

Research Article

Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification

Ashir Javeed,¹ Sanam Shahla Rizvi ,² Shijie Zhou,¹ Rabia Riaz,³ Shafqat Ullah Khan,⁴ and Se Jin Kwon ⁵

¹School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China

²Raptor Interactive (Pty) Ltd., Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa

³Department of CS&IT, University of Azad Jammu and Kashmir, Muzaaffarabad 13100, Pakistan

⁴Department of Electronics, University of Buner, Buner, Pakistan

⁵Department of Computer Engineering, Kangwon National University, Samcheok 25806, Republic of Korea

Correspondence should be addressed to Se Jin Kwon; sjkwon@kangwon.ac.kr

Received 16 March 2020; Revised 4 May 2020; Accepted 29 July 2020; Published 26 August 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Ashir Javeed et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diagnosis of heart disease is a difficult job, and researchers have designed various intelligent diagnostic systems for improved heart disease diagnosis. However, low heart disease prediction accuracy is still a problem in these systems. For better heart risk prediction accuracy, we propose a feature selection method that uses a floating window with adaptive size for feature elimination (FWAFE). After the feature elimination, two kinds of classification frameworks are utilized, i.e., artificial neural network (ANN) and deep neural network (DNN). Thus, two types of hybrid diagnostic systems are proposed in this paper, i.e., FWAFE-ANN and FWAFE-DNN. Experiments are performed to assess the effectiveness of the proposed methods on a dataset collected from Cleveland online heart disease database. The strength of the proposed methods is appraised against accuracy, sensitivity, specificity, Matthews correlation coefficient (MCC), and receiver operating characteristics (ROC) curve. Experimental outcomes confirm that the proposed models outperformed eighteen other proposed methods in the past, which attained accuracies in the range of 50.00–91.83%. Moreover, the performance of the proposed models is impressive as compared with that of the other state-of-the-art machine learning techniques for heart disease diagnosis. Furthermore, the proposed systems can help the physicians to make accurate decisions while diagnosing heart disease.

1. Introduction

The heart is a vital organ in the human body that is liable for blood circulation. The heart is responsible for oxygen and energy supply to all organs of the body including itself. Heart disease causes the abnormal blood circulation in the body that might be fatal for human life. Hence, if the heart stops its normal functionality, the whole system will be dead. From the literature, various risk factors are identified that cause the heart disease. The risk factors of heart diseases are classified into two major types such as the risk factors that can alter,

e.g., smoking and physical exercise, and the risk factors that cannot alter, e.g., gender, age, and patient's family history [1]. The diagnosis of heart through conventional medical methods is quite difficult, complex, time consuming, and costly. Therefore, the diagnosis of heart disease is worst in developing countries due to lack of state-of-the-art examination tools and medical experts [2, 3]. Additionally, the invasive medical procedure for examination of heart failure is formed on various tests suggested by physicians, after studying the medical history of the patient and analyzing the relevant symptoms [4]. Angiography is considered as the

gold standard among the medical tests for diagnosis of heart failure. Heart disease cases are affirmed through angiography as it is the best practice for diagnosis of heart disease. Moreover, angiography has side effects as well as higher cost for diagnosis of heart disease and demands extraordinary technical expertise [5, 6]. Therefore, machine learning and data mining techniques are needed to design the expert systems for resolving the problems of angiography.

To address the abovementioned problems, researchers have designed different noninvasive diagnosis systems by exploiting machine learning based predictive models. These models include logistic regression, naive Bayes, *k*-nearest neighbor (KNN), decision tree, support vector machine (SVM), artificial neural network (ANN), and ensembles of ANN for heart failure disease classification [1, 7–18]. Robert Detrano utilized logistic regression for heart failure risk prediction and attained classification accuracy of 77%. Newton Cheung utilized various predictive models consisting C4.5, naive Bayes, BNND, and BNNF algorithms. The accuracies of proposed algorithms were 81.11%, 81.48%, 81.11%, and 80.95%, respectively, for precise classification of patients and healthy subjects. A. Khemphila and V. Boonijjing proposed a classification technique based on multilayer perceptron (MLP) in addition to back-propagation learning algorithm and biomedical test values for diagnosing the heart disease through a feature selection algorithm. Information gain is utilized to filter features through elimination of the features which do not contribute for precise results. Total number of thirteen features is reduced to eight by using a feature selection algorithm. For the classification, ANN is used as a classifier. The accuracy of training dataset was 89.56%, while for data validation, the accuracy of 80.99% was reported.

Recently, Paul et al. proposed a fuzzy decision support system (FDSS) in order to detect the heart disease [19]. They proposed a genetic algorithm based on FDSS that has five key components such as preprocessing of the dataset, effective features selection through diverse methods, weighted fuzzy rules that are set up through genetic algorithm, generated fuzzy knowledge used to build FDSS, and heart disease prediction. The proposed system obtained the accuracy of 80%. Verma et al. proposed a hybrid model for coronary artery disease (CAD) diagnosis [20]. The proposed method consists of jeopardizing factor identifiers adopting a correlation based subset (CFS) selection with particle swarm optimization (PSO) search model and *K*-means. Supervised learning algorithms such as multilayer perceptron (MLP), multinomial logistic regression (MLR), fuzzy unordered rule induction algorithm (FURIA), and C4.5 are then utilized to design CAD cases. The accuracy of the proposed approach was 88.4%. The proposed model enhanced the efficiency of classification techniques from 8.3% to 11.4% of Cleveland dataset. Shah et al. proposed a technique based on the feature extraction for reducing feature dimensions [21]. The proposed approach used probabilistic principal component analysis (PPCA). Projection dimensions are extracted through PPCA that compliments high covariance and also helps to eliminate feature dimension. Parallel analysis (PA) helps in the selection of projection vectors. The feature

subset of reduce feature vector is input to the radial basis function (RBF) kernel-based support vector machines (SVMs). Two types of classification are categories into heart patient (HP) and normal subject (NS) through RBF-based SVM serves. The proposed model is tested against accuracy, specificity, and sensitivity on the datasets of UCI, i.e., Cleveland. The accuracy of the proposed model for Cleveland dataset was 82.18%, 85.82%, and 91.30%, respectively.

