the GA fitness function to maximize both accuracy and class balancing. The developed framework is tested through nested stratified cross-validation on different classifiers such as MLP,TabNet, Random Forest, SVM, Logistic Regression, and XGBoost. The main contributions of this work are summarized as follows:

- A hybrid multi-objective feature selection framework combining GA and Gini index to guide feature selection toward high accuracy and class balance.
- Integration of greedy search to automatically identify optimal hyper parameters.
- A robust evaluation protocol using nested stratified cross-validation to avoid data leakage.
- Analysis of feature importance in clinical context.
- The method was evaluated using the MLP, TabNet and Comprehensive comparison with strong baseline models.
- Providing a comprehensive evaluation using nested stratified cross-validation, confidence intervals, and statistical significance testing.

II. LITERATURE REVIEW

In feature selection, retaining high-quality, relevant data is critical, as the accuracy of the underlying dataset establishes the baseline for the model's evaluation metrics. For this purpose numerous methods have been applied in various domains. First, Escamila et al. suggested a dimensional reduction technique to discover characteristics of heart disease. Techniques for feature selection, including chi-square feature selection and principal component analysis (CHI-PICA), were employed to extract features[5]. The methodology employed was carried out by applying different classifiers like decision trees, gradient boosted tree, logistic regression, MLPs, naive Bayes, and random forests to three

DOI: **10.5281/zenodo.20345015**

datasets – Cleveland, Hungary, and CH. In all the three datasets, the maximum accuracy recorded was that of the random forest classifier which scored 89.7%, 91.0%, and 94.4% respectively. The study confirmed that due to the short sample size, it is difficult to widen the scope to encompass all heart conditions.

However, on the Cleveland Clinic cardiology data set, Dissanayake et al. used a variety of feature selection methods, including ANOVA, Chi-square, mutual information, forward feature selection, backward feature selection, redundant feature removal, Lasso regression, and Ridge regression, along with classification algorithms like decision tree, random forest, support vector machine, K-nearest neighbor, logistic regression, and Gaussian naive Bayes. The achieved accuracy was 88.52% with the decision tree classifier[6].

Murthy et al. [7] enhanced prediction of heart disease with the use of an integrated random forest feature sensitivity and correlation (RF-FSFC) approach. The trait association analysis phase looks at whether qualities are connected with coronary heart disease risk, while the sensitivity-based trait selection procedure ranks traits according to their value. The proposed methodology reached an accuracy reached 81.16%. However, feature sensitivity requires extensive analysis, which may affect the model's efficiency.

GanesaLingam and others predicted heart failure using feature selection techniques [8]. They applied two feature selection techniques, forward selection and backward selection, using the Cleveland heart disease datasets. To increase performance, they employed classification techniques as Naive Bayes (NB), Decision Trees (DT), Logistic Regression SVM (LR-SVM), and Logistic Regression (LR). The Decision Tree algorithm's accuracy was 88.51%. However, this method may not always find the global optimal input.

Jusia et al. [9] enhanced the feature selection procedure for heart disease prediction utilizing the Cleveland data set, the particle swarm optimization (PSO) algorithm, and K-NN and Decision Tree algorithms. The experimental results demonstrated that the particle swarm optimization algorithm performed well when applied to the combined datasets and the SF-2 feature subset, achieving an accuracy of 89.09%. However, its limitations become apparent when a swarm reaches a certain stage of convergence; its movement may slow, affecting the swarm's ability.

Fahad et al. employed the Cleveland dataset's arithmetic optimization approach to choose features. Using filter techniques for data set prepossessing, a novel AOA for feature selection, and a multi layer perceptron neural network to classify the data set into two classes, the prediction model achieves an accuracy of 88% on the datasets under consideration. However, when dealing with a lot of features, it can decrease the algorithm's efficiency [10].

Manikandan et al. suggested better cardiac disease prediction by identifying relevant features from the Cleveland Clinic Heart data set using the Boruta feature selection method [11]. Until every item has been approved or rejected, remove any features that are superfluous from the feature list. examined machine learning techniques, such as logistic regression, decision trees, and support vector machine (SVM) learning, to diagnose cardiac abnormalities; logistic regression yielded the highest accuracy of 88.52%.

Sharma [12] conducted a comparison to see how effectively various feature selection techniques and machine learning algorithms predicted cardiac disease. They discovered that the best accuracy of 73.74% was obtained when an XG Boost classifier and the wrapper strategy were combined. The study highlights the importance of choosing the right features to improve prediction accuracy.

Ahmad et al. [13] The paper compares various feature selection strategies, including Chi-Square, analysis of variance, and mutual information, and shows how these methods affect the effectiveness of deep learning and machine learning algorithms for heart disease prediction. According to the findings, Mutual Information (MI) outperformed other techniques, particularly neural networks, with an accuracy of 82.3%.

Ahmed et al. [14] improved the prediction of cardiovascular risk by combining multiple neural constructs and feature filtering techniques. This method achieved an accuracy of 89.2%. Although the robustness was improved compared to single-algorithms, but this the method's was complexity and the need for large training samples.

Patel et al. [15] investigated a transformer-based framework heart disease prediction that uses self-attention mechanisms to extract complex nonlinear patterns from on a Cleveland dateset. They evaluated the model on a Cleveland dateset, achieving an accuracy of 88.5%. However, due to a small and unbalanced data set , they suggested that future work would utilize a larger data set.

Zhang and Chen [16] applying recursive feature elimination integrated with conventional neural networks (CNN) for coronary heart disease diagnosis. After applying approach reached an accuracy of 87.8% on the combined data set and reduced dimensional and computational complexity.

Gupta and Sharma [17] used genetic programming for feature filtering with a deep learning classifier for heart disease prediction. Their hybrid model resulting in 86.7% accuracy on benchmark datasets. but this study noted results was instability across different data set splits.

Many deep learning algorithms have been successfully applied in predictiveanalytic .instead of processing tabular datasets, some algorithms, such convolution neural networks (CNN) and recurrent neural networks (RNN), are made to handle image and sequential data [18]. The model may choose the most pertinent features thanks to TabNet's sequential attention method, which was created especially for tabular data [19]. This feature allows Tab Net to detect complex nonlinear relationships while simultaneously performing feature selection. Furthermore, Tab Net provides an interpret able framework by highlighting the contribution of each feature during the decision-making process c ompared to traditional deep neural networks, Tab Net demonstrated competitive performance on tabular datasets while maintaining computational efficiency and improving interpret-ability[20] .Therefore ,Tab Net was chosen in this study because it offers a balanced combination of predictive power, automatic feature selection, andinterpret-ability , making it particularly suitable for diagnosing heart disease.

