Journal of Education for Pure Science- University of Thi-Qar
Vol.14, No.3 (Sept., 2024)
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI: https://doi.org/10.32792/jeps.v14i3.538

A New Hybrid Model for Heart Disease Prediction Using Machine Learning Algorithms Optimized by Modified Whale Optimization Algorithm

Zainab Shaker MatarAl-Husseini[1] ,HUDA HADI TALIP[2], Muna Ghani[3] and Ali H. Naser[4]

[1]Faculty of Archaeology, Department of Archeology and Islamic Civilization, University of Thi-Qar, Nasiriyah, Iraq.

[2,4] Thi-Qar Education Directorate.
[3]Computer science department, University of Thi-Qar, Nasiriyah, Iraq.

1) zainab.shaker@utq.edu.iq , https://orcid.org/0000-0002-1927-8569
2) msc.hudaalsafy@gmail.com
3) munaghan23@utq.edu.iq
4) prog.ali95@gmail.com

* Corresponding email: zainab.shaker@utq.edu.iq

Received 02/04 /2024, Accepted 14/06 /2024, Published 01/ 09 /2024

Abstract

Heart disease is a kind of cardiovascular disease (CVD) which globally considered death's number one cause. Data Science plays an important role in processing large data amounts in the healthcare domain. There are several problems that could hinder appropriate cardiac monitoring, including limited of medical dataset, lack of depth analysis, and feature selection. In this paper, we exploit the method of Fast Correlation-Based Feature Selection (FCBF) for filtering extra features for developing heart disease classification quality. Next, we implement the classification according to various algorithms of classification like Decision tree, Naïve Bayes, Logistic Regression, K-Nearest Neighbour (KNN), Random Forest, Support Vector Machine (SVM), and a Multilayer Perception which is optimized with the modified whale optimization algorithm evaluation version named Modified Whale Optimization

Algorithm (MWOA). The proposed model optimized with FCBF as well as MWOA obtain an 83.26% accuracy score by Logistic Regression.

Keywords Heart disease, disease prediction model, machine learning Classification, Feature selection, whale optimization algorithm

1. Introduction

In the last few years, cardiovascular/ heart disease is the most common cause of death in the world keeps death the basis in the world. Huge numerical are caused by heart disease which are typically found. A lot of predisposing features like professional and personal habits as well as genetic accounts of predisposition for the diseased heart. Different habitual factors of risk like stress, smoking, physical inactivity, alcohol, and caffeine overuse with other physiological factors such as high cholesterol or blood, hypertension, obesity, and pre-existing conditions of the heart are factors predisposing to the disease of the heart. Proper and effective early heart disease medical diagnosis plays an important role in taking preventive measures for avoiding death [1].

High-risk individuals of early heart disease identification as well as an improved diagnosis by utilizing the model of prediction have been recommended generally for reducing the rate of fatality as well as improving decision-making for the treatment and prevention further. The model of prediction which is performed in a clinical decision support system (CDSS) can be utilized for helping clinicians measure heart disease risk and also present accurate treatments for controlling risk further. Additionally, several papers have stated that CDSS implementation can develop the quality of decisions, clinical decision-making, and preventive care [2].

Algorithms depend on training and test data consistency, feature selection techniques used like data mining can help in preparing data for presenting more proper prediction. Meanwhile, related features are chosen, and hybrid models and classifiers can be used for predicting disease occurrence opportunities. Several problems exist that might avoid proper heart disease prediction such as feature selection, depth analysis lack, applications of ML algorithm, and limited datasets medical [3].

In this paper, we exploit the method of FCBF for filtering extra features for developing heart disease classification quality. Next, we implement the classification according to various algorithms of classification like Logistic Regression, KNN, Random Forest, Naïve Bayes, Decision tree, SVM, Multilayer Perception optimized with the modified whale optimization algorithm evaluation version named as MWOA [4].

