IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 4, 2020, accepted March 31, 2020, date of publication April 3, 2020, date of current version April 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985646

An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tompa and Stacked Genetic Algorithm

SYED ARSLAN ALI[1], BASIT RAZA[1], AHMAD KAMRAN MALIK[1], AHMAD RAZA SHAHID[1], MUHAMMAD FAHEEM[2], HANI ALQUHAYZ[3], AND YOGAN JAYA KUMAR[4]

[1]Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad 45550, Pakistan
[2]Department of Computer Engineering, Abdullah Gul University (AGU), Kayseri 38039, Turkey
[3]Department of Computer Science and Information, College of Science in Zulfi, Majmaah University, Al-Majmaah 11952, Saudi Arabia
[4]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Melaka 76100, Malaysia

Corresponding author: Basit Raza (basit.raza@comsats.edu.pk)

This work was supported by the Deanship of Scientific Research at Majmaah University.

ABSTRACT A rapid increase in heart disease has occurred in recent years, which might be the result of unhealthy food, mental stress, genetic issues, and a sedentary lifestyle. There are many advanced automated diagnosis systems for heart disease prediction proposed in recent studies, but most of them focus only on feature preprocessing, some focus on feature selection, and some only on improving the predictive accuracy. In this study, we focus on every aspect that may have an influence on the final performance of the system, i.e., to avoid overfitting and underfitting problems or to solve network configuration issues and optimization problems. We introduce an optimally configured and improved deep belief network named OCI-DBN to solve these problems and improve the performance of the system. We used the Ruzzo-Tompa approach to remove those features that are not contributing enough to improve system performance. To find an optimal network configuration, we proposed a stacked genetic algorithm that stacks two genetic algorithms to give an optimally configured DBN. An analysis of a RBM and DBN trained is performed to give an insight how the system works. Six metrics were used to evaluate the proposed method, including accuracy, sensitivity, specificity, precision, F1 score, and Matthew's correlation coefficient. The experimental results are compared with other state-of-the-art methods, and OCI-DBN shows a better performance. The validation results assure that the proposed method can provide reliable recommendations to heart disease patients by improving the accuracy of heart disease predictions by up to 94.61%.

INDEX TERMS Heart disease, prediction, deep belief network, genetic algorithm, Ruzzo-Tompa.

I. INTRODUCTION

Most of the deaths happening worldwide in the last few years have one primary reason behind them, which is heart disease. It is considered as the most common health problem, and its main cause is that the amount of blood required to meet the body needs is not pumped well enough by the heart [1]. It may be caused by unhealthy food, mental stress, genetic issues, or lack of exercise. It attacks a person slowly, and is very difficult to deal with unless it has progressed to a dangerous level. If the required remedial measures are not taken, it may adversely affect the patient. Hence, heart disease prediction is necessary so that people can take care of their health as soon as a heart disease symptom is observed [2]. The most common way to get all the information related to a patient is to use a hospital information system to manage the health of the patient in a precise way.

The associate editor coordinating the review of this manuscript and approving it for publication was Alicia Fornés.

Heart disease is mainly caused by risk factors categorized as changeable factors and non-changeable factors [3], [4]. Changeable factors include blood pressure, cholesterol level, etc., while non-changeable factors include a patient's history, age, and sex. There is a substantial amount of hidden information in a patient's medical data, but a major challenge in hospitals is that a large amount of this data is not being used adequately to support clinical decisions [5]. This extensive information can also be used to predict heart disease accurately in the early stages. Although the diagnosing of heart diseases is done by invasive cardiology-based techniques like angiography, it relies on a patient's medical history and a high level of technical expertise to analyze the concerned symptoms. Additionally, it is a complex process with high costs, so noninvasive techniques like decision support systems based on machine learning and deep learning must be used to overcome these issues.

Many researchers in recent years have produced different methods and models to design an expert system that should be capable of predicting heart disease in a patient in the early stages. There has been significant improvement in the quality of these proposed diagnostic systems. This motivated us to propose the OCI-DBN approach that is capable of handling the overfitting and underfitting problems and the network configuration and optimization problems so that the diagnosis of heart disease can be improved.

In our proposed OCI-DBN approch, the main objective is to resolve those problems that were the main reasons for the degraded performance of diagnostic systems, which prevented the predictive accuracy from reaching the level that is required. One problem is how to train the data so that the model does not overfit or underfit. Another problem is how to find the optimal configuration of the network so that the optimization problem can be resolved. When the model learns too much from the training data, it overfits because it picks even small details from the training data, and when applied to the testing data, the results are not adequate [6], [27]. On the other hand, when the model has not learned enough from the training data, it underfits, and as a result, both the training data and testing data show poor results. Not setting up the right configuration for the network and having few irrelevant features that do not contribute enough are the main reasons behind these problems. Because of these problems, the computational complexity increases, and the prediction time for heart disease increases. To remove the features that are noisy or redundant, we use the Ruzzo-Tompa algorithm because of its effectiveness in searching the optimized features. The next step is to propose an improved DBN model for classification and to optimize the network configuration. We use a stacked genetic algorithm (SGA) that is capable of fine-tuning the network to find an optimal solution and improve system performance.

