



OPEN

Empirical exploration of whale optimisation algorithm for heart disease prediction

Stephen Akatore Atimbire, Justice Kwame Appati[✉] & Ebenezer Owusu

Heart Diseases have the highest mortality worldwide, necessitating precise predictive models for early risk assessment. Much existing research has focused on improving model accuracy with single datasets, often neglecting the need for comprehensive evaluation metrics and utilization of different datasets in the same domain (heart disease). This research introduces a heart disease risk prediction approach by harnessing the whale optimization algorithm (WOA) for feature selection and implementing a comprehensive evaluation framework. The study leverages five distinct datasets, including the combined dataset comprising the Cleveland, Long Beach VA, Switzerland, and Hungarian heart disease datasets. The others are the Z-AlizadehSani, Framingham, South African, and Cleveland heart datasets. The WOA-guided feature selection identifies optimal features, subsequently integrated into ten classification models. Comprehensive model evaluation reveals significant improvements across critical performance metrics, including accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve. These enhancements consistently outperform state-of-the-art methods using the same dataset, validating the effectiveness of our methodology. The comprehensive evaluation framework provides a robust assessment of the model's adaptability, underscoring the WOA's effectiveness in identifying optimal features in multiple datasets in the same domain.

Keywords Heart disease, Whale optimization, Feature selection, Accuracy, Swarm intelligence

Abbreviations

WOA	Whale optimization algorithm
FS	Feature selection
HD	Heart disease
ML	Machine learning
DL	Deep learning
KNN	K-nearest neighbor
RF	Random forest
DT	Decision tree
RNN	Recurrent neural network
LSTM	Long short-term memory
XGBoost	Extreme gradient boosting
NB	Naïve Bayes
SVM	Support vector machine
MLP	Multilayer perceptron
CSRS	Cuckoo search with the rough set

Heart Disease (HD) is of utmost importance due to the heart's critical role among other human organs. HD has high death rates worldwide, with approximately 17.9 million people dying from heart conditions in 2019¹. Heart diseases account for 32% of global deaths, with heart attacks and stroke alone making more than 85% of recorded deaths. Over 75% of cardiovascular deaths in 2019 occurred in underdeveloped nations, accounting for 38% of deaths under 70 years¹. Since cardiovascular diseases are fatal, their early detection will enable medical professionals to provide timely healthcare to patients to avert death.

Because of a scarcity of ultra-modern examination tools and medical experts, conventional medical methods for diagnosing heart diseases are challenging, complicated, time-consuming, and exorbitant, making the

Department of Computer Science, University of Ghana, Accra, Ghana. ✉email: jkappati@ug.edu.gh

diagnosis of heart diseases difficult and sometimes unavailable, especially in developing countries². Machine and deep learning methods have been recently used to analyze clinical data and make predictions³.

Machine learning (ML) provides cost-efficient alternatives where already collected patient data serve as a data mine to perform predictive analysis for diagnostic purposes. To improve the accuracy of ML models, some existing works have focused on using various classifiers or their enhanced forms^{4–7}. Related works confirm that the feature selection reduces data dimensionality and improves model performance significantly⁸. Hence, some studies have utilized various methods to improve performance by varying the feature selection methods^{9,10}.

However, some works that utilize feature selection are fraught with redundant features that impact metrics recorded. This is affirmed when wrapper methods are used over filter methods and when embedded methods are used over filter and wrapper methods. It also explains why works, including feature selection, may only record better performance on some datasets if the technique is efficient. In addition, though the researchers do not present the reason some existing works have not reported on specific metrics, studies such as Hicks et al.¹¹ have posited that in a clinical setting, a subset of metrics may give an erroneous outlook of how a model performs and not enabling holistic model performance evaluation. There is an avenue for more scientific work on feature selection methods capable of improving other metrics besides the accuracy metric. This helps to affirm the reliability of the model performance as the unavailability of multiple evaluation metrics is an indication of an unbalanced model not capable of being thoroughly assessed.

This study proposes the use of the whale optimization algorithm (WOA) as a swarm-inspired feature selection algorithm on five (5) heart datasets on ten (10) models (classical ML, ensemble and deep learning models) for the selection of relevant datasets features. The approach contributes to the body of knowledge in the heart disease domain by providing a comprehensive assessment of five different datasets (in the same domain), ten different models and five evaluation metrics. The proposed methodology also validates the robustness of the WOA algorithm on five datasets of variable sizes in the same domain compared to most works, which do not test their methodologies on multiple datasets in the same domain.

Related works

Introduction

Identifying significant features (or redundant features) in a dataset remains a critical activity in modelling¹². Feature selection helps to determine relevant subsets from the existing features¹³ and is recommended for use where one needs to understand the selected features¹⁴. Excessive dataset features cause over-fitting, reduce model efficiency, and impair model generalization¹⁵. Feature selection (FS) improves the modelling process¹⁶. Selected feature selection techniques in literature are mostly grouped into filter, wrapper, hybrid, embedded, and recently, swarm intelligent methods.

Filter-based feature selection methods

Filter-based methods do not consider the dependence of features on one another; only the intrinsic properties are considered¹⁷. Ghosh et al.¹⁸ used the Cleveland heart dataset alongside the Decision Tree, K-Nearest Neighbour (KNN), and Random Forest (RF) classifiers, recording the highest accuracy of 93.36% with RF. The most optimal features were selected using relief before being passed to the classifiers for training and testing. Relief selected six of the 13 features, resulting in a significant achievement of accuracy. Narsimhulu et al.¹⁹ proposed a Filter Based Feature Selection (FBFS) to detect relevant features and remove the redundant ones. RF recorded the highest accuracy of 95.08%, validating the positive influence of the filter feature selection method on model accuracy. Filter methods can compute the scoring function quickly and efficiently²⁰. Filter methods do not consider feature interdependence and are not dependent on the classifiers²¹.

Wrapper based feature selection methods

To resolve some of the shortcomings of filter-based methods, the wrapper methods use classifiers and consider feature interdependence, increasing computational time²². Evaluating different feature subsets requires retraining and testing²³ compared to the filter feature selection algorithms. El-Sayed employed the Genetic Algorithm (GA) wrapper method on the Cleveland heart disease dataset, obtaining 89.07% and 67.22% for binary and multiclass, respectively. First, GA was used to reduce the attributes from the dataset, and Linear Discriminant Analysis (LDA) was used for classification. MultiLDA and multi-classifiers were then used again after GA for multiclass datasets. In their study, GA used alongside the LDA classifier outperformed KNN, Support Vector Machine (SVM) and Naïve Bayes (NB). Sequential Backward Selection (SBS) is also applied to ascertain more significant features, increasing classifier accuracy while decreasing computational time in the study of Haq et al.¹⁰. The Cleveland heart disease database was used in the study, recording an accuracy of 90% with six features selected using SBS for feature selection. The KNN classifier with the SBS method outperforms the KNN classifier used single-handedly. The Recursive Feature Elimination (RFE), another wrapper method, is used in the study of²⁴. Feature selection is performed using the SVM-RFE before using KNN to find the best features and reduce computing time. KNN without FS had an accuracy of 82.65% and 86.33% with feature selection. The Weighted KNN without FS recorded 83.45% and 90.88% with feature selection. The results confirm that with feature selection, the accuracy of the KNN or the weighted KNN experiences a significant improvement. Although wrapper methods perform better than filter methods, they are computationally expensive²³.

Embedded based feature selection methods

Embedded methods search for classifier-specific optimal features and keep track of feature interdependence with less complexity than the wrapper method^{25,26}. Embedded methods are less prone to overfitting than wrapper methods and are computationally costlier than the filter method²⁷. For the prediction of HD, the LASSO feature

selection was compared to the relief feature selection method by Ghosh et al.²⁸ by using the Cleveland, Long Beach VA, Switzerland, Hungarian and Statlog datasets. The study achieved the best accuracy of 99.05% with LASSO and the Random Forest Bagging classifier. In a related study, Zhang et al.²³ proposed the LinearSVC algorithm, an embedded method with a Deep Neural Network using the Cleveland UCI dataset. The work resulted in an accuracy of 98.56%. The study confirms the advantage of embedded methods of feature selection over the former techniques. The embedded feature approach mitigates the shortcomings of the filter and wrapper methods by engaging with the classifier and accounting for feature dependencies²⁶. Its drawback, however, is its slow performance²⁹.

Hybrid based feature selection methods

The hybrid methods use two or more methods together for feature selection tasks. The study of³⁰ hybridized Cuckoo Search with the rough set to form the Cuckoo Search with the Rough Set (CSRS) model in their paper. Data from 603 patients are divided into train data for 332 and test data for 271. Cuckoo Search is applied to determine the most optimal features from the training data. Eight features with highly significant values are chosen after 1000 iterations. With optimal features selected, CSRS obtained 93.7% accuracy, outperforming other Cuckoo Search integrated models by 3.37%. Also, a three-phase feature selection approach is proposed by³¹. In their paper, a three-way feature selection method is proposed for reducing the feature set of the arrhythmia dataset. The three cancer datasets achieved 100% accuracy, and 94.50% was achieved on the arrhythmia dataset. The method selects features with the best accuracy not dependent on the filter methods and the classifiers in the first phase, fusing four techniques: Mutual Information (MI), ReliefF (RFF), Chi-Square (CS), and Xvariance (XV) using three classifiers, KNN, SVM, and NB. The original data set is then subjected to the XGBoost | algorithm, with the top features chosen based on accuracy. In the second phase, the top features in the previous stage are correlated using the Pearson Correlation Coefficient (PCC), and strongly correlated features are dropped, ensuring that the feature subsets have maximum relevance. Again, XGBoost is used to obtain the top features, which are then sent to Phase 3. The optimal feature set is finalized in the third phase, using WOA. Arrhythmia, leukaemia and two others form part of the four datasets used in the trials. The features with the most negligible significance in the dataset had the lowest rankings in information, reliance, and distance. It is also observed that the selected features were reduced by 150 and 1286 times for the arrhythmia and leukemia datasets, respectively. The proposed method enhances the removal of noisy data which are discarded and improves accuracy.

Aside from the authors' report of limited data, one of the tri-stage's flaws is the costly computation required to obtain accuracy during the first phase and when WOA is applied.

Arroyo and Delima³² propose a genetically optimized neural network for HD risk prediction. The Cardiovascular disease dataset with 70,000 records and 12 features was employed for ANN modelling. The authors' work resulted in higher accuracy. However, determining the correct number of layers and neurons takes time and effort. The study of³³ also presented techniques based on the KNN, SVM, NB, RF, and a Multilayer Perceptron (MLP) optimized by Particle Swarm Optimization (PSO) merged with Ant Colony Optimization (ACO). The study achieved a maximum of 99.65% accuracy, and the research further enhances the position that with an optimal result set, the accuracy of a study can be significantly improved³⁴. Proposed a hybrid feature selection through Information Gain, Correlation, Chi-Squared, and Relief-F. Using the KNN classifier on a heart failure dataset. The approach recorded 84.61% accuracy. The computational and efficiency challenges inherent in a particular algorithm can be incorporated and reflected in the hybrid algorithm, making the hybrid method very complicated to implement. Therefore, individual algorithms' challenges must be intentionally handled to improve efficiency and accuracy in hybrid methods.

Swarm intelligent based feature selection methods

Swarm intelligence optimization methods have emerged as robust methods for feature selection in many fields. "Swarm Intelligence (SI) is a type of artificial intelligence that is based on collective behaviours in decentralized and self-organized systems"³⁵. SI is usually inspired by an organized pattern of a random behaviour of a population known as the agents. The intelligent behaviour results in a recognizable pattern helpful in solving optimization problems. Some standard swarm algorithms include the Cuckoo Optimization Algorithm (COA), Ant Colony Optimization (ACO), Bat Algorithm (BA), Grey Wolf Optimization (GWO), Salp Swarm Algorithm (SSA), Marine Predator (MP) and whale optimization algorithm (WOA).