The remaining sections of the paper are below: Section 2 reviews several previous works. Section 3 provides the proposed method. Section 4 discusses the results of the test. the final section provides the paper's conclusion.

2. Related work

In this section the researcher into heart Disease Prediction Using Machine Learning Algorithms Optimized.

Khourdifi et al. [5], in this study, the authors used the FCBF method to redundant features in order to improve the quality of heart disease classification. After that classify the data using a variety of classification algorithms including Random Forest, SVM, KNN, Naïve Bayes, and Multilayer Perception, Artificial Neural Network optimized by Particle Swarm Optimization (PSO) in conjunction with Ant Colony Optimization (ACO) techniques.

Rajdhan et al. [6], in this study, the authors utilize a dataset of heart disease accessible in the ML repository of UCI. +The suggested work uses a variety of data mining techniques, including logistic regression, naive bayes, random forest, and decision trees, to estimate the likelihood of heart disease and to categorize the patient's risk level. As a result, they offer a comparison analysis through performance analysis of different machine learning algorithms. Trial findings confirm that, when compared to other ML algorithms, the Random Forest approach has achieved the highest accuracy.

Bharti et al. [7], in this study, The UCI Machine Learning Heart Disease dataset analysis and outcomes are compared using a variety of machine learning and deep learning algorithms. The dataset has fourteen fundamental features that are used for analysis. By using a confusion matrix and accuracy, many promising outcomes are obtained and validated. The dataset is standardized to produce better results, and a number of unnecessary features are controlled by using Isolation Forest. Additionally, the manner that the paper can be combined with various multimedia technologies—such as mobile devices—is covered.

Jindal et al. [8], in this study, prepared the prediction system of heart disease for predicting whether the patient is likely to be diagnosed with heart disease or not by utilizing the patient medical history. They utilized various ML algorithms like KNN and logistic regression for predicting and classifying patients by heart disease. The effective way was utilized for regulating the way a model can be applied for improving Heart Attack prediction accuracy in each person. proposed model strength was promising and could predict heart disease evidence in the specific person by utilizing Logistic Regression and KNN that illustrated the good accuracy compared with previously utilized classifiers like naive bayes and so on.

Rani et al. [9], in this study, the authors used the system of hybrid decision support which can assist in early heart disease detection according to clinical patient parameters. Writers have utilized multivariate imputation with the chained algorithm of equations for controlling missing amounts. the hybridized algorithm of feature selection integrating recursive feature elimination and Genetic Algorithm (GA) has been utilized for proper feature selection from an accessible set of data. for data pre-processing, standard scalar methods and SMOTE (Synthetic Minority Oversampling Technique) has been applied. In the last proposed hybrid system development stage, it had been utilized the SVM, logistic regression, naive bayes, AdaBoost classifiers, and random forest. it had been recognized that system has given the most appropriate results by the random classifier of the forest.

Sowmiya et al. [10], in this study, it had been presented the improved technique and new selection of features as well as a method of classification for predicting mortality in patients with congestive heart failure. Via the technique, the rate of death because of heart disease will be gradually reduced. The algorithm of ant colony optimization (ACO) is applied to choose the best feature for the classifier of hybrid (KNN). The proposed technique is compared to basic methods of classification like Naïve Bayes, decision tree, SVM, KNN. Cleveland dataset UCI is applied for implementation. This study illustrates HKNN efficiency in the prediction system of heart disease. Essential features were analyzed and the classification is applied for achieving a better result.

Yadav et al. [11], in this study, it had been presented the methods of Lasso Regularization, Recursive Features Elimination, Pearson Correlation, and features selection used on the tree-based classifiers algorithms based on ML: Random Tree, M5P, and Reduced Error Pruning by a method of Random Forest ensemble. They assessed ROC,

precision, classification accuracy, and sensitivity value. They utilized the Repository dataset UCI for 14 features and 1025 samples. They recognize whether the individual is suffering from the issue of the heart not in ML algorithms of heart disease present different techniques for performing a medical set of data. Basic attributes were recognized with a correlation of Pearson, Lasso Regularization, and Elimination of Recursive Features by chosen basic attributes that they investigated by algorithms of Random Tree, developed, M5P classifiers, Random Forest, Reduced Error Pruning in heart disease. Due to the results, they recognize developed method of the Random Forest ensemble presents better accuracy in comparison with the other.