DOI: **10.5281/zenodo.20345015**

Mohammed et al. concentrated on identifying the key characteristics from electronic health records in order to leverage multi-modal data to forecast Alzheimer's disease in its early stages. This approach incorporated Fisher's score and greedy search as feature selection criteria with the SVM K-NN classifier, achieving 90% accuracy [21]. However, this method has the limitation that it selects the proposed features from a minimum set, rather than finding the optimal set.

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

Particle Swarm Optimization (PSO) has been used in other studies [23][24][25] for feature selection in breast cancer prediction. In essence, the feature selection algorithm is coupled with machine learning techniques. The findings showed that the PSO improves the diagnostic precision of breast cancer identification.

They proposed modern feature selection techniques for tabulated medical datasets. This method combined HAP-based feature selection with procedures such as Boruta, Recursive Feature Elimination (RFE), and embedded L1 regularization. Then selected features used through XG Boost, Light GBM, and Cat Boost,the FS adds overhead. The method significantly reduced feature dimensions and achieved accuracy of up to 92.3% on cardiac datasets[26]. The study emphasized that relying on manually set thresholds for SHAP estimation limits real-time applicability, especially when dealing with large datasets.

They proposed a modern feature selection technique using mutual information scoring and Boruta as pre-processing steps. With the Tab PFN and XG Boost algorithms Evaluations Results indicated that Tab PFN algorithms achieve superior performance 92% compared to XG Boost algorithms 90%[27]. However, the limitation of this method lies in the high cost of pr-training.

Current feature-selection-based heart disease prediction methods rely primarily on classification accuracy. It results in biased algorithms within medical datasets. In addition, prior research has mostly considered feature selection and hyper parameter tuning as two independent tasks, affecting the performance of the entire algorithm negatively. To address this limitation, a hybrid framework was developethat integrates a Gini into the genetic algorithm's fitness function, transforming feature selection into a multi-objective optimization problem. This results in simultaneous optimization of classification accuracy and class balance, crucial for reliable medical diagnosis.

Additionally, a greedy search strategy for hyper-parameter optimization was proposed, allowing for the co-optimization of subsets of features. This unified design distinguishes the proposed method from traditional packaging and filtering approaches. Previous research has not combined a single framework to integrate the a genetic algorithm, Gini based fitness , and greedy hyper-parameter optimization in the context of heart disease prediction using tabular datasets. Therefore, this work presents a process for improving both predictive performance and clinical reliability.

III. BACKGROUND

A. TabNet Model

Many deep learning algorithms have been successfully applied in predictive analytic. Instead of processing tabular datasets, some algorithms, such convolution neural networks (CNNs) and recurrent neural networks (RNNs), are made to handle image and sequential data [28]. The model may choose the most pertinent features thanks to Tab Net's sequential attention method, which was created especially for tabular data [29]. This feature allows TabNet to detect complex nonlinear relationships while simultaneously performing feature selection. Furthermore, TabNet provides an interpreter framework by highlighting the contribution of each feature during the decision-making process compared to traditional deep neural networks, TabNet demonstrated competitive performance on tabular datasets while maintaining computational efficiency and improving interpret-ability[30]. In conclusion, therefore, TabNet was used in this study as it has a balance of being able to predict, select features automatically, and interpret, hence its suitability for heart disease diagnosis.

B. Multilayer Perceptron (MLP)

An artificial neural network referred to as MLP contains neurons with connections between different layers. One or more hidden layers, an output layer, and an input layer are all present. An MLP is used to simulate intricate interactions between inputs and outputs. The main components of Multi-Layer Perceptron (MLP)

- Input Layer: every neuron is associated with a feature.
- Hidden Layers: MLP can have any number of hidden layers, with any number of nodes in each layer. The input layer provides the data that these layers process.
- Output Layer: is responsible for producing the final prediction. A corresponding number of neurons will be present in the output layer if there are numerous outputs.

IV. METHODOLOGY

There are four experiments in the proposed methodology:

A. Classification without feature selection

B. Classification with feature selection

C. Classification using the Gini Index for feature evaluation

D. Classification using a greedy search approach.

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

DOI: **10.5281/zenodo.20345015**

A. Classification without feature selection

In this experiment, the six classification models are used without any optimization.

B. Feature Selection using Genetic Algorithm (GA)

Parameters of the Genetic Algorithm (GA) were set in order to guarantee repeatability of the feature selection. The choice of GA is based on its efficiency in conducting global optimization through the process of feature space search. Unlike conventional sequential algorithms that evaluate only one feature subset at any one time, GA is able to evaluate multiple feature subsets at once. Specifically, the genetic algorithm involved a population size of 200 individuals and ran for 50 generations. One-point crossover operation with probability p = 0.8 was adopted together with bit-flip mutation technique with mutation probability m = 0.05. Moreover, elitism scheme was used, according to which top 30% individuals were carried over to the next generation. In order to assure experimental repeatability, a seed number was set to be 42.A one-point crossover operator was applied to generate offspring, where pairs of parent chromosomes exchange segments at a randomly selected crossover point. After that Mutation was implemented using a bit-flip operator with a mutation probability of 0.05, because it provided a balance between exploration and exploitation, where selected genes could switch between 0 and 1. The algorithm was executed for a fixed number of 50 generations, which served as the termination criterion. The fitness function is based on only accuracy.

C. Genetic Algorithm (GA) with Gini Index

In order to enhance the quality of feature selection for the predictive model (MLP) or TabNet in heart disease prediction, the Gini index was employed in the genetic algorithm's fitness evaluation in this stage paper of the studies. The Gini index measures the level of inequality in the resulting groups after classification. The Gini index is included in the fitness function to provide class balance and minimize bias in the predictions. This is because in most cases where there are medical data sets, there tends to be imbalance in the classes. This makes use of only the accuracy measure lead to biased algorithms due to their preference for the majority class. The mathematical definitions in equation 1 of the Gini impurity are [31] and [32] :

$$\text{Gini}_i = \sum_{k=1}^{c} p^2 k$$

where C is the number of classes and pk is the probability of class k.