Usman et al.³⁶ propose the Cuckoo Search Algorithm (CSA) and another variant, the Cuckoo Optimization Algorithm (COA) for FS on the Eric, Hungarian, Stat log, and Z-Alizadeh datasets. Four classifiers are used in their study, namely Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Naïve Bayes (NB), and Random Forest (RF) classifiers. CSA outperformed COA by selecting fewer features and higher accuracy across all datasets. Other researchers, Al-Tashi et al.³⁷, proposed establishing optimal features to diagnose coronary artery disease using the Grey Wolf Optimization (GWO). The authors first use GWO to determine the significant features. The SVM is then utilized with the optimally selected features as input. The Cleveland UCI dataset was employed, and the model recorded 89.83% accuracy. The other evaluation metrics used are sensitivity, which was recorded at 93%, and 91% for specificity rates. In the study of Al-Tashi et al.³⁷, Grey Wolf Optimization (GWO) finds optimal features in the dataset and then evaluates the fitness function of GWO using SVM. The Cleveland UCI dataset used with the study recorded 89.83%, 93% and 91% in accuracy, sensitivity, and specificity, respectively. By combining GWO and the Naïve Bayes (NB) classifier, a new strategy for detecting cardiac disease is developed³⁸. The features of the heart disease dataset are discretized to increase the accuracy of the classifiers. GWO then automatically selects Naïve Bayes's weights to maximize NB's performance, achieving 87.45% accuracy with good values for Sensitivity, F-measure, and G-mean. The Cleveland UCI data was used in the project. The proposed (GWO-NB) method performed better than the standard Naïve Bayes classifier in

accuracy from the experiments. It also confirms the effectiveness of GWO on classifiers. To improve the GWO, the paper of Chakraborty et al.³⁹ proposed an enhanced form of GWO for feature selection. Cleveland, Long Beach VA, and Switzerland were utilized in their study. Others include the Hungarian and Statlog datasets. Bagging and boosting techniques produce hybrid classifiers with NB, RF, Decision Tree (DT), K-Nearest Neighbor (KNN), Neural Network, Gradient Boosting, and Adaptive Boosting (AdaBoost). RF achieved the best accuracy of 99.26% with Enhanced-GWO, an improvement of accuracy of 11.90% over the conventional model. The Whale Swarm Algorithm (WSA) proposed by David⁴⁰ was used to select significant features to determine the presence of cardiovascular disease. Using the Statlog dataset, the study recorded an average selection of 6 features after 100 iterations. Using the Logistic Regression (LR), Random Forest, Support Vector Machine and Gaussian Naive Bayes (GNB) classifiers, the study noticed that the Random Forest outperformed the other classifiers, reporting an accuracy of 85.7%. Another commonly used swarm intelligence algorithm within the heart disease domain is Particle Swarm Optimization (PSO). Shahid and Singh⁴¹ in their paper propose a model known as the emotional neural networks (EmNNs) that PSO has hybridized on the Z-Alizadeh Sani dataset. The researchers apply four distinct feature selection algorithms to optimize the performance of the suggested model. The proposed method selected a total of 22 features. The model outperforms all other models used in the study, recording the highest averages for accuracy, precision, sensitivity, specificity, and F1 score at cross-validation, i.e., 88.34%, 92.37%, 91.85%, 78.98%, and 92.12%, respectively, which are comparable to existing methods in the literature. Their work is improved by introducing a novel multi-objective PSO (MOPSO) proposed by Asadi et al.⁴². Using the Statlog, Cleveland, SPECT, SPECTF, VA Long Beach, and Eric datasets, the study recorded improved accuracy better than when feature selection was not used. Wankhede et al.⁴³ propose a Decision Function-based Chaotic Salp Swarm (DFCSS) method to determine the most significant features after data pre-processing. The relevant attributes are then provided to an enhanced Elman neural network classifier. The experiment demonstrates that the proposed method outperformed existing methods with 98.7% and 98.0% accuracy for CVD and UCI datasets, respectively. The Salp Swarm Algorithm selects valuable features from the UCI Cleveland and heart-failure-clinical-records datasets in Sureja et al.⁴⁴. The study recorded 98.75% and 98.46% accuracy for the respective datasets. Swarm-based feature selection approaches thus have the potential to enhance classifier performance drastically.

This research explored the use of WOA for feature selection for the following reasons. First, WOA is easy to implement due to the few internal numbers of parameters⁴⁵, outperforms other swarm algorithms such as the Particle Swarm Optimization (PSO) and gravitational search algorithm (GSA) as recorded in the study of Mirjalili and Lewis⁴⁶ when used for engineering design problems. It has provided outstanding results for optimization problems in domains such as wireless resource allocation and gold price forecasting (outperforming PSO, Grey Wolf Optimization and genetic algorithm^{47,48}). The study of Ay et al.⁴⁹ also affirms the creditable performance of WOA compared with cuckoo search (CS), flower pollination algorithm (FPA), and Harris Hawks Optimization (HHO) algorithms, other metaheuristic algorithms. Also, WOA is not widely explored in the heart disease domain; hence, its creditable performance in different fields, simplicity, and excellent output against other metaheuristic methods make it a choice worth selecting for FS in the HD domain that guarantees good results.

Proposed methodology

The methodology for this work and the steps used for feature selection, model training, and model evaluation are discussed. Figure 1 depicts the proposed methods employed for this work.

Datasets

The datasets leveraged in this study are summarized in Table 1. The activities performed on the datasets are discussed in the data pre-processing subsection following.

Dataset visualization before feature selection

For each of the five datasets, four visualization techniques were employed to gain insights into the data's structure prior to feature selection.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is utilized to identify and visualize outliers, to aid the understanding of the data's distribution and identifying anomalous observations. Correlation Heatmaps was utilized to examine the relationships between features while.

Boxplots was used for to visualize feature distribution. Histograms and density plots give insight on the target variable's distribution across different classes or values.

The dataset visualizations reveal in the DBSCAN plot a lack of clear, separable clusters, or it could be indicating a need to adjust the algorithm's parameters for a more meaningful clustering. The correlation heatmap provides insights into potential multicollinearity, which could influence feature selection and model performance. The boxplots on the other hand reveal the presence of outliers and the spread of the data. The target distribution plot shows a fairly balanced dataset. A general observation on the datasets reveal similar observations underpinning the need for the preprocessing tasks and feature selection. The visualizations for the datasets are shown in Figs. 2, 3, 4, 5, 6, 7 and 8 representing the combined, Z-AlizadehSani, Framingham, South African and the Cleveland UCI datasets respectively.

Data pre-processing

Data pre-processing is an essential activity in most ML pipelines. It includes tasks such as data cleaning, transformation, and organizing data before passing it to a model. This step is essential to improve data quality and make it suitable for building accurate and efficient models.

In this work, the target values indicating the existence of HD are coded as 0 or 1 (absent and present, respectively). Other categorical data fields, such as "famhist" in the SA heart dataset with values "absent" and "present",

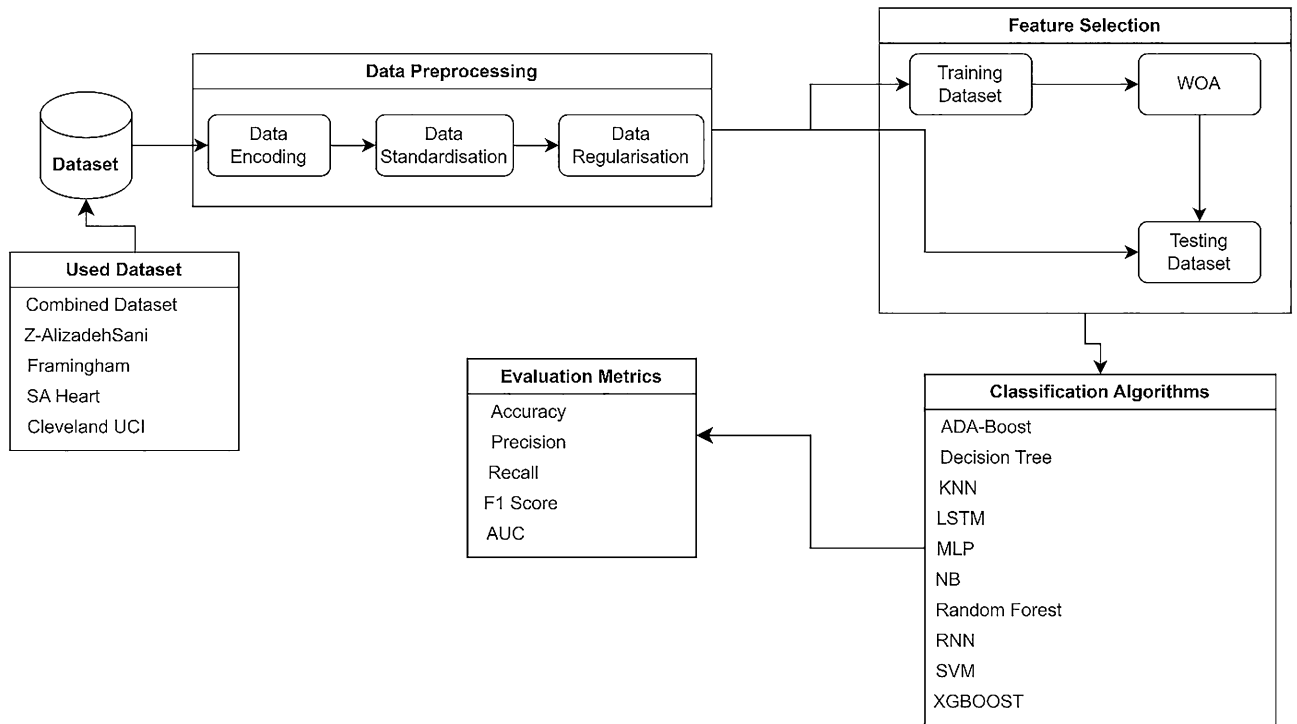


Figure 1. Methodology for the study.

#	Name of dataset	No. of observations	No. of features	Publicly available?
1	Combined dataset	1025	13	Yes
2	Z-AlizadehSani dataset	303	54	Yes
3	Framingham dataset	4133	15	Yes
4	South African heart dataset	462	9	Yes
5	Cleveland UCI heart dataset	303	13	Yes

Table 1. Datasets utilized in the study.

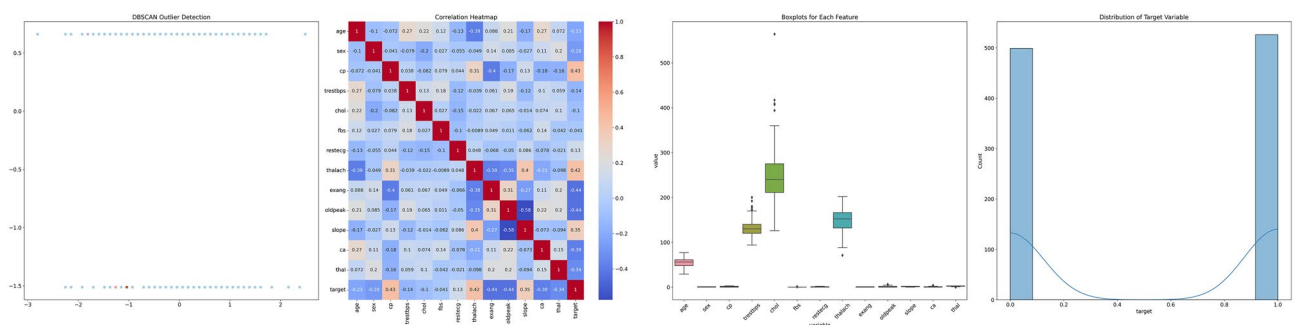


Figure 2. Combined dataset visualization before feature selection.

are coded as 0 for absent and 1 for present. “Male” and “Female” values for all datasets are encoded as 1 and 0, respectively. All other fields with similar values are coded with numerical values.

Data standardization and regularization

The standardization process utilizing the StandardScaler was employed. This involved adjusting the features present in the data to be transformed to possess a zero (0) mean and a one (1) standard deviation. Standardizing the features with StandardScaler has a lower error than MinMaxScaler because StandardScaler scales each feature with zero mean and unit variance, scaling them in an equivalent bell curve and handles missing values

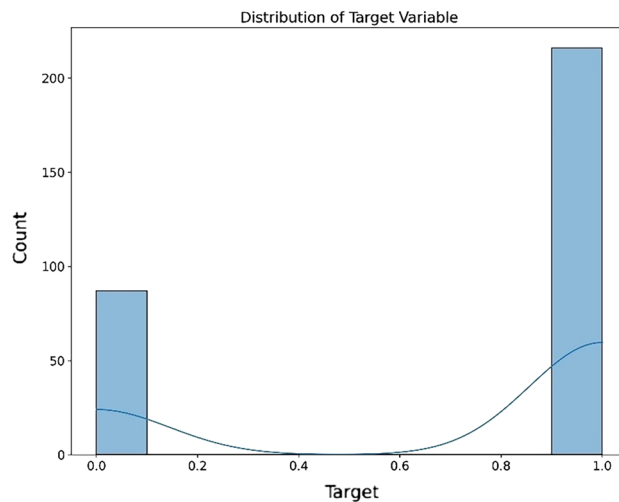


Figure 3. Z-AlizadehSani dataset target visualization before feature selection.