Abdeldjouad et al. [12], in this study, it had been presented a novel hybrid technique for predicting cardiovascular disease by utilizing various methods of ML-like Adaptive Boosting (AdaBoostM1), Fuzzy Hybrid Genetic Based Machine Learning (FH-GBML), Fuzzy Unordered Rule Induction (FURIA), Genetic Fuzzy System-LogitBoost (GFS-LB), Logistic Regression (LR), Multiobjective Evolutionary Fuzzy Classifier (MOEFC). Each classifier's results and accuracy have been compared with the best classifier which is selected for the more appropriate prediction of cardiovascular. To achieve this aim, they utilize 2 free software (Keel and Weka).

Singh et al. [13], in this study, it had been presented a novel technique after using 2 approaches. The technique assessed ML repository datasets UCI also targeted was predicting if the person had a disease heart/not on the features such as heartbeat, FBS, blood pressure, expand as well as the others by better accuracy rather than the other approaches. At first, they train logistic regression by whole features, after that they train logistic regression by powerful predictive features of power involved in part of eda, lastly, they train logistic regression after eliminating the least basic features. Second, they are using SVM. After that, they presented the design and logistic regression with a basic analysis of elements.

Shorewala et al. [14], in this study, it had been analyze ML effectiveness to predict disease of coronary heart. At first, analytics of data showed the patterns in basic attributes and data for binary classification of logistics. In addition to the k-NN, statistical technique played a crucial role also let for the efficient selection of features from a set of data.

3. Proposed Method

The study illustrates different ML algorithms analysis, algorithms which are utilized in this study are Random Forest Classifiers, Logistic Regression, and K nearest neighbors (KNN) that can be effective for medical analysts/ practitioners for correctly diagnosing the disease of Heart. The method is the process that involves stages that transform according to data in identified patterns of data for the user's knowledge the proposed approach was divided into five phases (Data gathering, Data Pre-processing, Fast Correlation-Based Feature selection, MWOA based Feature Optimization and Classification) the second step this extracts the important amounts and the third is the step of preprocessing in which explore data. Preprocessing of data copes with lost amounts, normalization, and data cleaning based on the algorithms utilized. After data pre-processing, the modified whale optimization algorithm evaluation version named MWOA was discussed for feature extraction of software usability. After optimization of the feature, the classifier is utilized for classifying pre-processed data that the classifier utilized in the presented model including Random Forest Classifier, KNN, and Logistic Regression. At last, the presented model is undertaken and we assessed our design on a performance and accuracy basis by utilizing different metrics of performance. In this study, the efficient prediction system of heart disease has been improved by applying various classifiers. The presented technique is illustrated in the following flowchart:

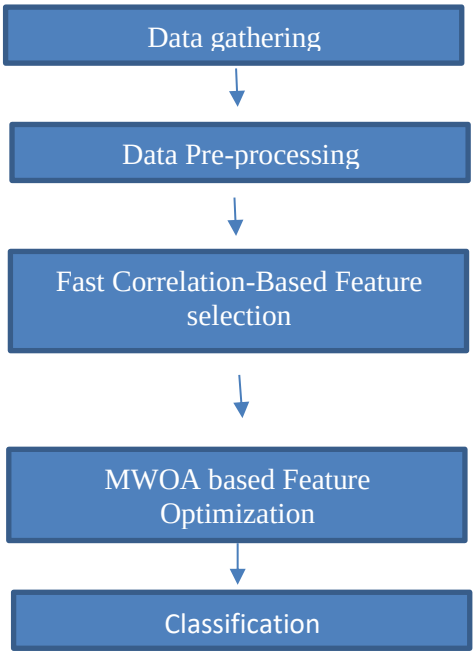

Figure 1. The general architecture of the proposed method

3.1 Data Gathering

The dataset is built by integrating 5 various sets of data (Hungary, Statlog, Cleveland, VA Long Beach, and Switzerland). It is involved in the framework.