When the Gini number is lower, the sample is more homogeneous, and vice versa. The objective is to minimize the disparity between classes while increasing the model's accuracy. The (1 − Gini) formula is 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 function of GA algorithm has been designed based on a multi-objective optimization perspective modified as equation 2 :

$$\text{Fitness}_i = \alpha * \text{Accur}_{i+}(1-\alpha)(1-\text{Gini}_i) \quad (2)$$

Where:Accuri: the accuracy of model.

Gini: value the model's results in the generation.

α: The balance factor between accuracy and balance.

The fitness function was defined as a combination of classification accuracy and class balance using the Gini index. However, the parameter α controls the trade-off between these two objectives. The sensitivity analysis was conducted using multiple values α ∈ {0.2, 0.4, 0.6 and 0.8}. The results showed trade-off: lower α values emphasized class balance, leading to higher sensitivity but reduced specificity and accuracy, whereas higher α values improved accuracy and specificity at the expense of a slight decrease in sensitivity. Intermediate values such as α = 0.6 provided a more balanced performance between sensitivity and specificity. These findings highlight that α plays a crucial role in controlling the bias of the model, which is important in medical diagnosis. However, in this work α is selected to be 0.8 to give higher importance to classification accuracy whereas the class imbalance is still considered.The feature selection process is framed as a multi-objective optimization problem.Were two objectives are optimized simultaneously: (1) maximizing classification accuracy, and (2) improving class balance,instead the fitness function is based on only accuracy , it will rely on both accuracy and balance optimization, allowing the genetic algorithm to search for subsets of features that achieve a balance between performance and class distribution. This is important because class imbalance can lead to biased predictions.

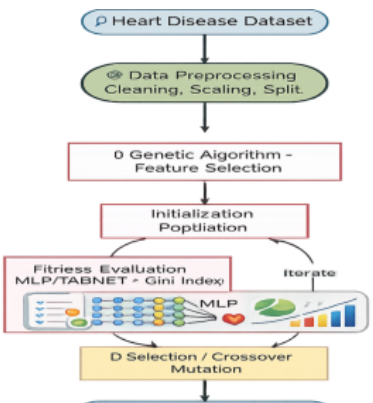
It is important to clarify that the GA + Gini framework performs pure feature selection rather than implicit feature extraction. Where the genetic algorithm selects optimal subsets from the features without creating new derived features, while Gini index is integrated into the fitness function to guide the search toward subsets that improve both accuracy and class balance. This is important in medical datasets where class imbalance may lead to biased algorithms with diagnostic decisions. According to the following steps:

1) Initial Population: a number of subsets of candidate features are generated.

2) Training algorithms for each feature subset: For each subset, a model is trained to calculate the accuracy and the Gini measure based on the prediction results. For each subset, the fitness is calculated based on the combination of accuracy and the (1−Gini) measure.

3) Selection: Individuals with the highest fitness are selected to pass to the next generation.

4) Crossover and Mutation: Genetic Algorithm operations are performed to generate feature subsets with the aim of further explore the solution space.



DOI: **10.5281/zenodo.20345015**

Fig 1. The proposed approach

D. Optimizing models using Greedy Search.

Optimizing hyper-parameters is essential for enhancing machine learning models performance. In this study, a greedy search strategy was using to tune the hyper parameters of the classification models ,this approach involves alliterative selecting hyper parameters that improve model performance, typically by evaluating hyper parameter and selects the configuration that maximizes the classification accuracy , greedy search is valued for its simplicity and efficiency, making it a practical choice when dealing with the complex and often high dimensional hyper parameter spaces of deep neural networks learning algorithms. By focusing on local improvements, it can effectively enhance the predictive within a reasonable computational budget.

Cross-validation Strategy

The framework adopted was nested stratified 5-fold cross-validation, where used the outer 5 folds for unbiased model evaluation, whereas the inner 3 folds were dedicated to feature selection using the (GA) and a greedy search strategy was applied for hyper-parameter tuning, all preprocessing, including feature scaling and SMOTE, were performed within the training of the inner folds only to avoid data leakage. The final performance represent the average performance across all outer folds.

V. RESULTS AND ANALYSIS

This section presents the experimental results. All performance metrics in the proposed framework were obtained using nested stratified 5-fold cross-validation. The reported values represent the mean performance across the outer folds. Where 95% confidence intervals were computed across the outer folds to assess the reliability of the proposed framework, in addition paired statistical significance testing using a paired t-test was performed to determine the differences of performance between algorithms were statistically significant, across the outer cross-validation folds.

A. Experimental Setup

All experiments were conducted out on a machine equipped with an Intel Core i7-10810U CPU environment and RAM was 32GB, operating under the Windows platform. The experiment implementation was using Python using machine learning libraries.

B. Dataset collection

Heart disease dataset, is used for heart disease prediction and is publicly available on Kaggle. was released by the UCI heart Disease Data Set. For scholars, it is a common freely accessible data set. Fig. 2 show distribution of patients based on diagnosis status.

The data set contains 1025 record of 13 features gathered over several time periods from patients during the Clinic Foundation study [33].The Heart Disease Data set's primary goal is to identify people who are at risk for heart disease early on. The features are listed in Table 1.

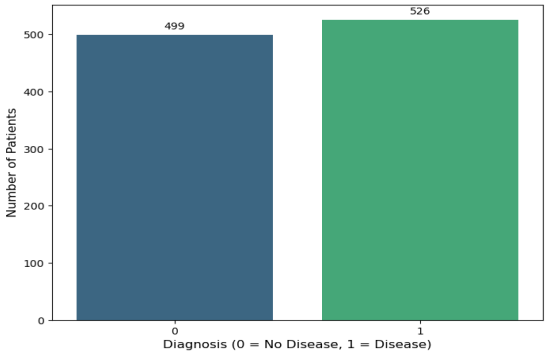


Fig 2. Distribution of patients based on diagnosis status

TABLE I. 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

C. Data preprocessing

Before model development, the data set inspected for missing values, feature distributions, and class imbalance. For improve data quality, missing values were handled using median `Simple Imputer` technique. After that, feature scaling was applied using the `Standard Scaler` during the training process. where transformed the features into scale with zero mean and unit variance, then used the Synthetic Minority Over-sampling Technique (SMOTE) for address Class imbalance. SMOTE was applied to the training folds within

DOI: **10.5281/zenodo.20345015**

the cross-validation to ensure fair model evaluation, as used a Nested Cross-Validation strategy was employed to prevent over fitting. The outer loop was used for final model evaluation, while the inner loop was utilized for feature selection and hyper-parameter optimization. In order to systematically assess the influence of each of the components in the proposed hybrid model, the experiment flow was divided into four subsequent steps. Step one included the classification stage without performing any feature selection. In step two, feature selection based on the genetic algorithm (GA) was implemented. The third stage was characterized by implementing the Gini index in the GA fitness function with the aim to optimize the classification results and balance of classes. In the last stage, GA, Gini index, and greedy hyper parameter optimization were combined. Such an experimental setup provided an opportunity for the systematic isolation of the effect of different techniques.