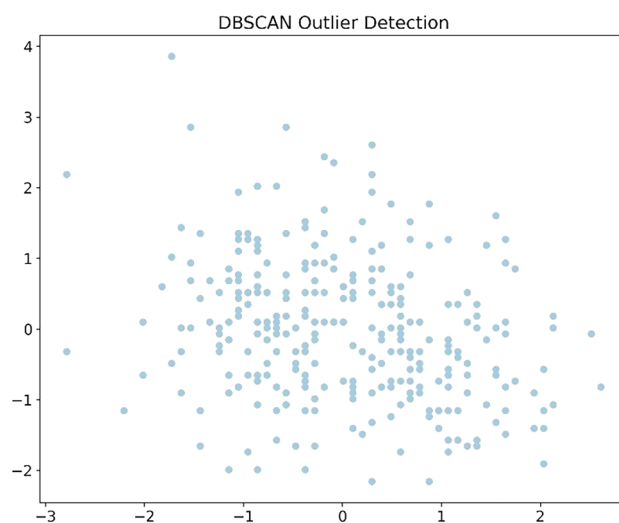


Figure 4. Z-AlizadehSani dataset DBSCAN outlier detection visualization before feature selection.

preventing overfitting^{50,51}. A Logistic Regression model with an L2 regularization penalty is fitted with the standardized data. Another parameter, “C”, which controls the trade-off between fitting the data well and the simplicity (parsimonious) of the model, is utilized in the logistic regression model. The Logistic Regression classifier fits the model to the scaled train data and the corresponding target. The model’s coefficients, which indicate each feature’s significance in making predictions, are stored in a variable applied to the training data to obtain a new, regularized training dataset. This regularized training data is utilized for model training. Afterwards, the stored coefficient is applied to the test data to regularize the test data and target. Logistic regression with L2 regularization has proven useful in prior studies to promote sparse solutions^{52,53}. The regularized train data is passed to the whale optimization algorithm to select the optimal features for model training modelling and testing. The WOA algorithm is discussed in the adjoining pages.

The whale optimization algorithm (WOA)

WOA is an optimization algorithm that draws inspiration from nature and imitates humpback whales’ hunting techniques developed by Seyedali Mirjalili⁴⁶. The main components of whale optimization are position updating, encircling prey, and searching for prey. The entire WOA steps are discussed in this section and modified to be utilized for feature selection, which is further explained.

Encircling prey

Encircling uses the first current solution as the target or near to it since the optimal design’s location is initially unknown. The remaining agents then update their locations guided by the location of the best agent. This behaviour is modelled by Eqs. (1) and (2).

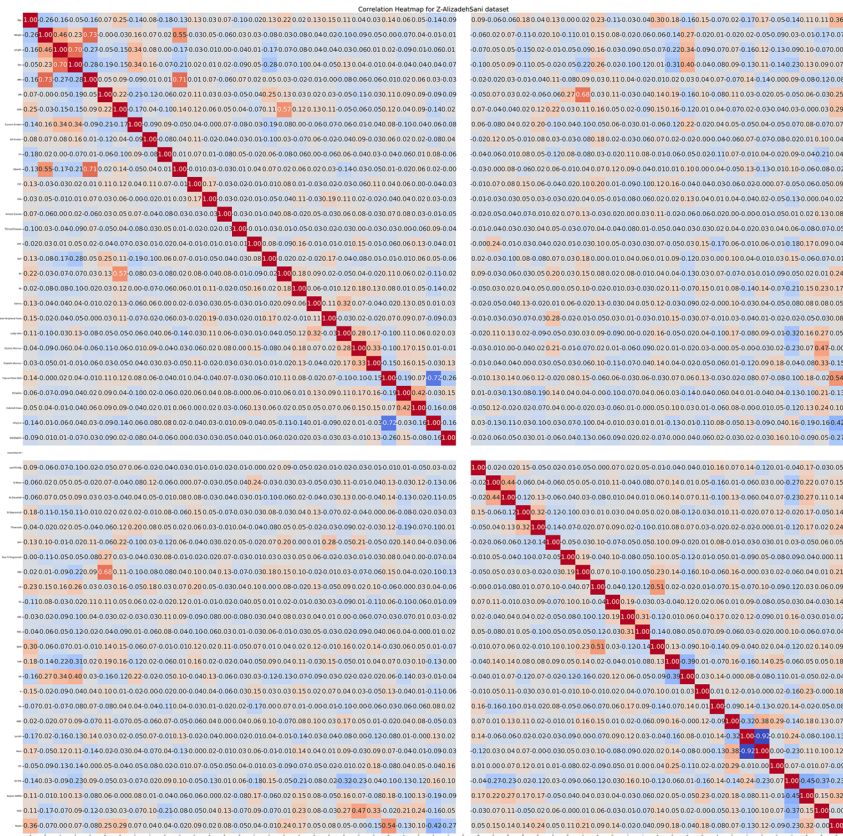


Figure 5. Z-AlizadehSani dataset correlation Heatmap visualization before feature selection.

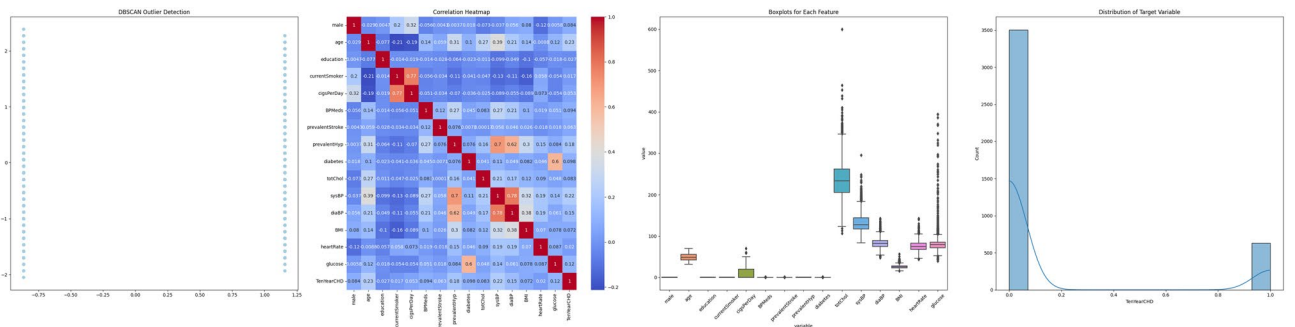


Figure 6. Framingham dataset visualization before feature selection.

$$\vec{D} = \left| \vec{C} \cdot \vec{X}^*(t) - \vec{X}(t) \right| \tag{1}$$

$$\vec{X}(t + 1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \tag{2}$$

where t is the current iteration, the coefficients \vec{A} and \vec{C} being vectors and X^* being the absolute position vector for the best solution of the iteration. \vec{A} and \vec{C} are determined using Eqs. (3) and (4).

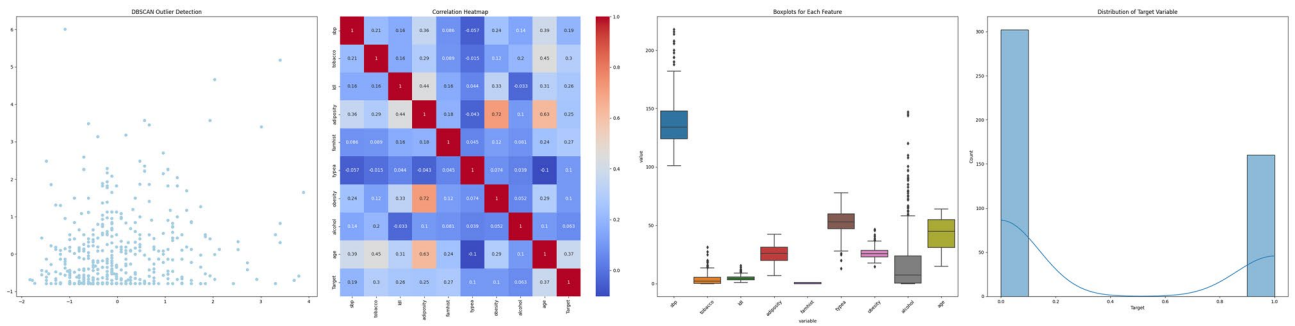


Figure 7. South African heart dataset visualization before feature selection.

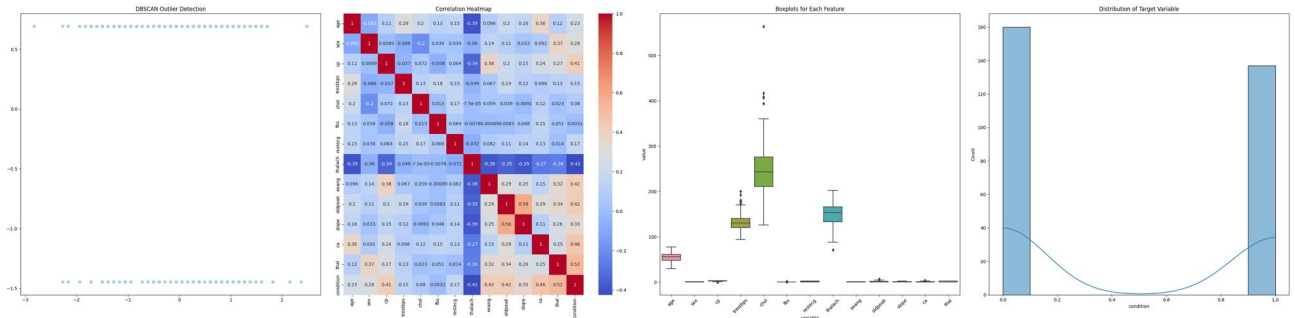


Figure 8. Cleveland UCI dataset visualization before feature selection.

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \tag{3}$$

$$\vec{C} = 2 \cdot \vec{r} \tag{4}$$

where \vec{a} linearly reduces from 2 to 0 and \vec{r} , a randomly generated vector with a value within 0, 1 is generated.

Bubble-net attacking method (exploitation phase)

Bubble-net attack (exploitation) uses encircling and spiral updating techniques to control the whale mechanisms in WOA. The shrinking encircling mechanism is governed by Eqs. (2) and (3), with the varying range of vector A relying on vector a. Equation (3) can compute this range. The vector a diminishes from 2 to 0 across numerous rounds, affecting vector A. When vector A lies within [-1, 1], the agent's following location is between the present and target positions. The spiral-based position update procedure begins by computing distances between the whale's present position and the positions of the targets. Based on this range, the spiral motion mimicking the humpback whale's swimming pattern is generated. The modelled pattern is explained in Eq. (5).

$$\vec{X}(t+1) = \vec{D}^{\vec{l}} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \tag{5}$$

The logarithmic spiral is defined by the constant b, and the value of l is chosen at random from [-1, 1]. There is a 50% likelihood of using shrinking or the spiral technique to attack the prey successfully. Equation (6) defines the equation.

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}^{\vec{l}} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \tag{6}$$

p is a number between 0 and 1.

Search for prey

The last process to be discussed in the WOA process is how the target is searched. Humpback whales disperse and explore the search space at random to identify the target, as Eq. (3) describes. When the value of vector A exceeds 1 or falls below -1, it prompts the whales to spread out and perform a random search. The goal of this phase is to integrate exploratory abilities into the WOA. During this stage, the following location of the agents is determined at random, irrespective of the existing best solution's value. Equations (7) and (8) explain the search process further.

$$\bar{D} = |\bar{C} \cdot \bar{X}_r - \bar{X}| \quad (7)$$

$$\bar{X}(t + 1) = \bar{X}_r - \bar{A} \cdot \bar{D} \quad (8)$$

Vector (X_r) is randomly positioned and drawn from the selected population. WOA begins with randomly generated solutions and iteratively refines the optimal solution using either random search agents or the best-performing one. The technique is based on three key parameters: vector a , vector A , and vector p . Vector a , which diminishes through 2 to 0, is utilized to maintain a suitable equilibrium between exploration and exploitation. Vector A determines whether to use a random or the best search agent for position updating. Suppose vector A is more significant than one (1); a randomly generated search agent is used to effect an update on the position. However, the current best solution is used if it is less than one. This assists the WOA in maintaining the proper equilibrium between exploring and exploiting. p facilitates the search agents to vary between cyclic and spiral movement, increasing their adaptability.