Most recently, Dwivedi tests the performance of different machine learning methods for the prediction of heart disease. The highest classification accuracy of 85% was reported based on logistic regression [22]. Amin et al. evaluate the different data mining methods and identify the significant features for predicting heart disease [23]. Predictive models were built from different combinations of features and well-known classification methods, e.g., LR, SVM, and *K*-NN. From experimental results, it was studied that the best performance of the data mining technique for classification accuracy was 87.4% for the heart disease prediction. Özşen and Güneş proposed an expert system developed from an artificial immune system (AIS) and achieved accuracy of 87% [24]. An expert system was proposed by Özşen and Güneş based on the artificial immune system (AIS). The accuracy of 87% was reported for the developed expert system. Polat et al. developed another similar system and obtained 84.5% accuracy [25]. Das et al. utilized a neural network ensemble model with the purpose of improving classification accuracy. His ensemble model obtained the classification accuracy of 89.01% [1]. Recently, Samuel et al. proposed a diagnostic system developed from ANN and Fuzzy AHP. The prediction accuracy of 91.10% was reported from the ANN and Fuzzy AHP diagnosis system [4].

As clear from the literature survey, ANN-based diagnostic systems have shown better performance on the heart disease data. Hence, we also attempt to design a diagnostic system based on neural network for heart disease detection. The development of various noninvasive diagnostic systems for heart disease detection motivates us to design an expert diagnostic system based on neural networks. From the empirical result, it is analyzed that proposed model shows promising performance. Hence, it can be used in clinics to make accurate decisions while diagnosing heart failure.

2. Materials and Methods

In previous studies, researchers used feature sets without eliminating irrelevant or noisy features. In this study, we propose a novel feature elimination method for removing noisy or irrelevant feature vectors and thus selected an optimal subset of feature vectors before feeding them to ANN or DNN. The proposed algorithm uses a window with adaptive size. The window size is initialized from one and is placed at the first feature of the feature vector. The feature or features to which the window points are eliminated while the remaining features constitute the subset of features that are supplied to the neural network for classification. To find the optimal configuration of the neural network for the subset of features, grid search algorithm is used. It is noteworthy that the previous studies utilized conventional ANN with only

one hidden layer for heart failure detection problem. However, in this study, we found out that deep neural networks with more than one hidden layers and trained with new learning algorithms show better performance. Additionally, this study evaluates the feasibility of features selection algorithm at the input level of DNN. The working of the proposed diagnostic system is clearly shown in Algorithm 1 and Figure 1.

2.1. Dataset Description. For this research, an online repository of machine learning and data mining from University of California, Irvine (UCI), for the heart disease dataset was used that is known as a Cleveland heart disease database. Data were gathered from the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation by Dr. Robert Detrano [26]. The dataset is comprised of 303 subjects. Furthermore, the number of subjects having missing values in the dataset is 6. In the dataset, 297 subjects have complete data values out of 303 subjects. Hence, the number of subjects that have complete data values is used for experiments. Moreover, each subject in dataset has 76 raw features. In the previous work, the researchers mostly used 13 prominent features out of 76 raw features of each subject for the diagnosis of heart disease. Therefore, mostly used 13 features for diagnosis of heart failure is considered for this study. Table 1 depicts the most commonly used 13 features of heart disease.

2.2. The Proposed Method. The proposed diagnostic system has two main components that are hybridized as one black-box model. The main reason for hybridizing the two components into one block is that they work in connection with each other. The first component of the system is a feature selection module, while the second component is a predictive model. Feature selection methods use data mining concepts to improve the performance of the machine learning models [27, 28]. The feature selection module uses a search strategy to find out the optimal subset of features which are applied to the DNN that acts as a predictive model. The feature selection module uses a window that scans the feature vector. The working of the proposed method can be depicted from the algorithm.

Initially, the size of the window is set to 1. And, the window is placed at the left most side of the feature vector with size n , i.e., having n number of features. Hence, initially, the feature is eliminated from the feature vector on that the window is placed and the remaining features constitute the subset of features which are supplied to the DNN for classification. The performance of the subset of features is saved. In the next step, the window floats towards the right direction. Again, that feature is eliminated on which the window is placed and the remaining features constitute feature subset whose performance is checked by the DNN model. The same process is repeated until the window reaches the last feature, i.e., the n^{th} feature. With this, the first round of window floating is completed. It is important to note that the features subset size is $n - 1$ in the first round.

In the next round, the window size is updated to 2. Hence, in this round, the window points towards the two features at a

time. Again, the window starts the floating process from the left most side of the feature vector and eliminates the first two features. The remaining $n - 2$ features constitute the features subset that is applied as the input to the DNN model for classification, and the results are compared with the best performance achieved on the previous subset of features. If the performance is better than the previous best performance, the best performance and optimal subset of features is updated. In the next iteration, the window floats towards the right direction and those two features are eliminated on which the window is placed. The remaining features constitute the subset of features which are applied to DNN. The same process is repeated until the window reaches to the right most side of the feature vector. This marks the end of the second round. In the third round, the window size is made 3 and the same process is repeated that was carried out for the first two rounds. Finally, at the $n - 1$ round, the window size is made $n - 1$. In this round, the window can float just once towards the right. And then, the whole process is ended. Finally, the subset of features that give us the best results is declared as the optimal subset of features. The whole process of features selection through adaptive floating window is clearly illustrated in Figure 2. Each time a subset of features is supplied to the DNN, the DNN architecture is optimized using the grid search algorithm. The performance of a DNN is highly dependent on its architecture [29]. Inappropriate DNN architecture will result in poor performance although there are chances that the DNN is applied with an optimal subset of features. The main reason for such a poor performance is that, if the DNN architecture selected for the classification is with insufficient capacity, then it will result in underfitting [30, 31]. In such a case, the DNN will show poor performance on both data, i.e., training data and testing data. However, if the DNN architecture has excessive capacity, it will overfit to the training data; thus, it will show better performance on the training data but poor performance on the testing data. Hence, we need to search optimal architecture of DNN that will show good performance on both testing and training data. To understand the relationship between DNN architecture and the capacity of DNN, we need to understand the formulation of DNN. The neural network is formulated as follows:

Neural networks are generated by the computational system based on mathematical models that simulate the human brain. The key element in the neural network model is known as perceptron or a node [32]. Nodes are shaped into groups which are called layers. Artificial neurons work on the same principal which is followed by the biological neuron. As an artificial neuron receives one or more inputs from the adjoined neurons, it then processes the information and transfers the output to the next perceptron. Artificial neurons are connected through a link that is known as weights. The input information χ_i is weighted either positive or negative during the computation of output. An internal threshold value ρ and weights are assigned for the solution of a problem under consideration. On every node, the result is calculated by multiplying the input values χ_n and associated weight w_n that is fine-tuned by the threshold value ρ . The output is then calculated through an activation function or transfer function (α) and is given in the following equation:

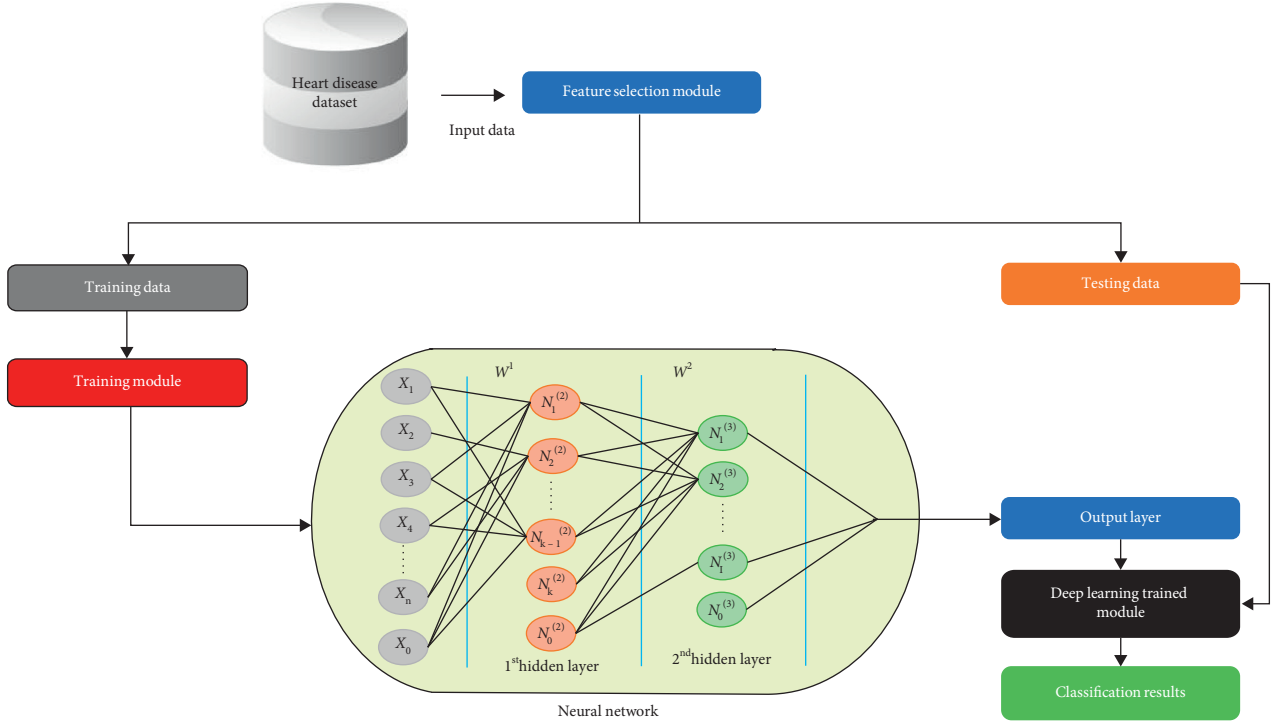


FIGURE 1: Block diagram of the newly proposed method.

Input: $\{N_F: \text{Features size}, \beta: \text{Hyperparameters}\}$
Output: $\{F_s: \text{Optimal Subset of features}, \beta_o: \text{Optimal}\}$
 (1) $W; 1, 2, 3, \dots, F$, where W is the window size
 (2) $\text{Best_Acc} = 0$
 (3) Initialize
 (4) **for** $1 = W_{\max}$
 (5) **for** $1 = \beta_{\max}$
 (6) Acc using F_s and β
 (7) **if** $(\text{Acc} > \text{Best_Acc})$
 Begin if
 $\text{Best_Acc} = \text{Acc}$
 Note down F_s and β as β_o
 End if
 (9) **END for**
 (10) **END for**
 (11) Display F_s, β_o and Best_Acc

ALGORITHM 1: FWAFE-DNN.

$$\delta_i = \alpha\left(\sum \omega_n \cdot \chi_n - \varrho\right). \quad (1)$$

The transfer can be linear or nonlinear. In the case of nonlinear function tangent, hyperbolic or radial basis form is applied. The sigmoid function, $\alpha(\delta_i)$, is done at the following layer as an output value (equation (2)). γ is related to the shape of the sigmoid function. The increase in parameter γ value strengthened the nonlinearity of the sigmoid function:

$$\alpha(\delta_i) = \frac{1}{[1 + \exp(-\gamma\delta_i)]}. \quad (2)$$

The neural network is obtained by connecting the artificial neurons. If the constructed neural network model has

only one hidden layer, we name it ANN [17]. However, if the constructed neural network model has more than one hidden layer, we name it DNN [17].