Limitations of PCA

As illustrated in Fig. 3, the first two principal components account for only 33% of the total variance. This relatively low value underscores the high dimensional and complexity of the heart disease data set. Furthermore, the significant overlap between classes suggests that the underlying relationships are likely non-linear, limiting the effectiveness of linear decomposition methods like PCA for clear class separation and outlier identification as shows in Fig 3 and 4.

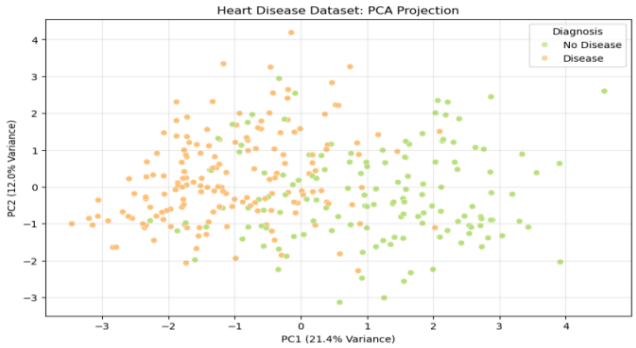


Fig 3. PCA: Visualizing cluster reparability

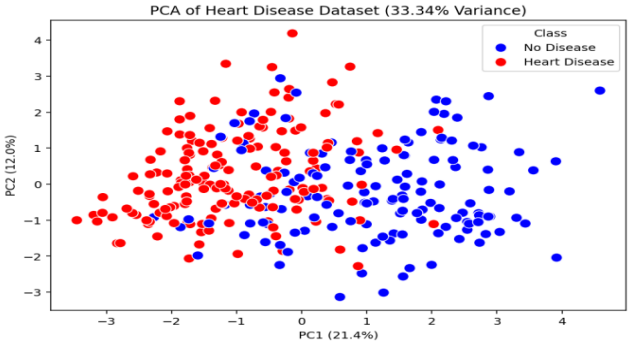


Fig 4. PCA: Multivariate outlier identification

Although the heart disease data set is of a reasonable size compared to other data set, it may suffer from problems the risk of over-fitting and reduces the generalization capability. To limit this, used the (GA) for feature selection, to a range of 3–8 features,this improve of model generalization.

Over fitting is a potential problem when using complex algorithms such as (MLPs) (up to 5000 training iterations) , TabNet and other models. To mitigate this in the MLP model, L2 regularization (alpha) controls over fitting, which is improved by using greedy search. In the TabNet model used early stopping (patience = 5) during training.The data-set contains unbalance data, which can lead to model bias towards the larger category. To address this issue, the data was stratified into two sets (training and testing) (stratify=y). Additionally, used evaluation metrics were accuracy, sensitivity (recall) and specificity. Finally, the dataset contains features with different distributions. This was handled using StandardScaler before training For models.

D. Results of Experiments

1) Experiment (a): Classification without Feature Selection

The six algorithms, the MLP, TabNet, Logistic Regression, XG Boost, Random Forest and support vector machine were used for the training data set. Without taking into account any feature selection method, the algorithms were assessed using all of the dataset's provided features. Table II lists the experiment's outcomes for every class.

TABLE II. 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)
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 show that RF classifier gives the highest accuracy of 86%. The learning curve of this model is displayed in Fig. 5.

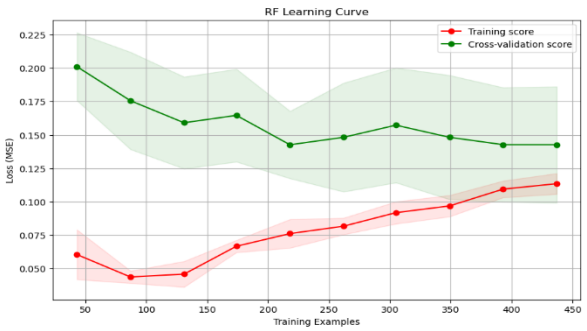


Fig 5. Learning curve of the Random Forest (RF) model

2) Experiment (b): Classification with feature selection

DOI: **10.5281/zenodo.20345015**

This experiment involves GA for feature selection using the criteria listed in Table III.

TABLE III. GA FEATURE SELECTION REQUIREMENT.

Process	Requirement
Attribute Evaluator	CFS Subset Evaluator (supervised, Class (nominal): 1 Outcome)
Search method	GA search
Requirement to Obtain the best fit features	Start set: no attribute Population size: 200 Number of iterations: 50 Mutation probability: 0.05
Random Seed	42

The results of all models are listed in Table IV. It is shown that, SVM outperformed the other models with an accuracy of 90.4% as shown in Table V, the results showed that, out of 13 features, the best-fit features are only eight using SVM classifier.

TABLE IV. PERFORMANCE OF ALL MODELS WITH (GA)

Model	Precision (mean ± std, 95% CI)	F1-score(mean ± std, 95% CI)	AUC(mean ± std, 95% CI)
MLP	0.796 ± 0.050 (0.734–0.858)	0.828 ± 0.019 (0.804–0.852)	0.883 ± 0.024 (0.853–0.913))
TabNet	0.823 ± 0.040 (0.773–0.873)	0.825 ± 0.008 (0.815–0.835)	0.913 ± 0.014 (0.896–0.930)
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	0.851 ± 0.035 (0.808–0.894)	0.866 ± 0.020 (0.841–0.891)	0.923 ± 0.012 (0.909–0.937)
RF	0.857 ± 0.025 (0.826–0.888)	0.873 ± 0.011 (0.859–0.887)	0.935 ± 0.019 (0.911–0.959)
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 V. FEATURE SELECTION TESTING RESULTS OF GA

Model	# Features	Name of selected feature
MLP	7	sex, cp, oldpeak, ca, thal, thalach, slope.
TabNet	8	sex, cp, trestbps, restecg, thalach, oldpeak, ca, thal
LR	8	sex,cp, trestbps, restecg, thalach, oldpeak,ca, thal
XGB	6	sex, cp, exang,ca, thal, slope
RF	6	sex, cp,exang, slope,ca,thal.
SVM	8	age, cp, restecg, exang, slope, ca,thal,sex

3) Experiment (c): Feature Selection Using GA and Gini Index

In this experiment, Gini measure is employed as part of GA fitness function, aiming to improve the feature selection and increase the efficiency of the models in predicting heart disease. This experiment considers the balance of performance between the two classes by introducing the (1-Gini) measure into the evaluation function. Table VI Lists a summary of the models performance.