Feature selection using WOA

The proposed modified WOA determines the best features using the whale optimization algorithm with input from the preprocessed data. The initial location of the whales (indicating feature selection or non-selection) is generated randomly as binary values. The algorithm then updates each whale's position in each iteration, shrinking the search space with each iteration. The new position is determined by combining the best position found thus far with random numbers. If the new position has better fitness, it becomes the best. The algorithm then returns the indices of the selected features, considered the best feature. The returned indices are passed to the next step to aid the prediction process. Each whale represents a potential solution in the context of feature selection, with a binary vector encoding the presence or lack of features. The algorithm updates each whale's location based on a linearly decreasing search space coefficient and the best position over a predetermined number of iterations (10, 20, 30, 40, 50 to 100) and agents (10, 20, 30, 40, 50 to 100).

Ten agents are paired against ten iterations for the WOA. This is repeated in additions of 10 (for agents and iterations). By training a Logistic Regression classifier penalized by an L1 regularization on the chosen feature sets, the fitness function is utilized to compute the performance of the provided subset of features. Each iteration updates the whale position. If the new position improves fitness, the position and fitness are updated accordingly. The maximized prediction performance on the validation with a minimized number of features is considered the optimal feature for that iteration. The best position, therefore, reflects the ideal subset of features so far as the algorithm has run through the required number of iterations. The Scaled training data and its target, the number of whales and iterations are the input factors used by WOA for feature selection. The entire process is captured in Algorithm 1.

Function WOA_FS (regularized_X_train, regularized_y_train, number_of_whales, number_of_iterations)

 Function fitness(features)

 Calculate fitness, penalize with features

 Return score

 End Function

 Function transfer function(x)

 Return $1 / (1 + e^{-x})$

 End Function

 Initialize whales, best position, best_fitness

 For i in range (num_iterations)

 Update a

 For j in range(num_whales)

 Calculate A, C, D

 Select new_position based on random value p

 Calculate new position binary, new_fitness

 Update the best position best fitness if the new fitness is better

 Update whales [j, :] if new_fitness is better

 End For

 End For

 Return indices where the best position is equal to 1

End Function

Algorithm 1.

Fitness function. Our study presents a novel approach to the fitness function in the whale optimization algorithm (WOA). Central to this method is integrating a logistic regression model with L1 regularization, a choice motivated by the model's inherent capacity for feature selection and sparsity⁵².

The logistic regression model, known for its effectiveness in binary classification problems, is enhanced with L1 regularization. This regularization technique introduces a penalty term equivalent to the absolute value of the magnitude of the coefficients. The primary advantage of incorporating L1 regularization is its tendency to produce sparse solutions, inherently performing feature selection by driving the coefficients of less significant features to zero^{54,55}. This aspect is particularly beneficial in our study, where model simplicity and interpretability are paramount.

In evaluating the fitness of the WOA, we adopt a cross-validation with five folds. This partitions the data into five subsets, iteratively using one subset for validation and the rest for training. Such a technique strengthens the model's validation on different data samples and mitigates the risk of overfitting, leading to a more reliable performance evaluation⁵⁶.

Also, while L1 regularization naturally promotes feature sparsity, our approach further penalizes the fitness score based on the number of features the logistic regression model selects. This penalty is designed to encourage the selection of the most relevant features and enhance interpretability.

With its unique integration of logistic regression with L1 regularization, k-fold cross-validation, and feature selection penalty, our proposed fitness function is a robust tool in the whale optimization algorithm. It adheres to the principles of parsimony and generalizability and aligns to achieve high predictive accuracy while maintaining model simplicity. This approach has significant implications for applications in various domains, particularly those involving high-dimensional datasets where feature selection is critical in model performance.

Transfer function. In the research, a sigmoid transfer is defined for use in WOA. The sigmoid function is mathematically depicted as Eq. (9).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

The sigmoid function can receive a numerical input in real numbers and subsequently act upon it by transforming input data into a finite range between 0 and 1. The sigmoid function converts the continuous output of the whale optimization algorithm into a binary format amenable to indicating the inclusion or exclusion of features. By employing a threshold value on the sigmoid, established at 0.5, we can ascertain the appropriateness of incorporating a feature, denoted as '1', or removing it as '0'.

Optimal features selected. WOA is a metaheuristic algorithm with stochastic characteristics; hence, it generates marginally distinct feature sets for every dataset in ten (10) different runs. WOA was executed ten times for every dataset. The selected frequency for each feature was determined by tallying the occurrences across ten separate runs indicating the prominence of each feature. The features that exhibit the highest occurrence among the various iterations are considered potential optimal features due to their higher consistency and relevance. To determine the final set of optimal features, a threshold value of 80% was established. This threshold ensures that only those features that appear in at least 80% of the experimental runs (a minimum of 8 out of 10 runs) will be selected. The optimal features chosen are subjected to individual evaluation by executing the classifiers. The performance of these features is documented for further analysis. The findings offer valuable insight into the effects of feature selection on model performance. Using the methodology above, the experiment endeavours to alleviate the stochastic nature inherent in the whale optimization algorithm (WOA) and produce a resilient collection of optimal features for diagnosing heart disease risk. Table 2 outlines the various datasets and the optimal features identified and selected.

Visualization after feature selection. The DBSCAN Outlier Detection scatter plots across the datasets illustrate the DBSCAN algorithm's ability to identify clusters and outliers. The clusters appear more distinct in the selected features compared to the broader spread seen in the full feature set. This suggests that feature selection has potentially removed noisy variables, allowing for more apparent patterns to emerge. The Correlation Heatmaps provide a detailed view of the interdependencies between features. After feature selection, the heatmaps are generally less cluttered, with fewer variables exhibiting strong correlations. This reduction in multicollinearity can benefit many machine learning models, as it tends to enhance model interpretability and performance. The

#	Dataset	Total features	No. of features selected	Optimal features selected
1	Combined	13	9	Age, cp, trestbps, fbs, thalach, exang, oldpeak, ca, thal
2	Z-AlizadehSani dataset	53	34	Age, sex, BMI, DM, HTN, current smoker, FH, CVA, DLP, lung rales, systolic murmur, typical chest pain, dyspnea, function class, atypical, nonanginal, exertional CP, Q wave, St elevation, St depression, tinversion, LVH, poor R progression, FBS, TG, LDL, HDL, ESR, HB, Na, Neut, PLT, EF-TTE, VHD
3	Framingham heart study	15	9	Male, age, cigspersday, prevalentstroke, prevalenthyp, totchol, sysbp, heartrate, glucose
4	SA heart dataset	9	7	Sbp, tobacco, ldl, adiposity, famhist, typea, age
5	Cleveland UCI	13	9	Age, cp, trestbps, fbs, thalach, exang, oldpeak, ca, thal

Table 2. Optimal features selected.

boxplots also highlight the distribution and variance of each feature within the datasets. Post-feature selection, the plots are fewer but more focused, often showing a reduced number of extreme outliers. This indicates that feature selection has likely discarded features with extreme values that could skew the model's learning process. The shapes of the target distributions are generally consistent before and after feature selection, signifying that the selected features maintain the original structure of the target variable. However, in some cases, the distribution appears more balanced post-feature selection, which may positively influence model performance, especially in classification tasks. In all, the visualization suggests that feature selection has streamlined the datasets, potentially improving the efficiency and efficacy of subsequent analyses. By focusing on the most informative features, we expect that the chosen subsets will provide clearer, more relevant insights and facilitate the development of more robust predictive models. Figures 9, 10, 11, 12, 13 and 14 present a visual representation of the various features after performing feature selection.

Training and testing models

After selecting the optimal features, the data is passed to the prediction model for training and testing. The models used are Support Vector Machine, Decision Tree, Random Forest, Multi-Layer Perceptron, Recurrent Neural Network, Adaptive Boosting, Long Short-Term Memory, Extreme Gradient Boosting, K-Nearest Neighbors, and Naïve Bayes classifiers. The choice of prediction models used ranges from classical ML classifiers, ensemble classifiers, and deep learning classifiers to get a broader perspective on performance.

Model evaluation metrics

The models are evaluated using accuracy, precision, recall, AUC and F1 score. The evaluation metrics utilized and the reasons for the choice are discussed next.

Accuracy. Accuracy measures the frequency with which a model accurately predicts outcomes. This metric provides a simple and intuitive understanding of how well the model performs regarding correct classifications^{57,58} making it a valuable choice for this work. The formula for calculating accuracy is provided in Eq. (10).

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

Precision. Precision primarily focuses on minimizing false positives to avoid unnecessary stress and medical interventions for patients⁵⁹. High precision indicates that the model accurately identifies heart disease cases,

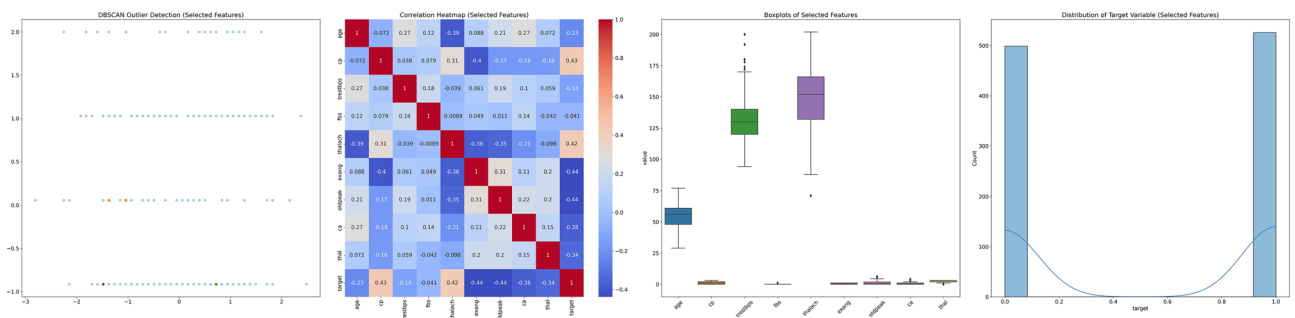


Figure 9. Combined dataset visualization after feature selection.

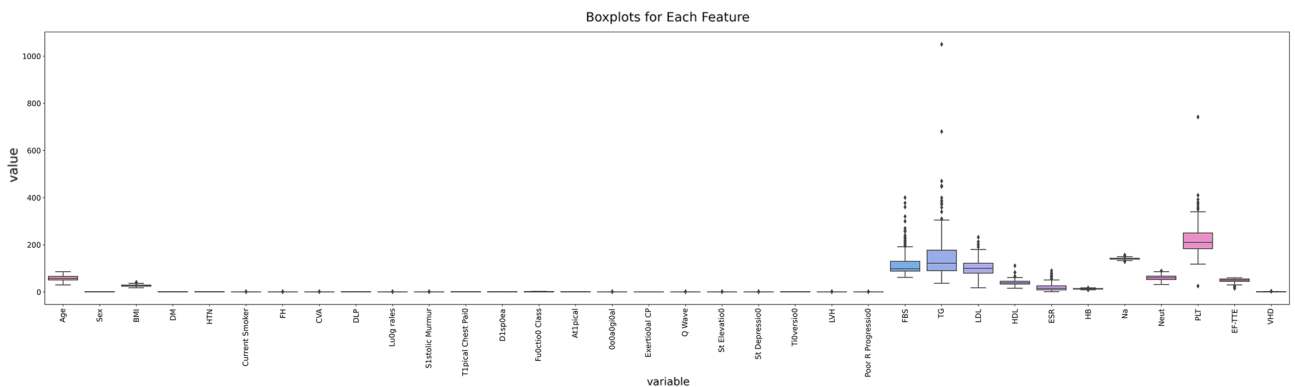


Figure 10. DBAZ-AlizadehSani dataset BoxPlot visualization after feature selection.