3. Validation Scheme and Evaluation Metrics

3.1. Validation Scheme. In earlier works, the performance of the expert diagnosis systems has been evaluated through holdout validation schemes. The dataset has to be partitioned into two parts: one is for training purpose, while another is used for testing. In the past, researchers have been using various train-test split percentages of data partitioning. Furthermore, Das et al. in [1] and Paul et al. in [33] used

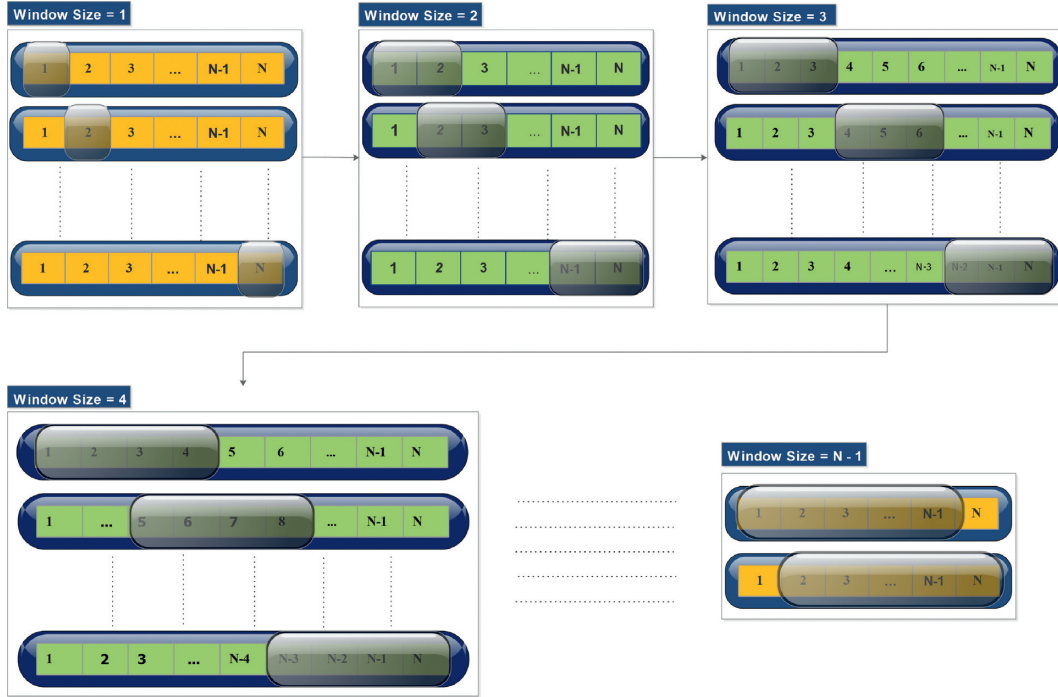


FIGURE 2: Feature selection by eliminating features using floating window with adaptive size.

holdout validation schemes in their research. They have partitioned the dataset into 70%–30% ratio, where 70% of the data set is utilized for the training purpose of the predictive model while 30% of the dataset is utilized for testing the performance of the predictive model. Therefore, we also utilized the same criteria of data partition for train-test purpose.

3.2. Evaluation Metrics. Various evaluation metrics such as specificity, sensitivity, accuracy, and Matthews correlation coefficient (MCC) are utilized for evaluating the performance and efficiency of the proposed model. The percentage of the precisely classified subjects is known as accuracy. Sensitivity is the accurate classification of the patients, whereas specificity is the absolute classification of healthy subjects. All the evaluation metrics are formulated in equations (3)–(6):

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

where TP stands for true positives, TN describes true negatives, FP shows false positives, and FN stands for false negatives.

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad (4)$$

$$\text{specificity} = \frac{TN}{TN + FP}, \quad (5)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (6)$$

The characteristic of binary classification is assessed using MCC for machine learning and statistics. The value of MCC is ranging between -1 and 1 . The -1 value of MCC denotes the total conflict between prediction and observation, whereas 1 shows the exact prediction, while 0 describes the classification as random prediction. Moreover, in this study, another evaluation metric, namely, the receiver operating characteristic (ROC) curve was also exploited. The ROC curve is a well-known metrics that is used to statistically evaluate the quality of a predictive model. The ROC curve provides area under the curve (AUC); thus, a model is considered a better model if its AUC is high.

4. Experimental Results and Discussion

In this session, two kinds of diagnostic systems are proposed. Moreover, experiments are done to test the performance of the proposed diagnostic system. In the first experiment, FWAFE-ANN is developed and stimulated, while in the second experiment, FWAFE-DNN is utilized. In the first experiment, the FWAFE algorithm is used to construct a subset of features. Furthermore, a subset of features is applied to ANN that is used as a predictive model. In the second experiment, FWAFE is used to construct a subset of features, whereas DNN is utilized for classification. All the experiments were simulated by using Python programming software package.

4.1. Experiment No. 1: Feature Selection by FWAFE and Classification by ANN. In this experiment, at the first stage, FWAFE is used, while in the second stage, ANN is used. The feature selection module eliminates noisy and irrelevant features by exploiting a search strategy, whereas the second model is deployed as a predictive model. The proposed

TABLE 1: Types of features of the dataset.

Feature no.	Feature description	Feature code	(Mean \pm std) _{Healthy}	(Mean \pm std) _{Patients}
1	Age (AGE)	FC ₁	52.64 \pm 9.52	56.84 \pm 7.42
2	Sex (SEX)	FC ₂	0.55 \pm 0.49	0.81 \pm 0.385
3	Chest pain type (CPT)	FC ₃	2.79 \pm 0.92	3.60 \pm 0.79
4	Resting blood pressure (RBP)	FC ₄	129.17 \pm 16.32	134.85 \pm 18.69
5	Serum cholesterol (SCH)	FC ₅	243.49 \pm 53.58	250.73 \pm 49.83
6	Fasting blood sugar (FBS)	FC ₆	0.14 \pm 0.35	0.15 \pm 0.35
7	Resting electrocardiographic results (RES)	FC ₇	0.84 \pm 0.98	1.14 \pm 0.97
8	Maximum heart rate achieved (MHR)	FC ₈	158.59 \pm 18.98	138.89 \pm 22.74
9	Exercise induced angina (EIA)	FC ₉	0.14 \pm 0.35	0.54 \pm 0.49
10	Old peak (OPK)	FC ₁₀	0.59 \pm 0.78	1.64 \pm 1.29
11	Peak exercise slope (PES)	FC ₁₁	1.41 \pm 0.59	1.83 \pm 0.56
12	Number of major vessels colored by fluoroscopy (VCA)	FC ₁₂	0.27 \pm 0.63	1.13 \pm 1.01
13	Thallium scan (THA)	FC ₁₃	3.78 \pm 1.55	5.90 \pm 1.70