Table VI depicts that adding Gini index to the GA fitness function has led to an increase in classification accuracy as well as class balance. Of all classifiers, SVM performed the most satisfactorily with 90.7% accuracy, showing that it is a suitable approach for capturing feature interactions and finding clinical features within tabular medical datasets.

The obtained best set of features, with the highest fitness are listed in Table VII. Both models gave 7 or 8 features out of 13. The difference between features selected by GA alone and GA with a Gini index (GA+Gini) depends on the fitness function. The genetic algorithm relies solely on accuracy, which may favor features that improve overall performance regardless of the class imbalances. Combining the genetic algorithm with a Gini index enhances both high accuracy and a balanced class distribution. Consequently, the genetic algorithm with a Gini index selects features that improve sensitivity and specificity, leading to better performance in medical diagnosis.

TABLE VI. SUMMARY OF THE MODELS PERFORMANCE WITH GA AND GINI INDEX

Model	Precision (mean ± std, 95% CI)	F1-score(mean ± std, 95% CI)	AUC(mean ± std, 95% CI)
MLP	0.837±0.027 (0.804-0.870)	0.855±0.018 (0.833–0.877)	0.916 ± 0.013 (0.899–0.933)
TabNet	0.911 ± 0.046 (0.854–0.968)	0.898 ± 0.050 (0.835–0.961)	0.957 ± 0.028 (0.922–0.992)
LR	0.833 ± 0.021 (0.806–0.860)	0.833 ± 0.021 (0.806–0.860)	0.833 ± 0.021 (0.806–0.860)
XGB	0.840 ± 0.037 (0.794–0.886)	0.840 ± 0.037 (0.794–0.886)	0.840 ± 0.037 (0.794–0.886)
RF	0.860 ± 0.011 (0.846–0.874)	0.860 ± 0.011 (0.846–0.874)	0.860 ± 0.011 (0.846–0.874)
SVM	0.904 ± 0.021 (0.879–0.929)	0.904 ± 0.021 (0.879–0.929)	0.904 ± 0.021 (0.879–0.929)

TABLE VII. FEATURE SELECTION TESTING RESULTS OF GA + GINI

Model	# Features	Name of selected feature
MLP	7	sex, cp, thalach, oldpeak, ca, thal, slope
TabNet	8	Age,cp, chol,, thalach, ca, thal, restecg, slope.
LR	8	sex, cp, trestbps, restecg, thalach, oldpeak, ca, thal
XGB	7	sex, cp, trestbps, exang, ca, thal, slope
RF	8	sex, cp, trestbps, exang, slope,ca, thal, oldpeak,
SVM	8	age, cp, restecg, exang, slope, ca, thal, thalach

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

The last stage of this work, greedy Search technique is used for hyper-parameter tuning with the goal of enhancing the models ability to forecast heart disease.

Before we present this experiment, we will give brief background about gready search optimizer, MLP and TabNet model.

DOI: **10.5281/zenodo.20345015**

a) Model Configuration and Hyper parameter tuning

A number of modifications have been introduced into the approach for hyper parameter tuning and performance evaluation of models. It involves using the nested stratified 5-fold cross-validation technique with the outer 5-folds for performing unbiased model evaluation and the inner 3-folds for tuning model parameters and feature selection. Specifically, a greedy search approach was used during the inner cross-validation to select the best hyper parameters for each considered model. In addition, care was taken to strictly separate the training, validation, and test data sets from each other in order to avoid any leakage from one fold to another. All preprocessing steps, such as feature scaling and SMOTE oversampling, were conducted solely on training folds prior to validation and testing of algorithms.

- *MLP Configuration*

The Multilayer Perceptron (MLP) classifier was configured by selecting the number of hidden layers, neurons, activation function, and regularization parameters. the hidden layer with 128 neurons and ReLU activation was used, the Adam optimizer with L2 regularization (alpha = 0.001) and early stopping to prevent over-fitting. To further optimizing performance, a greedy search strategy was using to tune the hyper-parameters of the classification algorithms.

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

The greedy search alliteratively selected parameter that improved accuracy. This process continued until no improvement was observed.

- *TabNet Configuration*

The TabNet classifier was configured using the PyTorch TabNet implementation with parameters, where Decision prediction dimension (nd) and Attention embedding dimension (na) were 32, as Number of decision steps (n_steps) was 4. Furthermore gamma was fixed 1.5 , the model was trained using the Adam optimizer with an initial learning rate of 0.01 ,a batch size of 256 and virtual_batch_size = 128 , ensuring sufficient optimization. where training was conducted for a maximum of 30 epochs, with early stopping applied using a patience of 5 epochs to mitigate over fitting. To enhance TabNet performance, a greedy search strategy was applied over the following parameters :

nd ∈ {16, 64} ,na ∈ {16, 64} ,n_steps ∈ {3, 5} .

The model evaluated all candidate values for a parameter at each iteration while keeping others fixed, selecting the value that maximized classification accuracy.

The initial parameter were selection based on commonly used configurations in the literatur e for tabular data classification tasks. For ensure a balance between computational efficiency and predictive performance, the greedy search was applied to adapt these parameters to the heart disease data-set.In addition to that all hyper-parameter tuning was applied on the training data only to avoid information leakage from the test set. This approach evaluates each parameter individually and selects the value that yields the most immediate improvement in performance, rather than relying on a potentially sub-optimal manual parameter selection process. Thus, it ensures an optimized parameter set that enhances model efficiency. After applying the algorithm, the optimal set of hyper-parameters that achieved the highest accuracy was 98.4% obtained. The results showed that, using the RF model, only eight features were optimal: sex, cp, trestbps, exang, slope, ca, thal and oldpeak.The best hyper-parameters after greedy search with RF:{,n_estimators: 100,max_depth: 1}. These parameters were selected as a result of the greedy Search strategy, which focuses on improving the cumulative performance of the model by gradually based on parameter impact on the outcomes.The results for Performance of algorithms using 5-fold nested cross-validation among the classifiers are shown in Table VIII the results showed that, out of 13 features, the best-fit features are only eight using RF classifier as in Table IX.