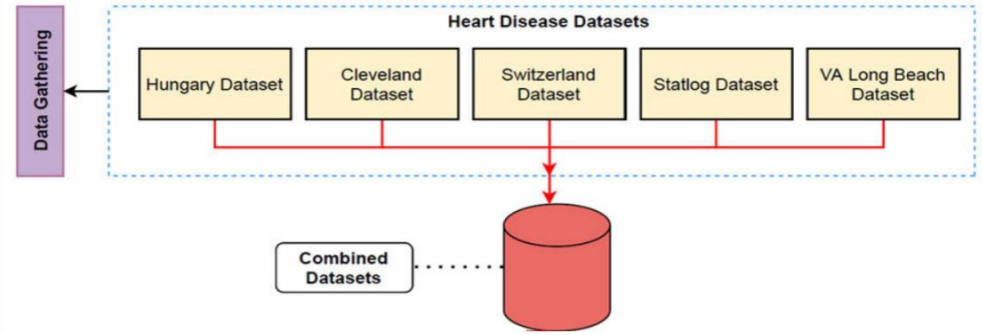

Figure 2. Data aggregation

3.2 Dataset Preprocessing

Dataset Preprocessing transforms raw data/encodes data into the form that can be interpreted simply with an algorithm. Methods of preprocessing utilized in the proposed work include:

1. Data cleaning is eliminated via the processes like filling in lost amounts, therefore, resolving data inconsistencies .
2. Data integration refers to the process of combining data from multiple sources into a unified, coherent format that can be used for various business.
3. Data transformation is the process of converting, cleansing, and structuring data into a usable format that can be analyzed to support decision making processes.
4. Data Reduction is the process in which an organization sets out to limit the amount of data it's storing .
5. Data Discretization the process of grouping continuous values of variables into contiguous intervals.

6. Data normalization the organization of data to appear similar across all records and fields.

3.3 Feature Selection

Is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction. Heart disease dataset, features number can achieve thousands of tens; a dataset of heart disease has 14 features. As numeric huge irrelevant number, as well as extra features, are included in such data of expression, the classification task of heart disease is made more complicated. If the full data are utilized for performing the classification of heart disease, the accuracy will not be appropriate, and costs, as well as the time of computation, will be high. So, the selection of features, as the stage of pre-treatment for the learning of machines, removes data unresolved, decreased sizing, raises the accuracy of learning, and develops results understanding. Present increment in data dimensionality includes important issues for choosing feature techniques according to effectiveness and efficiency. Feature selection's reliable method based on Fast Correlation is chosen for selecting the discriminatory features subset before the classification, with removing features by no /little effect, Feature selection based on Fast Correlation presents the great performance by complete feature correlation and redundancy consideration. The data were standardized, and attributes were selected using feature selection based on fast correlation.

3.4 Modified WOA for Usability Feature Selection

In WOA, mostly as an updated solution has relied on the present optimum solution of the candidate, we defined the weight of inertial ω in [0, 1] in WOA for achieving the Modified whale optimization algorithm (MWOA). MWOA algorithm stages include:

1. Whale’s population is initialized. The whale's position is initialized like each amount in the dataset also the inertia weight is multiplied. A score of the leader as well as the position of the Leader is initialized.
2. Every fitness feature or whale is investigated by utilizing the function of fitness as well and associating whales are added to the array. “a” is linearly reduced.
3. Whole constants are initialized.
4. Every whale's position is changed based on equations that are stated in the study. Accuracy is computed for every iteration also whole chosen whales are saved.