diagnostic system achieves accuracy of 91.11% using only a subset of features. The optimal subset of features is obtained for $n = 6$, $n = 7$, and $n = 11$ where n stands for the size of the feature subset. The simulation results are reported in Table 2. In the table, the last record displays a case where all the features are used, i.e., no feature selection is performed. It can be noticed that the best accuracy of 90% is achieved after optimizing the architecture of ANN by a grid search algorithm using all the features. Thus, it is evidently coherent that the proposed model is competent as it presents us better performance with the least number of features. The best performance of the proposed model is observed at 11 features for the peak training accuracy. Additionally, the feature selection module increases the performance of the optimized ANN by 1.11%. Moreover, F_e denotes the features that are eliminated from the features, space during the feature selection process. The results from distinct subsets of features and diverse hyperparameters are displayed in Table 2.

4.2. Experiment No. 2: Feature Selection by FWAFE and Classification by DNN. In this experiment, at the first stage, FWAFE is used, while at the second stage, DNN is implied. The feature selection module eliminates noisy and irrelevant features by exploiting a search strategy, whereas the second model is utilized as a predictive model. The proposed diagnostic system achieves an accuracy of 93.33% using only a subset of features. The optimal subset of features is obtained for $n = 11$ which includes FC₁, FC₂, FC₃, FC₄, FC₇, FC₈, FC₉, FC₁₀, FC₁₁, FC₁₂, and FC₁₃, i.e., by eliminating feature number 5 and 6. The experimental outcomes are displayed in Table 3. To validate the effectiveness of the proposed feature selection method, i.e., FWAFE, the experiment is performed using the DNN model on full features without using the feature selection module. The DNN architecture was optimized using grid search algorithm. The best accuracy of 90% was obtained using neural network with four layers. The size of 1st layer is equivalent to the number of features, 2nd layer consists 50 neurons, and 3th layer contains 2 neurons and output layer has only one neuron. In Table 3, the last row represents a case, whereas all features are utilized. Hence, it is evidently clear that the feature selection module boots the performance of DNN by 3.33%. Moreover, FWAFE-DNN

TABLE 2: Results of different subsets of features for the heart disease dataset.

n	F_e	N	Acc _{train}	Acc _{test} (%)	Sens. (%)	Spec. (%)	MCC
12	4	3	83.09	90.00	85.36	93.87	0.799
12	11	8	87.92	90.00	85.36	93.87	0.799
11	(5, 6)	4	82.60	91.11	85.36	95.91	0.822
11	(11, 12)	4	81.15	90.00	87.80	91.83	0.798
10	(4 to 6)	7	85.99	90.00	82.92	95.91	0.801
7	(3 to 8)	49	95.16	91.11	87.80	93.87	0.820
6	(6 to 12)	6	81.15	91.11	95.12	87.75	0.825
13	—	1	85.02	90.00	87.80	91.83	0.798

shows better performance than FWAFE-ANN. The results at distinct subsets of features on various hyperparameters are shown in Table 3. The ROC charts are utilized to analyze the performance of the proposed model. A method whose ROC chart has maximum area beneath the curve is considered the best. The ROC chart whose points are in the upper left corner is considered to be the best. Figure 3(a) shows the ROC chart of the proposed FWAFE-ANN diagnostic system, while Figure 3(b) denotes the ROC chart of the ANN-based diagnostic system. From the figure, it is evidently vivid that the feature selection module increases the performance of the ANN model owing to more area beneath the curve. Similarly, Figure 4(a) represents the ROC chart of the proposed FWAFE-DNN diagnostic system, while Figure 4(b) depicts the ROC chart of DNN-based diagnostic system. From the figure, it is clearly observed that the feature selection module also increases the performance of the DNN model.

4.3. Experiment No. 3: Results of Other State-of-the-Art Machine Learning Models. In this segment, a comparative analysis is done with other state-of-the-art machine learning models on biomedical datasets against our proposed model. The classifier selected for comparison are random forest (RF) classifier, randomized decision tree classifier, Adaboost ensemble classifier, SVM with radial basis function (RBF) kernel, and linear support vector machine (SVM). Table 4