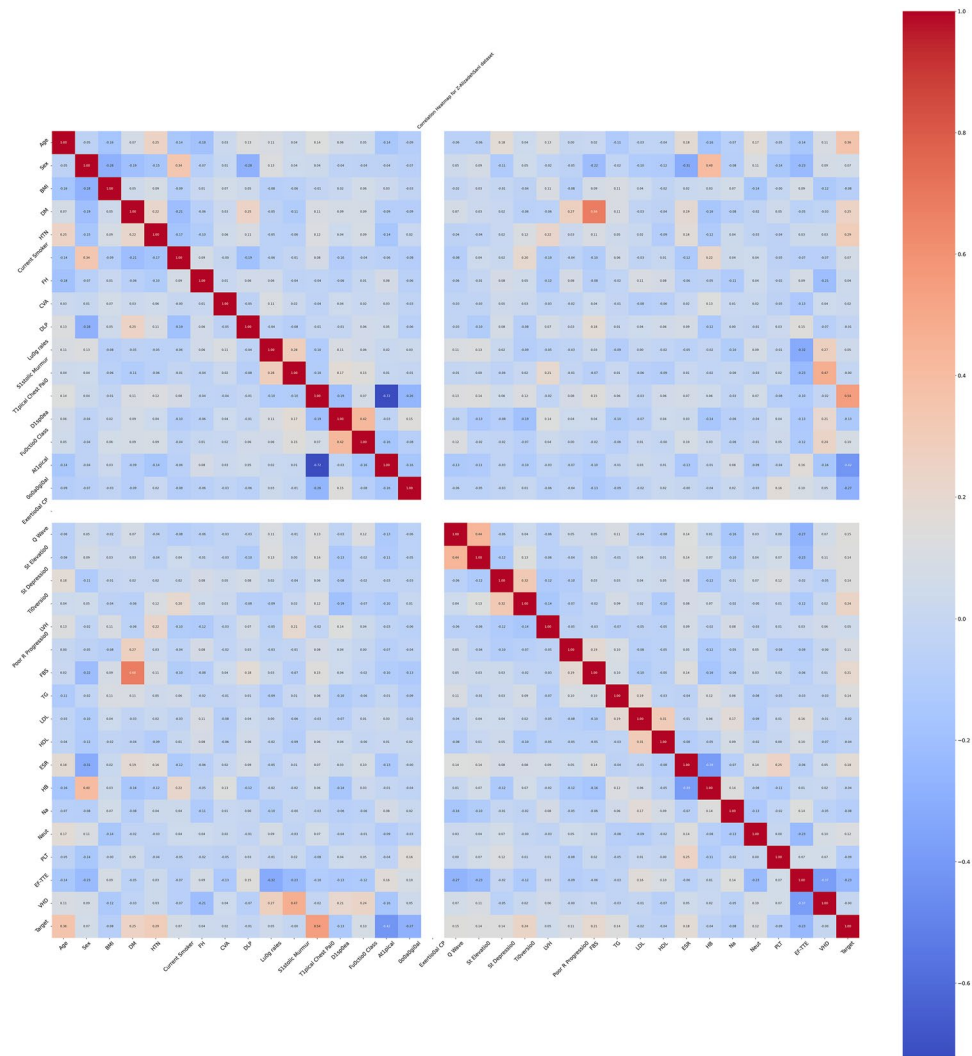


Figure 11. DBAZ-AlizadehSani dataset BoxPlot visualization after feature selection.

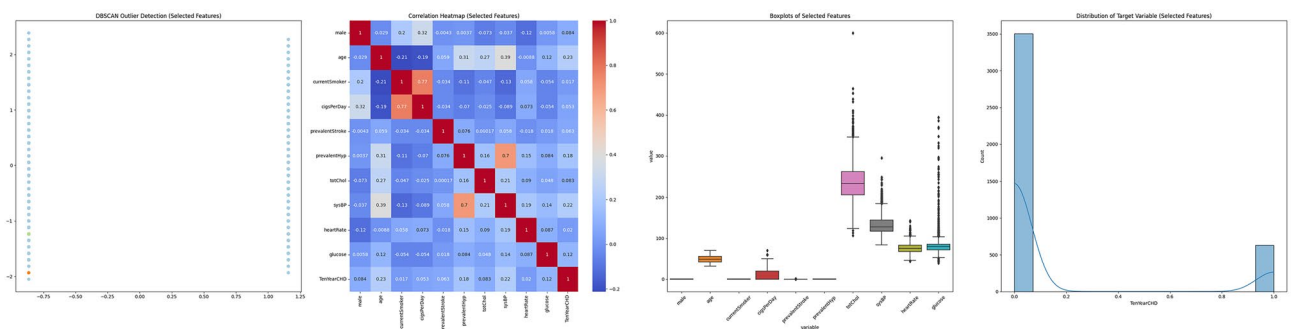


Figure 12. Framingham dataset visualization after feature selection.

facilitating informed clinical decision-making⁶⁰. This is crucial in building trust in the model’s diagnostic capabilities, hence its usage in our work. The precision formula is listed as Eq. (11).

$$\text{precision} = \frac{TP}{TP + FP} \tag{11}$$

#	Classifier	Hyperparameters
1	Adaptive boosting	n_estimators = 50, max_depth = 1, learning_rate = 1, algorithm = SAMME.R
2	Decision tree	max_depth = 3, random_state = 42
3	K-nearest neighbor	n_neighbors = 3, weights = uniform, algorithm = auto, leaf_size = 30, p = 2, metric = default
4	Multi-layer perceptron	Hidden_layer_sizes = 100, activation = relu, solver = adam, learning_rate = constant
5	Long short-term memory	batch_size = 32, epochs = 75, learning_rate = 0.001, LSTM Units = 50, LSTM layers = 7, optimizer = adam, activation = sigmoid
6	Naïve Bayes	Gaussian Naïve Bayes with var_smoothing of 1e-9
7	Random forest	n_estimators = 100, max_depth = 5
8	Recurrent neural network	Number of units in RNN = 32, activation function = sigmoid, optimizer = adam, loss function = binary_crossentropy, epochs = 20, batch_size = 32
9	eXtreme gradient boosting	Objective = binary:hinge, random_state = 42, n_estimators = 100, learning_rate = 0.3, max_depth = 6, subsample = 1, colsample_bytree = 1, min_child_weight = 1
10	Support vector machine	Kernel = linear, class_weight = balanced. All other hyperparameters are left in default

Table 3. Model hyperparameters.

Results and discussions

This section discusses the results obtained from the baseline studies and the experiments conducted with feature selection (WOA).

Baseline experiments without feature selection

The baseline experiments in this subsection serve as a critical foundation, providing valuable insights and setting a benchmark for the subsequent evaluation of feature selection methods. The experimentation results are captured in Tables 4, 5 and 6.

Dataset	Combined dataset					Z-AlizadehSani dataset				
	Classifier	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
AdaBoost	89.27	86.96	93.46	90.09	89.08	80.33	84.78	88.64	86.67	73.73
DT	84.39	84.40	85.98	85.19	84.32	81.97	86.67	88.64	87.64	76.67
KNN	91.22	92.38	90.65	91.51	91.25	60.66	72.73	72.73	72.73	51.07
MLP	85.56	85.10	88.04	86.40	85.45	70.49	77.94	86.59	78.17	57.71
LSTM	88.88	89.43	88.63	88.77	88.63	74.75	55.88	59.22	55.08	59.22
NB	85.37	89.72	83.48	85.26	85.63	73.77	68.18	93.75	72.08	72.74
RF	94.20	90.61	87.56	91.95	92.39	88.52	99.77	86.47	92.63	92.73
RNN	83.32	83.69	83.09	83.18	83.09	72.13	36.07	50.00	41.90	50.00
XGB	95.70	95.10	97.59	96.13	95.82	84.43	94.32	85.54	89.71	83.00
SVM	85.37	91.59	82.35	86.73	85.94	83.61	88.64	88.64	88.64	79.61

Table 4. Combined and Z-AlizadehSani datasets without FS.

Dataset	Framingham dataset					Cleveland-UCI				
	Classifier	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
AdaBoost	85.13	57.89	8.73	15.17	53.79	83.33	93.75	78.95	85.71	84.93
DT	84.52	33.33	1.59	3.03	50.51	68.33	88.00	57.89	69.84	72.13
KNN	82.47	32.73	14.29	19.89	54.50	60.00	81.82	47.37	60.00	64.59
MLP	84.75	49.30	5.87	10.21	52.40	81.17	94.69	74.74	83.16	83.50
LSTM	84.76	47.39	50.07	46.03	50.07	71.83	76.68	76.81	71.71	76.81
NB	82.95	18.25	37.70	57.49	62.13	83.33	76.32	96.67	83.03	83.33
RF	84.98	5.16	54.56	9.17	69.96	77.67	67.37	96.24	79.23	79.58
RNN	84.76	44.89	50.03	45.95	50.03	80.50	80.87	83.17	80.20	83.17
XGB	82.89	12.30	32.82	17.72	59.33	67.50	60.53	83.73	70.26	68.76
SVM	67.35	68.25	27.22	38.91	59.69	81.67	78.95	90.91	84.51	80.64

Table 5. Framingham and Cleveland UCI datasets without FS.

Dataset	SA heart dataset				
Classifier	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)
AdaBoost	69.89	58.33	43.75	50.00	63.68
DT	63.44	44.44	25.00	32.00	54.30
KNN	58.06	36.00	28.13	31.58	50.95
MLP	69.68	59.55	41.69	48.31	62.18
LSTM	64.49	56.07	52.11	50.70	54.16
NB	72.04	59.38	59.38	69.03	69.03
RF	70.97	41.25	61.56	49.36	67.66
RNN	68.71	65.78	59.51	57.86	59.51
XGB	65.59	45.31	50.24	47.63	61.43
SVM	0.741935	0.71875	0.6053	0.6571	0.7208

Table 6. SA heart datasets without FS. Key: Acc = Accuracy, Prec. = Precision, Rec. = Recall, F1 = F1 score.

Experiments with WOA feature selection

This subsection discusses the findings from experiments with WOA feature selection.

Evaluating completion time for WOA feature selection

In this subsection, the study delves into the completion time when WOA feature selection is employed. We find that the duration varies across different datasets. Factors such as the number of agents, iterations, and features in the data directly influence the completion time; more factors result in a longer duration. Moreover, the observations in each dataset also impact the model completion time. For instance, despite the Z-AlizadehSani and Cleveland UCI datasets containing 303 records, the former requires more time for feature selection. Taking the Framingham dataset as an example, the highest time consumption occurred with this dataset. The completion times for 10, 20, 30, 40, 50, and 100 agents/iterations were 7.56 s, 182.26 s, 383.06 s, 747.03 s, 1072.46 s, and 6656.57 s, respectively. The SA Heart dataset, with times of 0.46 s, 11.80 s, 30.05 s, 58.56 s, 91.20 s, and 547.01 s, had a longer completion time than the Cleveland dataset, even though the Cleveland dataset possesses three additional features. This discrepancy could be due to the marginally more significant number of records in the Cleveland UCI dataset. In summary, the time required for WOA to complete the feature selection process increases with the number of features and records in each dataset. Figure 15 presents the findings about feature selection time using Whale Optimization.

Average estimation of evaluation metrics

To obtain a general overview of the evaluation metrics, an average is computed for each metric after ten (10) runs for each of the five (5) datasets and all ten (10) models. The results show a general improvement across all metrics and datasets. The average metrics are elaborated in Figs. 16, 17, 18, 19, and 20, depicted per dataset.

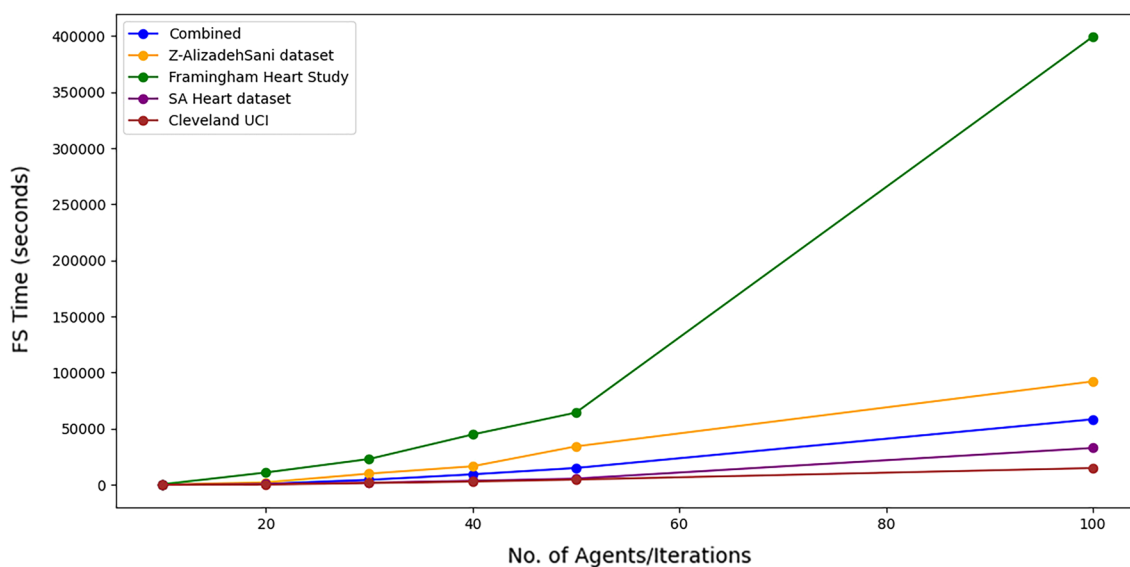


Figure 15. Feature selection time across datasets.