Modified equations of mathematics are presented below:

1. Search factors' positions are updated as below to encircle the mechanism of prey:

$$V = \left|C \cdot \omega Z^*(t) - Z(t)\right| \\ Z(t+1) = \omega Z^*(t) - A \cdot V \quad (1)$$

2. The search factors' position is updated as below in the spiral position of updating:

$$Z(t+1) = V' \cdot e^{bl} \cdot cos(2\pi l) + \omega Z^*(t) \quad (2)$$

MWOA pursues a similar way to select either of two mechanisms for hunting prey however by the updated equations of mathematics. We defined the variable p that depends on [0,1] that decides finally on the mechanism for being implemented. equations are as below:

$$Z(t+1) = \omega Z^*(t) - A \cdot V \quad if\ p < 0.5 \\ Z(t+1) = V' \cdot e^{bl} \cdot cos(2\pi l) + \omega Z^*(t) \quad if\ p \geq 0.5 \quad (3)$$

However, randomly preying humpback whales search, in addition to the method of bubble-net, this is important to notice that defined inertial weight stays stable for whole factors of search in each iteration. A MWOA has been presented for a selection of features.

3.5. Classification

The features of classification is presented as input to different algorithms of ML like techniques of Logistic Regression, Naive Bayes classification, Random Forest, and Decision Tree. The dataset of input is shared in 80 percent dataset of training as well as the remaining 20 percent in a dataset of the test. A dataset is a set of data that is utilized for training the design. Dataset of Testing is applied to check the trained model performance. Performance is calculated for every algorithm also it is analyzed according to various metrics utilized like F-measure scores, recall, accuracy, and precision as defined later. Various algorithms explored are as follows.

1. Random Forest

Random Forest is a supervised learning Algorithms of Forest are applied to classify and for regression. This makes a tree for data and also provides the predictions according to the tree. The learning algorithm of Random Forest can be applied to huge sets of data also it can create similar results while huge sets of record amounts are lost. Created instances from the tree of the decision can be stored and can also be utilized on the other data. There are 2 steps in a random forest, first, generate the random forest and predict by utilizing the classifier of the random forest generated in the first step.

2. Decision Tree

The Decision Algorithm The tree is shaped like a flowchart, with branches on the outside representing the results and nodes on the inside providing dataset attributes. The selection of the Tree of Decision is based on its dependability, simplicity, speed, and minimal requirements for data preparation. In the Tree of Decision, the class label prediction is initialized from the tree root. The amount of the root feature is compared to the record's feature. In addition to jumping to the next node based on the comparison result, the relevant branch is followed to the amount.

3. Logistic Regression

One classification approach that is frequently used for binary classification problems is a regression of logistic. Rather of directly fitting a hyperplane or line, the logistic regression algorithm uses the logistic function to squeeze the output of a linear equation between 0 and 1. About logistic regression. Because it has thirteen independent variables, logistic regression works well for classification.

4. Naive Bayes

The Naïve Bayes algorithm adheres to the Bayes rule. The most fundamental presumption and crucial factor in categorization is the independence of dataset features. Predicting and holding the best while maintaining the independence assumption is quick and simple with this method. A theorem of Bayes computes later event (A) probability according to several basic event B probabilities which are shown with P(A/B) as illustrated below:

P(A|B) = (P(B|A) P(A)) / P(B)

5. Vector machine (SVM)

Let instances of training have dataset Data = {y_i, x_i}; i = 1, 2, . . . , n that $x_i \in R^n$ provide a vector of i th, $y_i \in R^n$ present target case. linear SVM recognizes optimum form f (x) = w^T x + b hyperplane that w is the dimensional vector of coefficient also b is the offset. it is performed by solving subsequent issue optimization.

$$Min_{w,b,\xi_i} \frac{1}{2}w^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t.\ y_i\left(w^T x_i + b\right) \geq 1 - \xi_i, \xi_i \geq 0, \quad \forall_i \in \{1, 2, \ldots, m\} \tag{4}$$