TABLE VIII. THE PERFORMANCE OF THE MODELS WITH GREEDY SEARCH

Model	Precision (mean ± std, 95% CI	F1-Score (mean ± std, 95% CI	AUC (mean ± std, 95% CI
MLP	0.951 ± 0.044 (0.896–1.000)	0.948 ± 0.031 (0.910–0.986))	0.989 ± 0.009 (0.978–1.000)
TabNet	0.947 ± 0.021 (0.920–0.974)	0.936 ± 0.018 (0.913–0.959)	0.982 ± 0.005 (0.976–0.988)
LR	0.834 ± 0.022 (0.807–0.861)	0.867 ± 0.008 (0.858–0.876))	0.911 ± 0.011 (0.898–0.924)
XGB	0.968 ± 0.033 (0.927–1.000)	0.971 ± 0.028 (0.936–1.000))	0.995 ± 0.006 (0.988–1.000)
RF	0.981 ± 0.032 (0.941–1.000)	0.985 ± 0.029 (0.949–1.000)	0.998 ± 0.005 (0.992–1.000)
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)

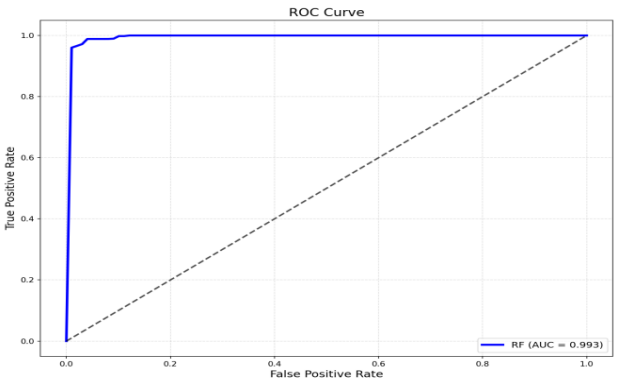


Fig 6..ROC curve of RF classifier

DOI: **10.5281/zenodo.20345015**

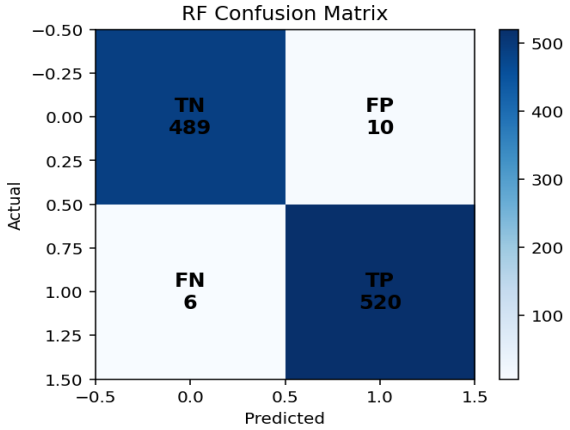


Fig 7. Confusion matrix of RF classifier

The Receiver Operating Characteristic (ROC) curve and the confusion matrix of the best classifier in this stage in shown in Fig. 6 and 7. It demonstrates the effectiveness and capability of the RF classifier in distinguishing performance between positive and negative heart disease cases.

TABLE IX. FEATURE SELECTION AFTER ADDING GREEDY SEARCH AND GA + GINI.

Model	# Features	Name of selected feature
MLP	7	sex, cp, thalach, oldpeak, ca, thal, slope
TabNet	8	age, cp, chol, thalach, ca, thal, restecg, slope
LR	8	sex, cp, trestbps, restecg, thalach, oldpeak, ca, tha
XGB	7	sex, cp, oldpeak, slope, ca, thal, trestbp
RF	8	sex, cp, trestbps, exang, slope, ca, tha', oldpeak
SVM	8	age, cp, restecg, exang, slope, ca, thal, thalach

The evaluation's findings showed that the prediction model that used RF as the classifier was highly accurate at forecasting heart disease. Among all algorithms, TabNet required the highest computational runtime ,approximately 4422.70 minutes due to its sequential attention and deep learning training procedure. AS the integration of nested cross-validation, GA-based feature selection, and greedy hyper-parameter optimization increased the computational cost of the model.

In contrast, Logistic Regression required the lowest runtime approximately 30.49 minutes because of its simple linear lower computational complexity. The model converged required fewer optimization steps compared to other algorithms.The reported 95% confidence intervals indicate the reliability of the proposed framework. where Narrow confidence intervals demonstrate that the performance metrics are consistent to data partitioning. while wider intervals indicator higher variability in model behavior. The proposed framework achieved narrow confidence intervals in the optimized RF after that TabNet and XGB models, confirming generalization capability of the proposed framework.

To ensure the statistical significance between different stages the hybrids framework, a paired t-test was conducted across all algorithms using 5-fold cross-validation results. The significance level was set to 0.05($p < 0.05$).Additionally, incorporating feature selection significantly improved performance in stage 2 compared to using all features in stage1. Although Gini optimization increased the classification performance, the performance increase achieved in the greedy optimization stage was greater as shown in Table X. This confirms that performance reliable and are not due to random variation.

TABLE X. PAIRED T-TEST BETWEEN STAGES ALL ALGORITHMS

Models	S1 vs S2 (p-value)	S2 vs S3 (p-value)	S3 vs S4 (p-value)	S1 vs S4 (p-value)
MLP	No(p=0.0748)	Yes (p = 0.0245)	Yes (p = 0.0062)	Yes (p = 0.0305)
TabNet	No (p = 0.2486)	Yes (p = 0.0372)	No (p = 0.1146)	No (p = 0.0715)
LR	Yes (p = 0.0412)	No (p = 0.5734)	No (p = 0.2056)	Yes (p = 0.0371)
XGB	No (p = 0.4259)	No (p = 0.0705)	Yes (p = 0.0013)	Yes (p = 0.0001)
RF	No (p = 0.6330)	No (p = 0.5543)	Yes (p = 0.0012)	Yes (p = 0.0049)
SVM	No (p = 0.0508)	No (p = 0.3046)	Yes (p = 0.0125)	Yes (p = 0.0358)

GA was used for feature selection, significantly improving both classifiers by identifying the most irrelevant features.