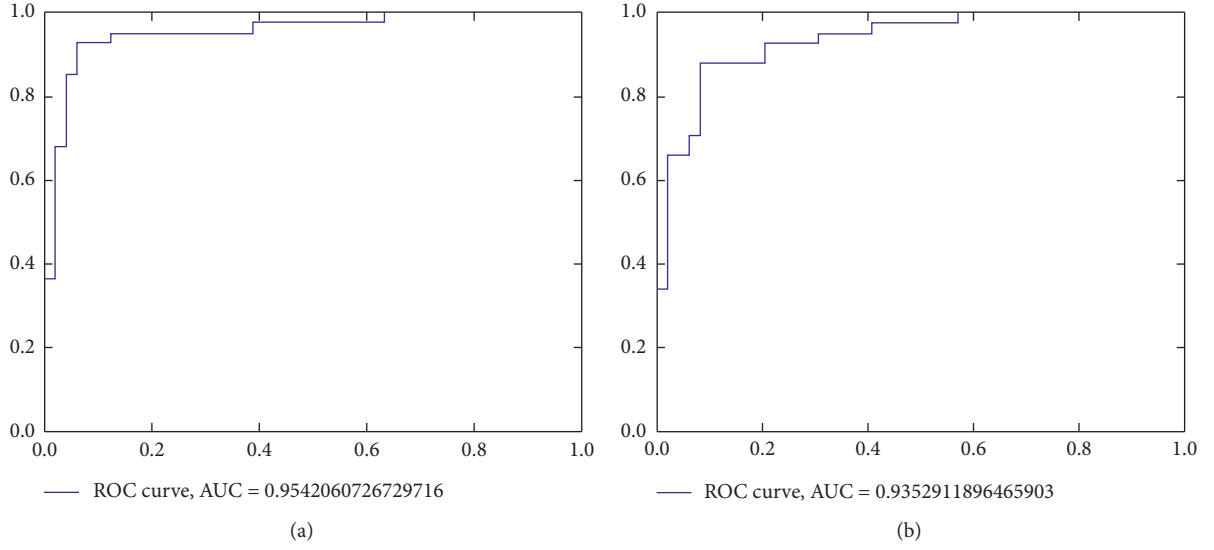


FIGURE 3: The ROC chart of the proposed ANN-based method. The ROC chart of the (a) proposed FWAFE-ANN diagnostic system and (b) conventional ANN-based diagnostic system.

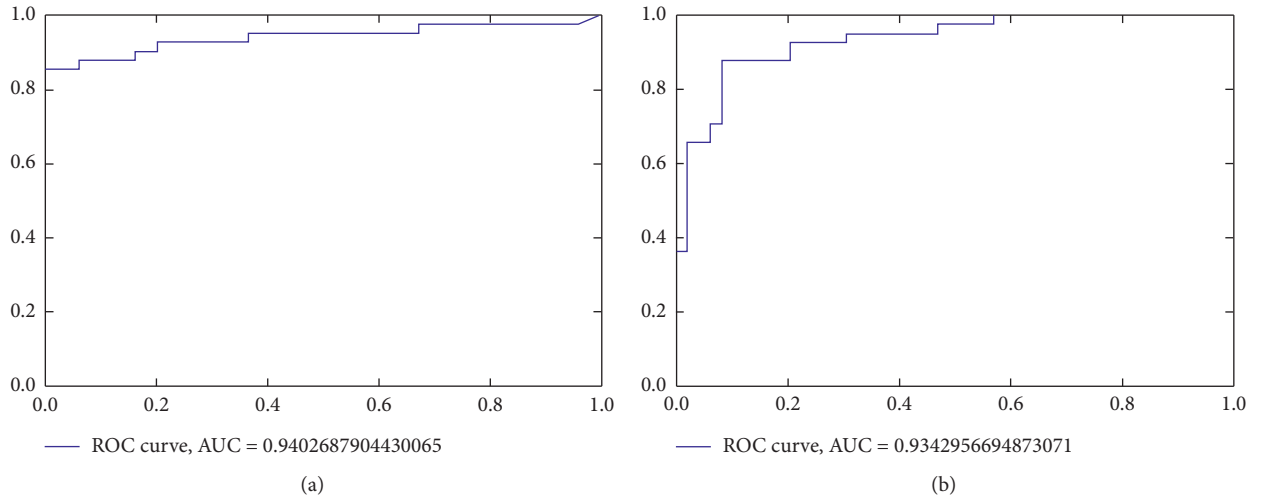


FIGURE 4: The ROC chart of the proposed DNN-based method. The ROC chart of the (a) proposed FWAFE-DNN diagnostic system and (b) conventional DNN-based diagnostic system.

TABLE 3: Results of different subsets of features for the heart disease dataset.

n	F_e	N_1	N_2	Acc_{test}	$Acc_{train}(\%)$	Sen. (%)	Spec. (%)	MCC
12	1	2	19	85.99	90.00	85.36	93.87	0.799
12	2	2	64	85.55	90.00	82.92	95.91	0.801
12	4	3	4	85.50	92.22	82.92	100	0.851
12	5	2	2	83.09	91.11	87.80	93.87	0.820
12	6	2	35	84.05	91.11	85.36	95.91	0.822
12	7	6	8	85.99	90.00	90.24	89.79	0.799
12	10	5	66	87.43	91.11	92.68	89.79	0.822
12	11	3	25	88.40	91.11	85.36	95.91	0.822
12	12	2	12	84.54	91.11	90.24	91.83	0.820
11	(5, 6)	50	2	83.57	93.33	85.36	100	0.872
6	(3 to 9)	16	42	93.23	92.22	90.24	93.87	0.843
13	—	2	4	85.02	90.00	87.80	91.83	0.798

TABLE 4: Performance of various predictive models on the heart disease dataset.

Model	Hyperparameters	Acc _{test}	Acc _{train}	Spec.	Sens.	MCC
Adaboost	$N_e = 25$	85.55	90.33	87.75	82.92	0.708
Adaboost	$N_e = 50$	86.66	94.68	87.75	85.36	0.731
Adaboost	$N_e = 75$	84.44	94.68	85.71	82.92	0.686
Adaboost	$N_e = 100$	82.22	97.58	81.63	82.92	0.643
Random forest	$N_e = 10$	81.11	98.06	87.75	73.17	0.619
Extra tree	$N_e = 10$	68.88	100.0	65.30	73.17	0.383
SVM (linear)	$C = 0.055$	90.00	84.05	93.87	85.36	0.799
SVM (RBF)	$C = 5, G = 0.2$	90.00	84.54	93.87	85.36	0.799
Proposed	(50, 2)	93.33	84.05	100	85.36	0.872

TABLE 5: Classification accuracies of the proposed method and other methods in the literature that used heart disease dataset.