6. K-Nearest Neighbor (K-NN)

Algorithm of K -nearest neighbors is the supervised algorithm of classification. That group's things are based on the nearest neighbor. This is learning based on instance. Attribute distance computation shape neighbors are assessed by applying the distance of Euclidean. This utilizes the named points set and also utilizes them on the way to marking the other point. Data are organized according to the similarity between them, also this is feasible for filling lost data amounts by utilizing K-NN. Different methods of prediction are used for the set of data since the lost amounts are filled. This is feasible for gaining better accuracy by using different algorithm combinations. The K-NN algorithm is simple to use and doesn't require any design work or additional assumptions. The approach is adaptable and can be used for classification, regression, and search. Features that are irrelevant and noisy impair accuracy Even so, the simplest algorithm is K-NN.

7. MLP Classifier

Multilayer Perception (MLP) is a supervised multilayer feedforward artificial neural network model. This determines input; it has an output layer for prediction and an input layer for obtaining attributes. The middle layer is referred to as the concealed layer. Each hidden layer neural network uses the activation function to modify input in addition to taking the final layer's weighted input total. In a nonlinear path, the activation function maps the

relationships between input and output. Every attribute input is determined by the MLP classifier according to the class level.

4. Results and Discussions

First, the dataset used in the research will be introduced and the existing variables will be explained one by one. The data used in this study were extracted from the database of California University[1]. This database is called heart disease, which has five identical sub-datasets related to heart patients called Hungarian data, Cleveland data, VA data, Switzerland data, and Stat log. The difference between these datasets is that they are each collected by a specific institution. Each of these datasets has four grades of disease for patients and each of which refers to the severity of heart disease. The number of records in this dataset is described in Table 1.

Table 1- The distribution of class values in four datasets

Database name	Total records
Cleveland	**303**
Hungarian	**294**
Switzerland	**123**
Long Beach VA	**200**
Statlog	**270**

According to the table above, the Cleveland dataset has more records. Since numerous articles have used only this dataset for modeling in their research, moreover, other datasets have large missing values for some variables, and therefore this dataset is chosen for modeling in this research. We now describe the variables in this dataset.

The Cleveland dataset has 303 records and 76 variables and only 14 variables can be used among them (13 conditional variables and one decision variable), which exist in the published dataset. All 14 variables are described below:

[1] http://archive.ics.uci.edu/ml/datasets/Heart+Disease

- Age of the patient (age): This variable is continuous and related to the patient's age at the time of hospitalization.
- Patient gender (sex): This variable represents the patient's gender and has two values of 0, which means female, and 1 indicates a male.
- Type of chest Pain: This variable is used to determine the type of pain that occurs in heart patients. Values associated with this variable are 1) prevalent angina, 2) non-prevalent angina, 3) non-angina pain, and 4) asymptomatic.
- Steady blood pressure: This variable is a continuous variable that is stored in the patient information file at hospital admission.
- Cholesterol: The cholesterol variable is a continuous variable that measures the amount of cholesterol in the blood.
- Fasting blood sugar: This variable indicates the patient's fasting blood sugar. If this value is more than 120 ml / db, the value is 1, otherwise, it equals 0.
- Electrocardiographic results: This variable has three values: 0) normal, 1) abnormal, and 2) left ventricular magnitude.
- Heart Rate: The maximum heart rate that a patient has experienced is recorded in this continuous numerical variable.
- Exercise-induced angina: This variable is a two-value variable in which a value of 1 indicates yes and 0 implies no.
- Depression: This variable is a continuous variable that represents the amount of anxiety caused by exercise associated with rest.
- Slope: This variable has three values: 1) upward slope, 2) smooth and without slope, and 3) downward slope.
- Number of vessels (Ca): The number of main vessels that have three values of 1, 2, and 3.
- Cardiac status (Thal): This variable has three values: 3= normal, 6= permanent defects, and 7= reversible defects.
- The decision or class variable: This variable is related to the prediction class label, which has 4 values. 0 means absence, heart disease, and values 1 to 4 show heart disease with different intensities.