Table XI lists a comparison with previous researches.It is clear that this work shows that good improvement achieved by the Gini index, which demonstration its ability to evaluate the purity of class distributions. while accuracy based on measure overall prediction this may be misleading in imbalanced datasets, where a model can achieve high accuracy by favoring the majority class. The Gini index solve this limitation by penalizing impurity in class separation. By integrating (1 − Gini) into the fitness function, the proposed method encourages the selection of feature subsets that improve both classification performance and class balance. This leads to enhanced sensitivity and specificity, making the model more reliable for medical diagnosis.

Finally, greedy search was applied to optimize the parameters of each classifier, providing additional performance gains overall, the proposed complete hybrid model (genetic algorithm + gini index + optimization of hyper-parameters) exhibits higher accuracy, sensitivity, and specificity . RF slightly outperformed across all metrics, demonstrating its ability to feature interactions in medical datasets.

These results underscore the complementary effects of feature selection, consideration of class purity, and hyper-parameter optimization in enhancing classifier performance.

In conclusion Fig. 8 shows a sumup of selected features getted by all the experiments done in this paper. The most common features that are selected by all stages are: cp, exang, slope, ca and thal. The .feature sex is selected by 3 experiment .Finally, Fig. 8 summarizes the features that were obtained from all of the experiments in this work. All phasee choose the

DOI: **10.5281/zenodo.20345015**

following features: cp, exang, slope, ca, and thal. Three experiments choose the feature sex.

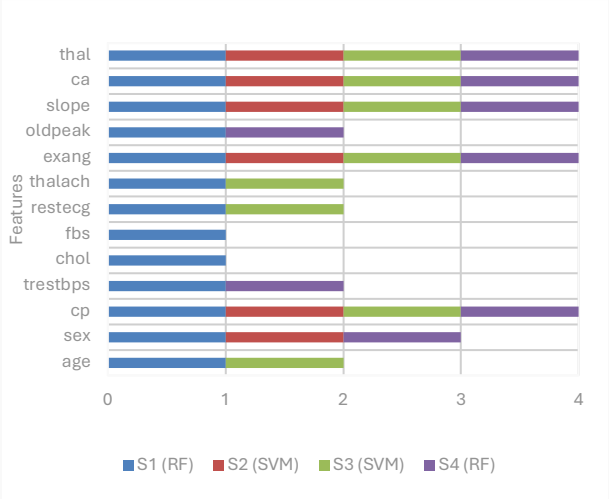


Fig 8. Selected features of all experiments

TABLE XI. COMPARISON WITH RELATED WORK.

Study	Method / Model	Feature Selection Technique	Dataset	Accuracy (%)	Computational Complexity
Escamila et al. [5]	Multiple ML Classifiers	CHI + PCA	Cleveland, Hungary, CH	89.7	High
Dissanayake et al. [6]	Decision Tree, RF, SVM, KNN	ANOVA, Chi-Square, MI, Lasso, Ridge	Cleveland	88.52	Medium
Murthy et al. [7]	RF-FSFC Framework	Sensitivity + Correlation Analysis	Cleveland	81.16	High
Ganesa Lingam et al. [8]	Decision Tree, LR-SVM	Forward and Backward Selection	Cleveland	88.51	Low
Jusia et al. [9]	KNN / Decision Tree	PSO-based Feature Selection	Cleveland	89.09	High
Fahad et al. [10]	MLP	Arithmetic Optimization Algorithm (AOA)	Cleveland	88.00	High
Manikandan et al. [11]	Logistic Regression, SVM	Boruta	Cleveland	88.52	Moderate
Sharma [12]	XGBoost	Wrapper-based Selection	Cleveland	73.74	High
Ahmad et al. [13]	Deep Learning Algorithms	Chi-Square, ANOVA, Mutual Information	Multi-center Clinical Dataset	89.2	Medium
Patel et al. [15]	Transformer-based Model	Self-Attention Feature Extraction	UCI Cleveland	88.5	High
Zhang & Chen [16]	CNN	Recursive Feature Elimination (RFE)	Combined Clinical Datasets	87.8	High
Gupta & Sharma [17]	Deep Learning Classifier	Genetic Feature Filtering	Benchmark Datasets	86.7	Medium
Grinsztajn et al. [23]	Transformer-based Foundation Model	In-context Learning	TabArena Benchmark (≤100K samples)	87.0	Medium
Cao et al. [26]	XGBoost, LightGBM, CatBoost	Boruta + RFE + Mutual Information + L1 Regularization	Standard Tabular Benchmarks	92.3	Medium
Proposed Method	**GA + Gini + Greedy + LR MLP RF SVM TabNet XGB**	**GA-based Multi-objective Feature Selection with Gini Optimization**	**Heart Disease Dataset (UCI/Kaggle)**	**85 94.6 98.4 96.8 93.6 97**	**Low High Medium High High High**

VI. CONCLUSION

The current research study designed a combined approach to develop a classifier for predicting heart diseases using GA-based feature selection, Gini index-based class balancing, and hyper parameters tuning. Different machine learning and deep learning models were used to test the developed framework, such as MLP, TabNet, XGBoost, Random Forest, Logistic Regression, and SVM. Experiments showed better results in comparison to those achieved by other machine learning approaches based on accuracy, precision, F1-Score, and AUC.

Out of all the models tested, RF was found to be the most effective. Using the Gini index as part of the GA's fitness function helped address issues related to class imbalance and bias, and using nested stratified cross-validation enhanced the reliability of the assessment process while preventing data leakage. There were still, however, several limitations.Despite the implementation of some techniques for dealing with over fitting, such as use of SMOTE within training folds, stringent splitting of data during preprocessing, early stopping, L2 regularization, and limited selection of features, external validation of the algorithms is required. In future, there will be efforts to validate the framework on a variety of clinical datasets and compare the developed methodology with other sophisticated frameworks like AutoML and Bayesian optimizations.

REFERENCES

[1] Nalluri, V. Saraswathi, R. Somula, K. Govinda, and E. Swetha, "Chronic heart disease prediction using data mining techniques," Advances in Intelligent Systems and Computing, vol. 1079, pp. 903–912, 2020, doi:10.1007/978-981-15-1097-7_76.

[2] R. Manji and P. S. Tappia, "Cost–effectiveness analysis of rheumatic heart disease prevention strategies," Expert Review of Pharmacoeconomics & Outcomes Research, vol. 13, no. 6, pp. 715–724, 2013, doi:10.1586/14737167.2013.847471.