Study (year)	Method	Accuracy (%)
ToolDiag, RA [35]	IB1-4	50.00
WEKA, RA [35]	InductH	58.50
ToolDiag, RA [35]	RBF	60.00
WEKA, RA [35]	FOIL	64.00
ToolDiag, RA [35]	MLP + BP	65.60
WEKA, RA [35]	T2	68.10
WEKA, RA [35]	1R	71.40
WEKA, RA [35]	IB1c	74.00
WEKA, RA [35]	K^*	76.70
Robert Detrano [35]	Logistic regression	77.00
Paul et al. [19]	FDSS	80
Khemphila and Boonijng (2011) [36]	MLP-backpropagation	80.99
Newton Cheung (2001) [37]	BNNF	80.96
Newton Cheung (2001) [37]	C4.5	81.11
Newton Cheung (2001) [37]	Naïve Bayes	81.48
Newton Cheung (2001) [37]	BNND	81.11
WEKA, RA [35]	Naïve Bayes	83.60
Šter and Dobnikar [38]	Fisher discriminant analysis	84.2
Šter and Dobnikar [38]	Linear discriminant analysis	84.5
Šter and Dobnikar [38]	Naïve Bayes	82.5–83.4
Polat et al. (2005) [25]	AIRS	84.50
Dwivedi (2018) [22]	LR	85
Ozsen et al. (2005) [39]	Kernel functions with AIS	85.93
Kahramanli and Allahverdi (2008) [40]	Hybrid neural network system	86.8
Polat et al. (2006) [10]	Fuzzy-AIRS-knn based system	87.00
Amin et al.(2019) [23]	LR, SVM, K -NN	87.4
Verma et al. (2016) [20]	CFS-PSO	88.4
Das et al. (2009) [1]	Neural network ensembles	89.01
Jankowski and Kadirkamanathan (1997) [41]	IncNet	90.00
Ali and Bukhari (2019) [12]	Mutual information + DNN	90
Kumar (2011) [42]	ANFIS	91.18
Shah et al. (2017) [21]	PPCA	91.30
Samuel et al. (2017) [4]	ANN-fuzzy-AHP	91.10
Kumar (2012) [43]	Fuzzy resolution mechanism	91.83
Ali et al. (2019) [34]	Stacked SVMs	92.22
Proposed method (2020)	Feature selection based on FWAFE + ANN	91.11
Proposed method (2020)	Feature selection based on FWAFE + DNN	93.33

denotes the results of the abovementioned models. The performance of each model with hyperparameters values is also depicted from Table 4. For Adaboost classifier, the hyperparameter N_e represents the maximum number of estimators at which the boosting is terminated. For the RF classifier, the hyperparameter N_e denotes the number of trees in the forest. The ensemble model based on

randomized decision trees used average for improving the prediction accuracy. In case of SVM, the width of the Gaussian kernel is denoted by G and soft margin constant is denoted by C . Lastly, the number of neurons in 1st hidden layer of DNN is denoted in the last record of the table (50, 2), i.e., $N_1 = 50$, and the number of neurons in 2nd hidden layer of the DNN, i.e., $N_2 = 2$. Moreover, the performance of the

proposed model is evidently precised and then the various state-of-the-art ensemble models as well as SVM model and are depicted from Table 4.

4.4. Comparative Study with Previously Reported Methods.

In this section, experimental results of the proposed method are compared with those of the other methods discussed in the literature. The performance comparison is based on the prediction accuracy. Hence, Table 5 tabulates the prediction accuracies of our proposed method and other previously proposed methods in the literature. From the experimental outcomes, it is evident that the proposed hybrid method shows promising performance on heart disease, while the main limitation of the proposed method is its high time complexity.

From Table 5, it can be seen that many studies proposed numerous methods for automated detection of HF. For example, Ali et al. developed a two stage system using linear SVM at the first stage for feature selection and linear discriminant analysis model for classification at the second stage and obtained 90% accuracy. In another study, Verma et al. in [20] utilized the correlation-based feature subset (CFS) for feature selection and particle swarm optimization (PSO) algorithm for k -means clustering. Their method produced an accuracy of 88.4%. Saqlain et al in [21] proposed probabilistic principle component analysis and obtained an accuracy of 91.30%. Ali et al. in [34] proposed a novel hybrid method for improving the heart disease prediction accuracy. Their proposed method utilized linear SVM for feature selection and another SVM (with linear and nonlinear kernels) for classification. Their proposed method produced 92.22% heart disease detection accuracy. Hence, based on comparison with these methods, it is clear that our proposed method is a step forward in improving heart disease detection accuracy.

5. Conclusions

In this paper, an effort has been made to design a two stage diagnostic system that can improve the prediction accuracy of heart risk failure prediction. Two types of systems were developed. Both systems used same feature selection method, while the first system used ANN for classification and the second system used DNN for classification. A classification accuracy of 91.11% was achieved with the ANN-based system, while an accuracy of 93.33% was obtained with the DNN-based diagnostic system. It was also observed that the proposed diagnostic system shows better performance than other state-of-the-art machine learning models. From the experimental results, it can be safely concluded that the proposed system can help the physicians to make accurate decision while diagnosing heart disease.

Data Availability

All the data used in this study are available at UCI machine learning repository.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2017R1D1A3B04031440 and by the Natural National Science Foundation of China under grant 61472066, Sichuan Science and Technology Program (Nos. 2018GZ0180, 2018GZ0085, 2017GZDZX0001, 281 2017GZDZX0002, 2018GZDZX0006, and 2018FZ0097).

References

- [1] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.
- [2] H. Yang and J. M. Garibaldi, "Automatic detection of protected health information from clinic narratives," *Journal of Biomedical Informatics*, vol. 58, pp. S30–S38, 2015.
- [3] L. A. Allen, L. W. Stevenson, K. L. Grady et al., "Decision making in advanced heart failure," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.
- [4] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ann and fuzzy_ahp for heart failure risk prediction," *Expert Systems with Applications*, vol. 68, pp. 163–172, 2017.
- [5] L. Ali, S. U. Khan, N. A. Golilarz et al., "A feature-driven decision support system for heart failure prediction based on statistical model and Gaussian naive bayes," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6314328, 8 pages, 2019.
- [6] L. Ali, S. U. Khan, M. Anwar, and M. Asif, "Early detection of heart failure by reducing the time complexity of the machine learning based predictive model," in *Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, IEEE, Swat, Pakistan, July 2019.
- [7] P. K. Anooj, "Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules," *Journal of King Saud University—Computer and Information Sciences*, vol. 24, no. 1, pp. 27–40, 2012.
- [8] İ. Babaoglu, O. Findik, and E. Ülker, "A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3177–3183, 2010.
- [9] S. B. Patil and Y. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," *European Journal of Scientific Research*, vol. 31, no. 4, pp. 642–656, 2009.
- [10] K. Polat, S. Şahan, and S. Güneş, "Automatic detection of heart disease using an artificial immune recognition system (airs) with fuzzy resource allocation mechanism and k -nn (nearest neighbour) based weighting preprocessing," *Expert Systems with Applications*, vol. 32, no. 2, pp. 625–631, 2007.
- [11] S. U. Khan, M. K. A. Rahim, and L. Ali, "Correction of array failure using grey wolf optimizer hybridized with an interior point algorithm," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 9, pp. 1191–1202, 2018.