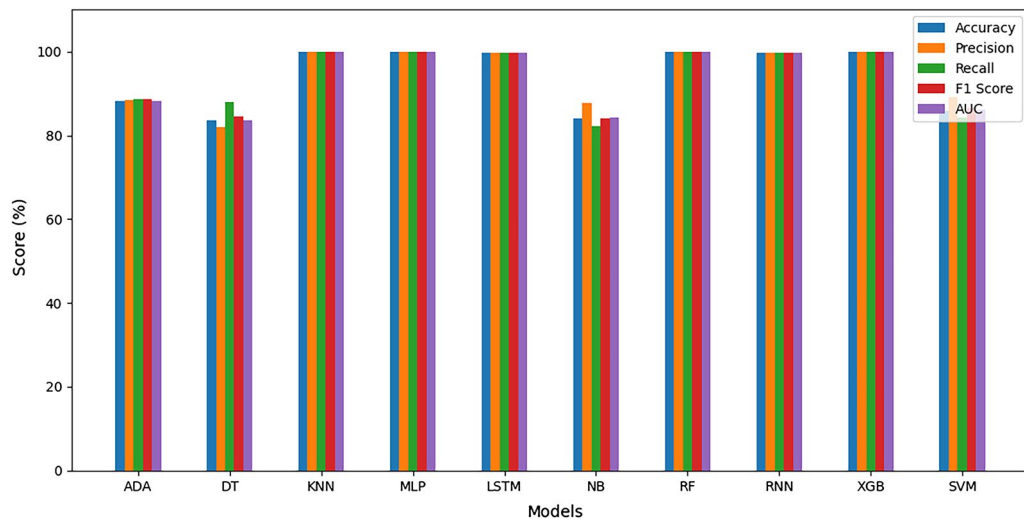


Figure 16. Average evaluation metrics for combined dataset.

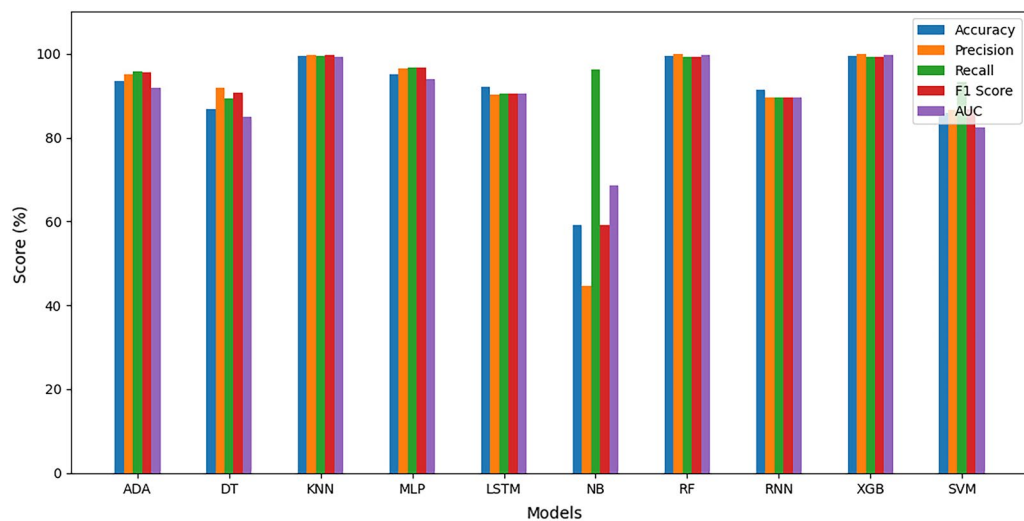


Figure 17. Average evaluation metrics for Z-AlizadehSani dataset.

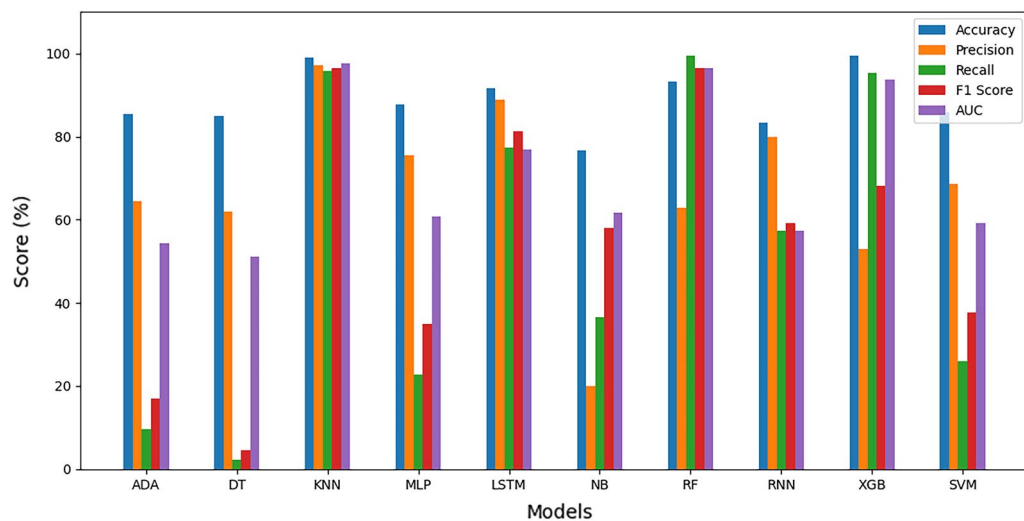


Figure 18. Average evaluation metrics for framingham dataset.

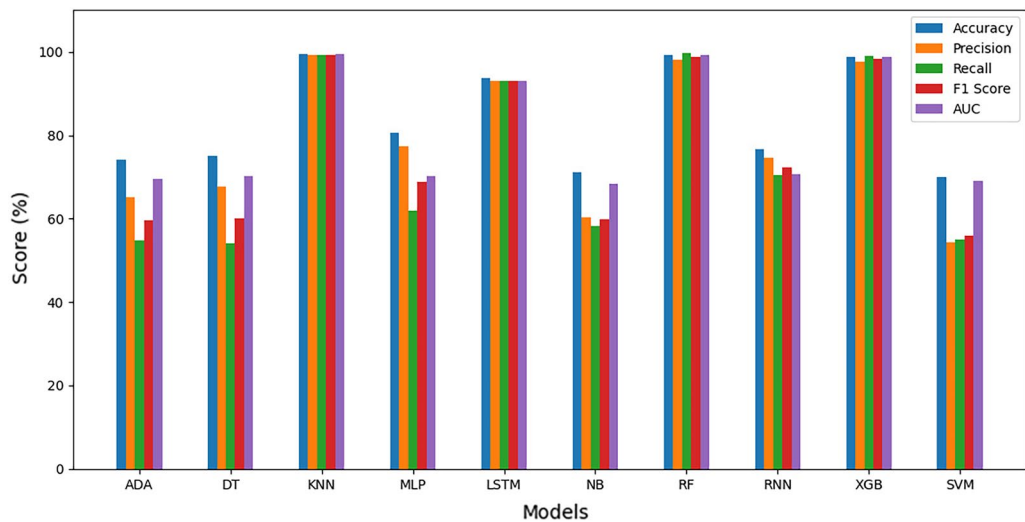


Figure 19. Average evaluation metrics for SA heart dataset.

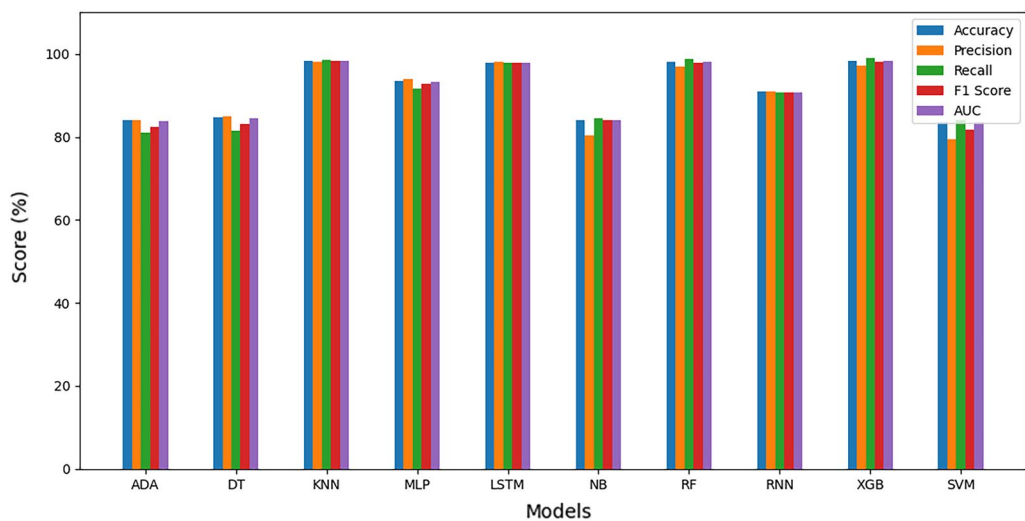


Figure 20. Average evaluation metrics for Cleveland heart dataset.

Evaluating metrics using optimal features

The KNN model performs best across all the metrics and for all the datasets. The results significantly improved over the same experiments on all datasets without WOA feature selection. In some instances, KNN with WOA records 100% across metrics (combined dataset), indicating the influence of WOA on a model's performance. XGB models perform very well across all the metrics on all datasets except the Framingham dataset. Though XGB performs lower on the Framingham dataset than on all the other datasets, the performance across all the valuation metrics is significantly improved over values recorded without WOA feature selection on the Framingham dataset. In essence, the model has improved due to the feature selection just lower than recorded for other datasets when feature selection is employed. The RF classifier also performs well on all the datasets except the Framingham dataset, which performs well on all the metrics apart from the lowest precision (4.78%). The LSTM model records the highest, 99.80%, on all the metrics except the AUC, which recorded 99.81%. The best performance is recorded on the combined dataset for all the models. Similar performance improvement is seen across all the other datasets for the deep learning models. Generally, the ensemble models (RF, XGB and AdaBoost) recorded consistent and improved performance across all the datasets on the evaluation metrics than all but KNN for the classical ML models. The deep learning models perform comparatively well, especially on larger datasets. It was also observed that though the SA heart dataset has a relatively small number of features (9 features), WOA selected seven, subsequently improving the model metrics across all the models. This shows that even if there are not many features, WOA can still select the optimal features if they exist. The experiments consistently improve the evaluation metrics when the results from the model with feature selection are compared to those without. The evaluation metrics are captured in Tables 7, 8 and 9.

Dataset	Combined dataset					Z-AlizadehSani dataset				
	Classifier	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
SVM	82.05	87.83	79.38	83.39	82.47	85.48	84.72	94.33	94.33	89.27
NB	81.85	84.79	80.80	81.81	81.94	60.40	46.76	95.28	95.28	60.24
RF	91.32	94.49	89.23	91.78	91.52	92.74	96.30	93.69	93.69	94.98
XGB	100.00	100.00	100.00	100.00	100.00	98.35	99.07	98.62	98.62	98.85
AdaBoost	87.02	86.59	88.40	87.49	86.99	92.74	94.50	95.37	95.37	94.93
DT	82.83	81.70	85.74	83.67	82.75	85.48	93.88	85.19	85.19	89.32
KNN	99.80	99.62	100.00	99.81	99.80	99.01	100.00	98.61	98.61	99.30
RNN	98.93	98.92	98.93	98.93	98.93	94.72	93.82	93.21	93.21	93.51
MLP	99.41	99.24	99.62	99.43	99.41	97.03	98.59	97.22	97.22	97.90
LSTM	99.80	99.80	99.81	99.80	99.81	94.72	92.90	94.58	94.58	93.68

Table 7. Model performance on combined and Z-AlizadehSani datasets using optimal features.