4.1. The parameters' initialization

To evaluate the quality of the proposed algorithm, we adjusted the parameters according to the parameters of the base paper [6]. In such a way that 80% of the data were considered for training and 20% of the data were considered for testing. Table 2 shows the settings for the parameters of the proposed method.

Table 2 Initial values for parameters in the proposed method

Parameters	Values
Training dataset	80%
Testing dataset	20%
Number of search agents	5
Maximum number of iterations	100
Omega	0.8

4.2 Evaluation Measures

Several Measures are used to evaluate classifications and the manner of their performance in predicting classes' labels. These criteria include precision (detection), sensitivity, Specificity, accuracy, and F-Measure.

Accuracy: is the proportion of correct prediction to the total sample. The higher the score, the greater the accuracy of the model's predictions

.

$$\text{Accuracy} = \left(\frac{\text{Correct predictions}}{\text{Total predictions}}\right)$$

Sensitivity: this is the parameter of performance that measures the system's ability for making the appropriate positive predictions.

$$\text{Sensitivity} = \left(\frac{\text{True positives}}{\text{True positives} + \text{false negatives}}\right)$$

Specificity: is the rate to of identified negative samples to all negative samples indicating the model's ability to identify negative sample.

$$\text{Specificity} = \left(\frac{\text{True negatives}}{\text{True negatives} + \text{false positives}} \right)$$

Precision: Is the ratio of true positive Precision sample to all positive predictive samples. The higher, the score the greater the number of positive samples predicted by the model to be true.

$$\text{Precision} = \left(\frac{\text{True positives}}{\text{True positives} + \text{false positives}} \right)$$

F-Measure: F-Measure integrates precision and sensitivity results by applying the mean of harmonic.

$$\text{F - Measure} = 2 \times \frac{\text{Sensitivity} \times \text{precision}}{\text{Sensitivity} + \text{precision}}.$$

The gmean is also defined as follows:

$$\text{gmean} = \sqrt{\text{sensitivity} \times \text{specificity}} \times 100$$

In addition to the mentioned criteria, the terms negative (N), positive (P), true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are summarized in a matrix called the confusion matrix, which is shown in the figure below.

Table 2. The confusion matrix

		Predicted class		
		yes	no	total
Real class	Yes	TP	FN	P
	No	FP	TN	N
	Total	P’	N’	P+N

4.3.The results' evaluation

We implemented the tests by the issue of dataset named the issue of heart disease: 5-class. We utilized 20 percent data as the data for testing for the issue of 5-class. It is because of reality that design requires much more data for the present problem because of huger facilities number amount in output. The technique here illustrates that the presented proposed technique can present the results that can be compared with the other articles. Results Comparison by applying 3 various kinds is illustrated in Table 5. In comparison with present techniques as well as the results of the test, we recognized that the optimized model of us outperforms rather than other schemes in the classification and prediction of heart disease. We take advantage of the selected features of FCBF given the algorithm of MWOA. Therefore, getting a higher accuracy of classification rather than present schemes. The presented optimized scheme with MWOA and FCBF obtained an 83.26 percent score of accuracy by regression of Logistic.

Table 3- The comparison results of the proposed method with the basic paper for the 5-class problem

Method	Accuracy Test
MP-SVM Random forest [6]	73.33
KNN, random forest classifier, and SVM [14]	70.00
Proposed method	83.26

Table 4- The results of evaluation criteria for the proposed method

Method	Accuracy	sensitivity	specificity	precision	gmean	F-score
RandomForest	81.97	76.36	86.99	84	81.51	80
DecisionTree	82.40	79.09	85.37	82.86	82.17	80.93
LogisticRegression	83.26	80	86.18	83.81	83.03	81.86
NaiveBayes	80.69	79.09	82.11	79.82	80.59	79.45