- [12] L. Ali and S. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *IRBM*, .
- [13] N. A. Golilarz, A. Addeh, H. Gao et al., "A new automatic method for control chart patterns recognition based on convnet and harris hawks meta heuristic optimization algorithm," *IEEE Access*, vol. 7, pp. 149398–149405, 2019.
- [14] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart diseases diagnosis using neural networks arbitration," *International Journal of Intelligent Systems and Applications*, vol. 7, no. 12, p. 72, 2015.
- [15] G. Manogaran, R. Varatharajan, and M. Priyan, "Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4379–4399, 2018.
- [16] L. Ali, S. U. Khan, M. Arshad, S. Ali, and M. Anwar, "A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease," in *Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, IEEE, Swat, Pakistan, July 2019.
- [17] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34938–34945, 2019.
- [18] F. S. Ahmed, L. Ali, B. A. Joseph, A. Ikram, R. Ul-Mustafa, and S. A. C. Bukhari, "A statistically rigorous deep neural network (DNN) approach to predict mortality in trauma patients admitted to the intensive care unit," *Journal of Trauma and Acute Care Surgery*, 2020.
- [19] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. Akhand, "Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease," in *Proceedings of the 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 145–150, IEEE, Dhaka, Bangladesh, May 2016.
- [20] L. Verma, S. Srivastava, and P. Negi, "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data," *Journal of Medical Systems*, vol. 40, no. 7, p. 178, 2016.
- [21] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, "Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis," *Physica A: Statistical Mechanics and Its Applications*, vol. 482, pp. 796–807, 2017.
- [22] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," *Neural Computing and Applications*, vol. 29, no. 10, pp. 685–693, 2018.
- [23] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics and Informatics*, vol. 36, pp. 82–93, 2019.
- [24] S. Özşen and S. Güneş, "Attribute weighting via genetic algorithms for attribute weighted artificial immune system (awais) and its application to heart disease and liver disorders problems," *Expert Systems with Applications*, vol. 36, no. 1, pp. 386–392, 2009.
- [25] K. Polat, S. Sahan, H. Kodaz, and S. Günes, "A new classification method to diagnosis heart disease: supervised artificial immune system (airs)," in *Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN)*, Izmir, Turkey, June 2005.
- [26] D. Dua and C. Graff, *UCI Machine Learning Repository*, 2017.
- [27] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, "Reliable Parkinson's disease detection by analyzing hand-written drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model," *IEEE Access*, vol. 7, pp. 116480–116489, 2019.
- [28] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.
- [29] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.
- [30] T. Meraj, A. Hassan, S. Zahoor et al., "Lungs nodule detection using semantic segmentation and classification with optimal features," *Neural Computing and Applications*, .
- [31] L. Ali, I. Wajahat, N. A. Golilarz, F. Keshtkar, and S. A. C. Bukhari, "LDA-GA-SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Computing and Applications*, pp. 1–10, 2020.
- [32] F. S. Ahmad, L. Ali, Raza-Ul-Mustafa et al., "A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs)," *Journal of Ambient Intelligence and Humanized Computing*, 2020.
- [33] A. K. Paul, P. C. Shill, M. R. I. Rabin, and K. Murase, "Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease," *Applied Intelligence*, vol. 48, no. 7, pp. 1739–1756, 2017.
- [34] L. Ali, A. Niamat, J. A. Khan et al., "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019.
- [35] Datasets Used for Classification: Comparison of Results (2007), <http://www.is.umk.pl/projects/datasets.html>.
- [36] A. Khemphila and V. Boonjing, "Heart disease classification using neural network and feature selection," in *Proceedings of the 2011 21st International Conference on Systems Engineering*, pp. 406–409, IEEE, Las Vegas, NV, USA, October 2011.
- [37] N. Cheung, *Machine learning techniques for medical analysis. school of information technology and electrical engineering*, Ph.D. thesis, B. Sc. Thesis, University of Queensland, Brisbane, Australia, 2001.
- [38] B. Šter and A. Dobnikar, "Neural networks in medical diagnosis: comparison with other methods," in *Proceedings of the International Conference on Engineering Applications of Neural Networks*, pp. 427–430, London, UK, June 1996.
- [39] S. Ozsen, S. Gunes, S. Kara, and F. Latifoglu, "Use of kernel functions in artificial immune systems for the nonlinear classification problems," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 621–628, 2009.
- [40] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 82–89, 2008.
- [41] N. Jankowski and V. Kadirkamanathan, "Statistical control of rbf-like networks for classification," in *Proceedings of the International Conference on Artificial Neural Networks*, Springer, Lausanne, Switzerland, pp. 385–390, October 1997.
- [42] A. S. Kumar, "Adaptive neuro-fuzzy inference system for heart disease diagnosis," in *Proceedings of the International*

Conference on Information System, Computer Engineering & Application (ICISCEA 2011), pp. 91–99, Singapore, 2011.

- [43] A. S. Kurnar, “Diagnosis of heart disease using fuzzy resolution mechanism,” *Journal of Artificial Intelligence*, vol. 5, no. 1, pp. 47–55, 2012.