Dataset	Framingham dataset					Cleveland-UCI				
	Classifier	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
SVM	65.71	83.16	83.16	83.16	83.16	83.16	84.65	87.58	87.58	86.09
NB	83.79	83.50	83.50	83.50	83.50	83.50	85.83	86.51	86.51	84.59
RF	85.53	90.24	90.24	90.24	90.24	90.24	93.31	88.76	88.76	90.98
XGB	90.08	96.97	96.97	96.97	96.97	96.97	97.83	97.07	97.07	97.45
AdaBoost	85.63	83.16	83.16	83.16	83.16	83.16	89.30	90.35	90.35	89.82
DT	84.93	85.52	85.52	85.52	85.52	85.52	85.77	88.98	88.98	87.34
KNN	99.44	98.32	98.32	98.32	98.32	98.32	99.41	98.82	98.82	99.11
RNN	85.97	90.91	90.91	90.91	90.91	90.91	91.95	91.97	91.97	91.96
MLP	87.15	91.92	91.92	91.92	91.92	91.92	95.12	95.87	95.87	95.49
LSTM	92.52	98.99	98.99	98.99	98.99	98.99	94.68	94.76	94.76	94.72

Table 8. Model performance on framingham and cleveland datasets using optimal features.

Dataset	SA heart disease				
	Classifier	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
SVM	68.83	73.13	53.67	53.67	61.90
NB	71.65	61.88	58.58	58.58	69.08
RF	80.30	57.50	80.00	80.00	66.91
XGB	97.40	95.63	96.84	96.84	96.23
AdaBoost	76.41	68.35	59.38	59.38	63.55
DT	74.68	68.07	50.63	50.63	58.06
KNN	99.57	100.00	98.75	98.75	99.37
RNN	76.19	73.84	72.24	72.24	72.85
MLP	85.06	85.83	68.13	68.13	75.96
LSTM	82.90	81.00	82.22	82.22	81.50

Table 9. Model performance on SA heart. Key: Acc = Accuracy, Prec. = Precision, Rec. = Recall, F1 = F1 score.

Comparative studies

This section performs a comparative study with other related works, comparing the metrics used in the study with values from the metrics of the optimal features used in this work. Works used are the most recent studies in the domain that utilize the same datasets. Table 10 outlines recent studies with the same five datasets used in this work. We compared the features selected by the respective algorithms and their metrics reported and compared it with results from the optimal features in Tables 6, 7 and 8. The results when WOA is used show significant improvement.

#	Author	#Features	Dataset	FS method	Metric
1	Wadhawan and Maini ⁶⁶	9	Combined	Optimal feature subset selection algorithm (OFSSA)	Accuracy: 99.666%, Precision: 99.832%, Recall/sensitivity: 99.832%, F1: 99.832%, AUC: N/A
2	Ghosh et al. ¹⁸	10	Combined	Filter (relief and lasso)	Accuracy 99.05%
3	Kolukisa and Bakir-Gungor ⁶⁷	25	Z-AlizadehSani	Hybrid method (exhaustive and probabilistic ensemble)	Accuracy: 91.14%, Precision: 93.53%, Recall/sensitivity: 94.41%, F1: 93.7% 99.832%, AUC: 94.1%
4	Fajri et al. ⁶⁸	24	Z-AlizadehSani	Hybrid (Bee swarm optimization algorithm combined with Q-learning for optimizing)	Accuracy: 90.1%, Precision: N/A, Recall/sensitivity: N/A, F1: N/A, AUC: 94.1%
5	Kolukisa and Bakir-Gungor ⁶⁷	9	Cleveland	Hybrid method (exhaustive and probabilistic ensemble)	Accuracy: 85.47%, Precision: 86.22%, Recall/sensitivity: 82.96%, F1: 83.9% 99.83%, AUC: 91.1%
6	El-Shafiey et al. ⁶⁹	7	Cleveland	Hybrid (Genetic algorithm and particle swarm optimization)	Accuracy: 91.40%, Precision: N/A, Recall/sensitivity: N/A, F1: N/A, AUC: N/A
7	Fajri et al. ⁶⁸	6	Cleveland	Hybrid (Bee swarm optimization method and Q-learning)	Accuracy: 84.4%, Precision: N/A, Recall/sensitivity: N/A, F1: N/A, AUC: 90.1%
8	Ay et al. ⁴⁹	N/A	Cleveland	Swarm algorithms (Cuckoo search (CS), Flower pollination algorithm (FPA), and Harris hawks optimization (HHO))	Accuracy: N/A, Precision: N/A, Recall/sensitivity: N/A, F1: 99.72%, AUC: 98.0%
9'	Budholiya et al. ⁷⁰	N/A	Cleveland	None	Accuracy: 91.80%, Precision: N/A, Recall/sensitivity: 85.71%, F1: 90.56%, AUC: N/A
10	Owusu et al. ⁷¹	10	Cleveland	Filter (Extra tree classifie)	Accuracy: 99.75%, Precision: 94%, Recall/sensitivity: 87%, F1: 90%, AUC: 93.2%
11	Mienye and Su ⁷²	N/A	Framingham	None	Accuracy: 91%, Precision: 92%, Recall/sensitivity: 90%, F1: 91%, AUC: N/A
12	Rahim et al. ⁷³	6	Framingham	Filter (feature importance)	Accuracy: 99.1%, Precision: N/A, Recall/sensitivity: N/A, F1: N/A, AUC: N/A
13	Krishnani et al. ⁷⁴	N/A	Framingham	None	Accuracy: 96.71%, Precision: 98.94%, Recall/sensitivity: 94.4%, F1: 96.61%, AUC: N/A
14	Mahmoud et al. ⁷⁵	N/A	Framingham	Embedded method	Accuracy: 85.05%, Precision: N/A, Recall/sensitivity: N/A, F1: 91.90%, AUC: N/A
15	Nalluri et al. ⁷⁶	N/A	Framingham	None	Accuracy: 85.86%, Precision: N/A, Recall/sensitivity: N/A, F1: N/A, AUC: N/A
16	Anuradha and David ⁷⁷	6	SA heart	Filter	Accuracy: 78.49%, Precision: 78%, Recall/sensitivity: 91%, F1: 84%, AUC: N/A
17	Khdair and Dasari ⁷⁸	N/A	SA heart	None	Accuracy: 78.1%, Precision: 78.4%, Recall/sensitivity: 77.6%, F1: 78%, AUC: N/A
18	Gokulnath and Shantharajah ⁷⁹	7	Cleveland UCI	Wrapper (genetic algorithm)	Accuracy: 88.34%, Precision: N/A, Recall/sensitivity: N/A, F1: N/A, AUC: N/A
19	Cenitta et al. ⁸⁰	13	Combined	Swarm (squirrel search feature selection algorithm)	Accuracy: 98.38%, Sensitivity: 98.66%, Specificity: 98.10%, F1: 98.32%
20	Zhang et al. ²³	12	Cleveland UCI	Embedded (LinearSVC Algorithm)	Accuracy: 98.56%, Precision: 97.84%, Recall: 99.35% Specificity: N/A, F1: N/A
21	Haq et al. ¹⁰	6	Cleveland UCI	Wrapper (sequential backward selection)	Accuracy: 90.00%, Precision: N/A, Recall: N/A, Specificity: N/A, F1: N/A
22	Proposed model	9	Combined	Swarm (whale optimization algorithm)	Accuracy: 100.00%, Precision: 100.00%, Recall/sensitivity: 100.00%, F1: 100.00%, AUC: 100.00%
23	Proposed model	35	Z-AlizadehSani	Swarm (whale optimization algorithm)	Accuracy: 99.01%, Precision: 100.00%, Recall/sensitivity: 98.61%, F1: 98.61%, AUC: 99.30%
24	Proposed model	9	Framingham	Swarm (whale optimization algorithm)	Accuracy: 99.44%, Precision: 98.32%, Recall/sensitivity: 98.32%, %, F1: 98.32%, %, AUC: 98.32%
25	Proposed model	7	SA heart	Swarm (whale optimization algorithm)	Accuracy: 98.32%, Precision: 99.41%, Recall/sensitivity: 98.82%, F1: 98.82%, AUC: 99.11%
26	Proposed model	9	Cleveland UCI	Swarm (whale optimization algorithm)	Accuracy: 99.57%, Precision: 100%, Recall/sensitivity: 98.75%, F1: 98.75%, AUC: 99.37%

Table 10. Comparative studies with recent works.

The results show superior metrics recorded when WOA feature selection is used and the impact of WOA on the model performance. It is worth noting that this work measures five evaluation metrics over five different datasets using ten models, providing an exhaustive list of metrics capable of judging the performance reported for every model.

Table 10 lists the most recent works in the HD domain, with the last five being the results obtained from this study on the optimal features.

Limitations of the study

The study has data limitations as image heart disease datasets are out of the scope of this work. The experiments in this study are not performed for more than 100 agents and 100 iterations for the whale optimization algorithm due to computational constraints. Finally, the work is limited to only the heart disease domain.

Conclusion

The original WOA is designed for continuous optimization. In this work, however, we implement the solution space into a binary form, necessitating the transformation of the output from continuous to binary using a sigmoid transfer function. A thresholding process is then used to convert the values to a binary format. We specify a fitness function which involves training a logistic regression model using the selected features and evaluating its performance through cross-validation. Since the aim is feature selection, the logistic regression model is penalized using an L1 (Lasso) to produce more sparse models, where a subset of coefficients becomes exactly zero, technically performing feature selection. We then introduce a penalty term to discourage the selection of more features. The whale optimization algorithm for feature selection shows improved heart disease risk prediction evaluation metrics. With fewer yet relevant features, models significantly improved across all five metrics. This contributes to the ongoing efforts to enhance the effectiveness of cardiovascular health risk, providing valuable insights for future studies in this field.

Future works can consider using WOA on image datasets for feature extraction. How WOA compares to other swarm intelligence algorithms in the heart disease domain is an area worth researching. Extending the research to more than one dataset in domains other than Cardiovascular health is a useful area for future studies. In addition, future studies can consider more hyperparameter tuning of all the models to improve the results obtained.

Data availability

Combined dataset <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>, Z-AlizadehSani dataset <https://www.kaggle.com/datasets/tanyachi99/zalizadeh-sani-dataset-2csv>, Framingham dataset <https://www.kaggle.com/datasets/captainozlem/framingham-chd-preprocessed-data>, South African Heart dataset <https://hastie.su.domains/ElemStatLearn/datasets/SAheart.data>, Cleveland UCI dataset <https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci>.