[3] G. Sivapalan, T. Gasparyan, and K. Nundy, "ANNet: A lightweight neural network for ECG anomaly detection in IoT edge sensors," IEEE Transactions on Biomedical Circuits and Systems, vol. 16, no. 1, pp. 24–35, 2022, doi:10.1109/TBCAS.2021.3098538.

[4] S. Yeom, I. Giacomelli, and M. Fredrikson, "Privacy risk in machine learning: Analyzing the connection to overfitting," in Proc. IEEE 31st

DOI: **10.5281/zenodo.20345015**

Computer Security Foundations Symposium (CSF), pp. 268–282, 2018, doi:10.1109/CSF.2018.00029.

[5] J. Escamila, A. Karen, and A. El Hassani, “Classification models for heart disease prediction using feature selection and PCA,” Informatics in Medicine Unlocked, vol. 19, Art. no. 100335, pp. 1–10, 2020, doi:10.1016/j.imu.2020.100335.

[6] K. Dissanayake and J. Johar, “Comparative study on heart disease prediction using feature selection techniques on classification algorithms,” Applied Computational Intelligence and Soft Computing, vol. 2021, Art. ID 5581803, pp. 1–17, 2021, doi:10.1155/2021/5581803.

[7] S. Murthy, “A novel feature selection approach with integrated feature sensitivity and feature correlation for improved disease prediction,” Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 9, pp. 12005–12019, 2022, doi:10.1007/s12652-022-03750-y.

[8] V. Ganesalingam, S. Suthakaran, and V. Ravi, “Enhancing prediction of heart failure using feature selection techniques,” in Proc. 14th ASBIRES, pp. 234–242, 2022.

[9] [9] A. Jusia, A. Rahim, and H. Yani, “Improving performance of KNN and C4.5 using particle swarm optimization in classification of heart diseases,” Jurnal RESTI, vol. 8, no. 3, pp. 640–649, 2024, doi:10.29207/resti.v8i3.2587.

[10] [10] F. Alghamdi, H. Al-Manaseer, and G. Jaradat, “Multilayer perceptron neural network with arithmetic optimization algorithm-based feature selection for cardiovascular disease prediction,” Machine Learning and Knowledge Extraction, vol. 6, no.2,pp.987–1008,2024,doi:10.3390/make6020056.

[11] G. Manikandan, B. Pragadeesh, and V. Manojkumar, “Classification models combined with Boruta feature selection for heart disease prediction,” Informatics in Medicine Unlocked, vol. 44, Art. no. 101446, 2024, doi:10.1016/j.imu.2024.101446.

[12] V. Sharma, S. Yadav, and M. Gupta, “Heart disease prediction using machine learning techniques,” in Proc. ICACCCN, pp. 1–6, 2024.

[13] B. Ahmad, J. Chen, and H. Chen, “Feature selection strategies for optimized heart disease diagnosis using ML and DL models,” arXiv preprint, arXiv:2501.12345, 2025.

[14] [14] M. Ahmed, F. Li, and Z. Wang, “Integrated ensemble deep learning with feature filtering for cardiovascular risk assessment,”Journal of Medical Systems, vol. 49, no. 3, Art. no. 22, pp. 1–12, 2026.

[15] [15] S. Patel, A. Kumar, and D. Singh, “Heart disease prediction using transformer-based feature extraction and classification,” IEEE Access, vol. 12, pp. 50273–50284, 2024.

[16] Y. Zhang and Y. Chen, “Deep feature optimization for coronary heart disease diagnosis using CNN and feature selection,” Applied Soft Computing, vol. 115. no. 109245, 2025.

[17] [R. Gupta and R. Sharma, “Optimized deep learning-based feature selection for heart disease prediction,” in Proc. Intelligent Computing Conference, pp. 115–126, 2026.

[18] G. Borisov, I. Seßler, B. Leemann, and G. Kasneci, “Deep neural networks and tabular data: A survey,” IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 4, pp. 1–21, 2023, doi:10.1109/TNNLS.2022.3181373.

[19] S. Ö. Arik and T. Pfister, “TabNet: Attentive interpretable tabular learning,” in Proc. AAAI Conference on Artificial Intelligence, vol. 35, no. 8, pp. 6679–6687, 2021, doi:10.1609/aaai.v35i8.16826.

[20] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, “Revisiting deep learning models for tabular data,” Advances in Neural Information Processing Systems, vol. 34, pp. 18932–18943, 2021.

[21] N. Mohammed and P. Thiyagarajan, “Feature selection using efficient fusion of Fisher score and greedy searching for Alzheimer’s classification,” Journal of King Saud University – Computer and Information Sciences, vol. 34, no. 9, pp. 4993–5006, 2022, doi:10.1016/j.jksuci.2021.04.008.

[22] A. Elisseeff and I. Guyon, “An introduction to variable and feature selection,” Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.

[23] [23] R. Kazerani, “Improving breast cancer diagnosis accuracy by particle swarm optimization feature selection,” International Journal of Computational Intelligence Systems, vol. 17, no. 1, Art. no. 44, 2024, doi:10.1007/s44196-024-00428-5.

[24] [24] S. Benghazouani, S. Nouh, and A. Zakrani, “Enhancing breast cancer diagnosis: A comparative analysis of feature selection techniques,” IAES International Journal of Artificial Intelligence, vol. 13, no. 4, pp. 4312–4322, 2024.

[25] S. Silaich and R. Yadav, “Breast cancer diagnosis using machine learning and particle swarm optimization,” Technical Journal, vol. 19, no. 4, pp. 256–263, 2025.

[26] K. Cao et al., “Prediction of cardiovascular disease based on multiple feature selection and improved OPS-XGBoost model,” Scientific Reports, vol. 15, 2025.

[27] H. Wang et al., “Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods,” Journal of Big Data, vol. 11, Art. no. , 2024.

[28] I. Goodfellow, Y. Bengio, and A. Courville,“ Deep Learning”. Cambridge, MA: MIT Press, 2016.

[29] V. Attari and R. Arroyave, “Decoding non-linearity and complexity: deep tabular learning approaches for materials science,” Digital Discovery, vol. 4, no. 10, pp. 2765–2780, 2025, doi:10.1039/D5DD00166H.

[30] R. Shwartz-Ziv and A. Armon, “Tabular data: Deep learning is not all you need,” Information Fusion, vol. 81, pp. 84–90, 2022.

[31] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Belmont, CA: Wadsworth, 1984.

[32] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York: Springer, 2009.

[33] J. Smith, “Heart Disease Dataset,” Kaggle.[Online].Available:https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

DOI: **10.5281/zenodo.20345015**