SVM	82.83	76.36	88.62	85.71	82.26	80.77
KNN	81.97	89.09	75.61	76.56	82.07	82.35
MLP	80.29	80.91	79.68	78.07	80.29	79.46

Conclusion

Heart disease death's basic cause is the detection delay of it. For reducing death, the study have presented the hybrid decision support system for heart disease. This study's basic contribution is proposing the support system of optimized decisions to detect disease heart with greater accuracy in comparison with present systems. Either this is the step of selection of classifier, selection of feature, or missing value; the study has recognized the best algorithms via the simulation also they utilized them in presenting a hybrid support system of decision. For comparing and testing the proposed system, the project has utilized 5 sets of data in the simulated area improved by utilizing Matlab. This has illustrated the better performance associated with the other hybrid support systems of decision recognized in the literature. Logistic Regression had given the best 83.26 percent accuracy. The proposed system is not a substitute for a doctor, this can be utilized in rural as well as far regions where the heart specialist doctor /other modern facilities of medical are not accessible. In addition, this can help the doctor in making fast decisions.

References

[1] Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. "Heart disease prediction using machine learning techniques." *SN Computer Science* 1, no. 6 (2020): 1-6.

[2] Fitriyani, Norma Latif, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. "HDPM: an effective heart disease prediction model for a clinical decision support system." *IEEE Access* 8 (2020): 133034-133050.

[3] Ghosh, Pronab, Sami Azam, Mirjam Jonkman, Asif Karim, FM Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, and Friso De Boer. "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques." *IEEE Access* 9 (2021): 19304-19326.

[4] Jain, Rishabh, Deepak Gupta, and Ashish Khanna. "Usability feature optimization using MWOA." In *International conference on innovative computing and communications*, pp. 453-462. Springer, Singapore, 2019.

[5] Khourdifi, Youness, and Mohamed Bahaj. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony

optimization." *International Journal of Intelligent Engineering and Systems* 12, no. 1 (2019): 242-252.

[6] Rajdhan, Apurb, Avi Agarwal, Milan Sai, Dundigalla Ravi, and Poonam Ghuli. "Heart disease prediction using machine learning." *International Journal of Research and Technology* 9, no. 04 (2020): 659-662.

[7] Bharti, Rohit, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. "Prediction of heart disease using a combination of machine learning and deep learning." *Computational Intelligence and Neuroscience* 2021 (2021).

[8] Jindal, Harshit, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath. "Heart disease prediction using machine learning algorithms." In *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012072. IOP Publishing, 2021.

[9] Rani, Pooja, Rajneesh Kumar, Nada MO Sid Ahmed, and Anurag Jain. "A decision support system for heart disease prediction based upon machine learning." *Journal of Reliable Intelligent Environments* (2021): 1-13.

[10] Sowmiya, C., and P. Sumitra. "A hybrid approach for mortality prediction for heart patients using ACO-HKNN." *Journal of Ambient Intelligence and Humanized Computing* 12, no. 5 (2021): 5405-5412.

[11] Yadav, DHYAN CHANDRA, and S. A. U. R. A. B. H. Pal. "Prediction of heart disease using feature selection and random forest ensemble method." *International Journal of Pharmaceutical Research* 12, no. 4 (2020): 56-66.

[12] Abdeldjouad, Fatma Zahra, Menaouer Brahami, and Nada Matta. "A hybrid approach for heart disease diagnosis and prediction using machine learning techniques." In *International conference on smart homes and health telematics*, pp. 299-306. Springer, Cham, 2020.

[13] Singh, Archana, and Rakesh Kumar. "Heart disease prediction using machine learning algorithms." In 2020 international conference on electrical and electronics engineering (ICE3), pp. 452-457. IEEE, 2020.

[14] Shorewala, Vardhan. "Early detection of coronary heart disease using ensemble techniques." Informatics in Medicine Unlocked 26 (2021): 100655.