Received: 24 August 2023; Accepted: 19 February 2024

Published online: 24 February 2024

References

- World Health Organization. *Cardiovascular Diseases 2020*. [Online] (Accessed 10 March 2022); https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1
- Ghwanmeh, S., Mohammad, A. & Al-Ibrahim, A. Innovative artificial neural networks-based decision support system for heart diseases diagnosis. *J. Intell. Learn. Syst. Appl.* **5**(3), 176–183 (2013).
- Staffini, A. *et al.* Heart rate modeling and prediction using autoregressive models and deep learning. *Sensors* **22**(1), 1–13 (2022).
- Anshori, M. & Haris, M. S. Predicting heart disease using logistic regression. *Knowl. Eng. Data Sci.* **5**(2), 188–196 (2023).
- Shah, D., Patel, S. & Bharti, S. K. Heart disease prediction using machine learning techniques. *SN Comput. Sci.* **1**, 1–6 (2020).
- Wang, Y., Pan, Z. & Dong, J. A new two-layer nearest neighbor selection method for kNN classifier. *Knowl.-Based Syst.* **235**, 107604 (2022).
- Verma, E. P. & Singh, E. P. Human heart disease prediction system using enhanced decision tree algorithm in data mining. *Int. J. Innov. Sci. Eng. Technol.* **8**(6), 1–7 (2021).
- Bharti, R. *et al.* Prediction of heart disease using a combination of machine learning and deep learning. *Comput. Intell. Neurosci.* **2021**, 11 (2021).
- Amin, S. M., Kia, Y. & Dewi, K. Identification of significant features and data mining techniques in predicting heart disease. *Telematics Inform.* **36**, 82–93 (2019).
- Haq, A. U., Li, J., Memon, M. H., Memon, M. H., Khan, J. & Mariam, S. M. Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection, in *IEEE 5th International Conference for Convergence in Technology (I2CT)* (2019).
- Hicks, S. A. *et al.* On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* **12**, 5979 (2022).
- Benítez-Caballero, M. J., Medina, J., Ramírez-Poussa, E. & Ślęzak, D. Bireducts with tolerance relations. *Inf. Sci.* **435**, 26–39 (2018).
- Zeniarja, J., Ukhifadhina, A. & Salam, A. Diagnosis of heart disease using K-nearest neighbor method based on forward selection. *J. Appl. Intell. Syst.* **4**(2), 39–47 (2019).
- Farahat, A. K., Ghodsi, A. & Kamel, M. S. Efficient greedy feature selection for unsupervised learning. *Knowl. Inf. Syst.* **35**(2), 285–310 (2013).
- Wang, S., Chen, J., Guo, W. & Liu, G. Structured learning for unsupervised feature selection with high-order matrix factorization. *Expert Syst. Appl.* **140**, 112878 (2020).
- Pathan, M. S., Nag, A., Pathan, M. M. & Dev, S. Analyzing the impact of feature selection on the accuracy of heart disease. *Healthc. Anal.* **1**, 100060 (2022).
- Bommert, A., Sun, X., Bischl, B., Rahnenführer, J. & Lang, M. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* **143**, 106839 (2020).
- Ghosh, P., Azam, S., Karim, A., Jonkman, M., & Hasan, M. Z. Use of efficient machine learning techniques in the identification of patients with heart diseases, in *5th International Conference on Information System and Data Mining (ICISDM 2021)* (2021).
- Narsimhulu, K., Ramchander, N. S., & Swathi, A. An AI enabled framework with feature selection for efficient heart disease prediction, in *2022 5th International Conference on Contemporary Computing and Informatics* (2022).
- Ditzler, G., Polikar, R. & Rosen, G. A sequential learning approach for scaling up filter-based feature subset selection. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(6), 2530–2544 (2017).
- Taha, A., Hadi, A. S. & Bernard Cosgrave, S. M. A multiple association-based unsupervised feature selection algorithm for mixed data sets. *Expert Syst. Appl.* **212**, 118718 (2023).
- Mostafa, S. A. *et al.* Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cogn. Syst. Res.* **54**, 90–99 (2019).
- Zhang, D. *et al.* Heart disease prediction based on the embedded feature selection method and deep neural network. *Hindawi* **2021**, 1–9 (2021).
- Hutamaputra, W., Mawarni, M., Krisnabayu, R. Y., & Mahmudy, W. F. Detection of coronary heart disease using modified K-NN method with recursive feature elimination, in *6th International Conference on Sustainable Information Engineering* (2021).
- Ang, J. C., Mirzal, A., Haron, H. & Hamed, H. N. A. Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **13**(5), 971–989 (2015).

26. Khaire, U. M. & Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J King Saud Univ. Comput. Inf. Sci.* **34**(4), 1060–1073 (2022).
27. Firdaus, F. F., Nugroho, H. A. & Soesanti, I. A review of feature selection and classification approaches for heart disease prediction. *Int. J. Inf. Technol. Electric. Eng.* **4**(3), 75–82 (2020).
28. Ghosh, P. *et al.* Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. *IEEE Access* **9**, 19304–19326 (2021).
29. Pavya, K. & Srinivasan, B. Feature selection techniques in data mining: A study. *Int. J. Sci. Dev. Res.* **2**(6), 594–598 (2017).
30. Acharjya, D. P. A hybrid scheme for heart disease diagnosis using rough set and cuckoo search technique. *J. Med. Syst.* **44**(1), 1–16 (2020).
31. Mandal, M., Singh, P. K., Ijaz, M. F., Shafi, J. & Sarkar, R. A tri-stage wrapper-filter feature selection framework for disease classification. *Sensors* **21**, 5571 (2021).
32. Arroyo, J. C. T. & Delima, A. J. P. An optimized neural network using genetic algorithm for cardiovascular disease prediction. *J. Adv. Inf. Technol.* **13**(1), 95–99 (2022).
33. Khouridfi, Y. & Bahaj, M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *Int. J. Intell. Eng. Syst.* **12**(1), 242–252 (2019).
34. Prayogo, R. D. & Karimah, S. A. Hybrid feature selection with K-nearest neighbors for optimal heart failure detection, in *2022 12th International Conference on System Engineering and Technology (ICSET), Bandung, Indonesia* (2022).
35. Rostami, M., Berahmand, K., Nasiri, E. & Forouzande, S. Review of swarm intelligence-based feature selection methods. *Eng. Appl. Artif. Intell.* **100**, 104210 (2021).
36. Usman, A. M., Yusof, U. K. & Naim, S. Cuckoo inspired algorithms for feature selection in heart. *Int. J. Adv. Intell. Inf.* **4**(2), 95–106 (2018).
37. Al-Tashi, Q., Rais, H., & Jadid, S. Feature selection method based on grey wolf optimization for coronary artery disease classification, in *International Conference of Reliable Information and Communication Technology* (2018).
38. Bakrawy, L. M. E. Grey Wolf optimization And Naive Bayes classifier incorporation for heart disease diagnosis. *Aust. J. Basic Appl. Sci.* **11**(7), 64–70 (2017).
39. Chakraborty, C., Kishor, A. & Rodrigues, J. J. Novel enhanced-Grey Wolf optimization hybrid machine learning technique for biomedical data computation. *Comput. Electric. Eng.* **99**, 107778 (2022).
40. David, V. K. Feature selection using Whale swarm algorithm and a comparison of classifiers for prediction of cardiovascular diseases. *Int. J. Res. Anal. Rev. (IJRAR)* **6**(2), 123–130 (2019).
41. Shahid, A. H. & Singh, M. A novel approach for coronary artery disease diagnosis using hybrid particle Swarm optimization based emotional neural network. *Biocybern. Biomed. Eng.* **40**(4), 1568–1585 (2020).
42. Asadi, S., Roshan, S. & Kattan, M. W. Random forest swarm optimization-based for heart diseases diagnosis. *J. Biomed. Inform.* **115**, 103690 (2021).
43. Wankhede, J., Kumar, M. & Sambandam, P. Efficient heart disease prediction-based on optimal feature selection using DFCSS and classification by improved Elman-SFO. *IET Syst. Biol.* **14**(6), 380–390 (2020).
44. Sureja, N., Chawda, B. V. & Vasant, A. A novel salp swarm clustering algorithm for prediction of the heart diseases. *Indones. J. Electric. Eng. Comput. Sci.* **25**(1), 265–272 (2022).
45. Lee, C.-Y. & Zhuo, G.-L. A hybrid Whale optimization algorithm for global optimization. *Mathematics* **9**, 1477 (2021).
46. Mirjalili, S. & Lewis, A. The Whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016).
47. Pham, Q.-V., Mirjalili, S., Kumar, N., Alazab, M. & Hwang, W.-J. Whale optimization algorithm with applications to resource allocation in wireless networks. *IEEE Trans. Veh. Technol.* **69**(4), 4285–4297 (2020).
48. Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H. & Jianhua, Z. Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resour. Policy* **61**, 250–260 (2019).
49. Ay, Ş., Ekinçi, E. & Garip, Z. A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. *J. Supercomput.* **79**, 11797–11826 (2023).
50. Mezher, M. A. Genetic folding (GF) algorithm with minimal kernel operators to predict stroke patients. *Appl. Artif. Intell.* **1**, 2022 (2022).
51. Nguyen, H. T., Cao, A. H., & Bui, P. H. D. Electrocardiogram-based heart disease classification with machine learning techniques, in *International Conference on Computational Collective Intelligence* (2023).
52. Deza, A. & Atamturk, A. Safe screening for logistic regression with ℓ_0 - ℓ_2 regularization, in *14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022)* (2022).
53. Qin, J. & Lou, Y. L_1 - L_2 regularized logistic regression, in *53rd Asilomar Conference on Signals, Systems, and Computers* (2019).
54. Emmert-Streib, F. & Dehmer, M. High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection. *Mach. Learn. Knowl. Extr.* **1**(1), 359–383 (2019).
55. Patil, A. R. & Kim, S. Combination of ensembles of regularized regression models with resampling-based lasso feature selection in high dimensional data. *Mathematics* **8**(1), 110 (2020).
56. Wong, T.-T. & Yeh, P.-Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **32**(8), 1586–1594 (2020).
57. Chicco, D. & Jurman, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genom.* **21**(1), 1–13 (2020).
58. Chicco, D., Tötsch, N. & Jurman, G. The matthews correlation coefficient (mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min.* **1**(14), 1–22 (2021).
59. Sukegawa, S. *et al.* Multi-task deep learning model for classification of dental implant brand and treatment stage using dental panoramic radiograph images. *Biomolecules* **11**(6), 815 (2021).
60. Seliya, N., Khoshgoftaar, T. M., & Hulse, J. V. A study on the relationships of classifier performance metric, in *2009 21st IEEE International Conference on Tools with Artificial Intelligence* (2009).
61. Ma, W. & Lejeune, M. A. A distributionally robust area under curve maximization model. *Oper. Res. Lett.* **48**(4), 460–466 (2020).
62. Sofaer, H. R., Hoeting, J. A. & Jarnevich, C. S. The area under the precision-recall curve as a performance metric for rare binary events. *Methods Ecol. Evol.* **10**(4), 565–577 (2018).
63. He, H. & Garcia, E. A. Learning from imbalanced data. *Trans. Knowl. Data Eng.* **21**(9), 1263–1284 (2009).
64. Ribeiro, M. T., Singh, S., & Guestrin, C. Why should I trust you? Explaining the predictions of any classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and data Mining* (2016).
65. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 and accuracy in binary classification evaluation. *BMC Genom.* **21**(1), 1–13 (2020).
66. Wadhawan, S. & Maini, R. ETCD: An effective machine learning based technique for cardiac disease prediction with optimal feature subset selection. *Knowl. Based Syst.* **255**, 109709 (2022).
67. Kolukisa, B. & Bakir-Gungor, B. Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis. *Comput. Stand. Interfaces* **84**, 103706 (2023).
68. Fajri, Y. A. Z. A., Wiharto, W. & Suryani, E. Hybrid model feature selection with the bee swarm optimization method and Q-learning on the diagnosis of coronary heart disease. *Information* **14**(15), 1–15 (2023).

69. El-Shafiey, M. G., Hagag, A., El-Dahshan, E. S. A. & Ismail, M. A. A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. *Multimed. Tools Appl.* **81**, 18155–18179 (2022).
70. Budholiya, K., Shrivastava, S. K. & Sharma, V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J. King Saud Univ. Comput. Inf. Sci.* **34**(7), 4514–4523 (2022).
71. Owusu, E., Boakye-Sekyerhene, P. & Appati, J. K. Computer-aided diagnostics of heart disease risk prediction. *Comput. Intell. Neurosci.* **2021**, 3152618 (2021).
72. Mienye, I. D. & Sun, Y. An improved ensemble learning approach for the prediction of heart disease risk. *Inform. Med. Unlocked* **20**, 100402 (2020).
73. Rahim, A. *et al.* An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE Access* **9**, 106575–106588 (2021).
74. Krishnani, D., Kumari, A., Dewangan, A., Singh, A., & Naik, N. S. Supervised machine learning algorithms prediction of coronary heart disease using, in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)* (2019).
75. Mahmoud, W. A., Aborizka, M. & Amer, F. A. E. Heart disease prediction using machine learning and data mining techniques: Application of framingham dataset. *Turk. J. Comput. Math0 Educ. (TURCOMAT)* **12**(14), 4864–4870 (2021).
76. Nalluri, S., Saraswathi, R. V., Ramasubbareddy, S., Govinda, K., Swetha, E. Chronic heart disease prediction using data mining techniques, in *Engineering and Communication Technology, Advances in Intelligent Systems and Computing*, 903–912 (2020).
77. Anuradha, P. & David, V. K. Feature selection and prediction of heart diseases using gradient boosting algorithms, in *Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)* (2021).
78. Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A. & Singh, G. Prediction of coronary heart disease using machine learning: An experimental analysis. *ACM* **12**(5), 28–36 (2019).
79. Gokulnath, C. B. & Shantharajah, S. P. An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Comput.* **22**, 14777–14787 (2019).
80. Cenitta, D., Arjunan, R. V. & Prema, K. V. Ischemic heart disease prediction using optimized squirrel search feature selection algorithm. *IEEE Access* **10**, 122995–123006 (2022).

Author contributions

S.A.A.—Initial Draft, Experimentation, Analysis; J.K.A.—Idea Initialization, Supervision, Analysis, Final Draft; E.O.—Data Mugging, Planning, Supervision, Final Draft.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.K.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024