In the final step, experiments were performed, and six evaluation metrics were considered, including accuracy, specificity, sensitivity, precision, Matthew's correlation coefficient (MCC), and F1 score. The results were validated using 10-fold cross-validation. For training and testing, Cleveland's heart disease dataset from an online data mining repository of the University of California, Irvine (UCI), was used. The dataset was chosen because most of the research on heart disease prediction use this dataset for finding the effectiveness of their model. The experiments were performed in Python on a PC with an Intel®Core^{TM}i7-4600M @2.90 GHz CPU.

The major research contributions are described below:

i. A novel approach, OCI-DBN, is proposed for classification and improvement in heart disease prediction that does not overfit or underfit. To handle the overfitting problem, many feature selection techniques were considered for selecting the optimal subset of features. Among these, Ruzzo-Tompa provided the best performance in the proposed OCI-DBN approach.
ii. The proposed OCI-DBN approach also solves the network optimization and configuration problem by finding the optimal width of each hidden layer in a DBN using an SGA. We embedded a local search operation after getting the optimal value from the simple genetic algorithm. The results are sent back for further improvement of the fine-tuning capability of the genetic algorithm. The process is repeated, and SGA finds an optimal number of layers and nodes to be used in DBN. An analysis is also performed for a RBM and trained DBN to depict the system working and also provided results with different network depths in DBN.
iii. We show that even with the small size Cleveland dataset, the performance of DBN is better than neural networks with one hidden layer when provided with an optimal subset of features and optimal numbers of layers and nodes.

The study is organized as follows. Section II presents the related work. The proposed approach is provided in Section III. The dataset description and evaluation metrics are presented in Section IV. The results of the experiments when the proposed approach is applied to the dataset are discussed in Section V, while the research conclusion and future work are presented in Section VI.

II. RELATED WORK

Many existing studies have presented different approaches, including different machine learning and deep learning techniques for prediction in general [19], [20] and particularly to diagnose and predict heart disease correctly [6], [7]. Many techniques including support vector machine (SVM), artificial neural network (ANN), fuzzy logic, deep neural networks (DNN), decision trees, and long short-term memory have been applied for identifying heart disease symptoms in patients. The quality of decision making improves significantly using these expert systems, and there has been a slight decrease in the death rate because of these expert diagnostic systems [6], [7].

A multimodal disease prediction algorithm is presented in [8], which is based on the structured and unstructured data acquired from the hospital, and their proposed approach

works well in predicting chronic disease. A combination of neural network and fuzzy logic was proposed in [9] that gives a contribution score for each attribute in predicting heart disease. Their results had improved accuracy when compared with the conventional neural network approach. Similarly, a heart disease prediction model is presented in [10] that uses important statistics of a patient's medical record. They ensembled three best models to improve the overall prediction accuracy. The techniques used were AdaBoost, boosted tree, and naïve Bayes, and the researchers showed that their proposed approach worked well in predicting heart disease patients with better accuracy than the conventional methods.

Paul *et al.* proposed an automatic fuzzy diagnostic system [11] based on five important steps that achieved a predictive accuracy of 92.31%. Recently a heart disease prediction model [12] was proposed using a neural network approach along with Ruzzo-Tompa. Two steps are performed in the proposed approach, the first is to minimize the data dimensionality using hybridized Ruzzo-Tompa, and the second is classification using a neocognitron neural network. Another hybrid method [13] was proposed by Zeinab *et al.* by using a genetic algorithm to enhance the initial weights of neural networks, which improves the performance and achieves a predictive accuracy of 93.85%. In the research presented in [14], the overfitting problem is addressed by first training the neural network with each feature ranking and then training it to the extent that it is capable of outputting a potential heart disease diagnosis. A combination of multiple kernel learning with an adaptive neuro-fuzzy interference system was proposed in [15] for heart disease diagnosis. A twofold approach is used in this study in which multiple kernel learning is used for parameter separation between patients and healthy people, while the results from this are given to adaptive neuro-fuzzy for classification. Another efficient and expert diagnostic system for heart disease is proposed in [16] using two SVM models. The first SVM is used for removing irrelevant features from the feature set, while the second SVM is used for prediction. For optimizing the two models, they used a hybrid grid search algorithm so that both the models can be optimized simultaneously.

III. PROPOSED OCI-DBN APPROACH

In this section, we present a novel approach to improve the performance of heart disease prediction by overcoming the overfitting and underfitting problems and the optimization problems, i.e., finding the optimal configuration for the deep neural network. We first describe the feature selection process through an effective evolutionary Ruzzo-Tompa algorithm. Then we present an optimally configured and improved DBN that can avoid the problems mentioned above and improve the heart disease prediction process.

A. RUZZO-TOMPA FOR FINDING THE OPTIMAL SUBSET OF FEATURES

The goal of the Ruzzo-Tompa algorithm is to find the optimal solution that can play an important role in the heart disease classification process. To find the optimal solution, the algorithm must select the optimal subset of features, i.e., the most important and relevant features to add more importance to the heart disease prediction process. Selecting crucial features is necessary because it reduces the complexity and computational time, which are both very important factors in developing diagnostic systems [3], [17]. The whole feature selection process using the Ruzzo-Tompa algorithm is shown in Figure 1, in which we have collected the input features of the heart disease dataset in the search space so that the searching process can start. As each feature of the heart disease dataset has its own characteristics and significance in classifying heart disease, the searching takes place in a continuous manner. The time taken for feature searching in search space is O(n) with O(n) space, where n is the feature length. In this study, a set of rules for getting the optimal subset of features from the heart disease dataset is followed.

1. Feature searching takes place from left to right, i.e., from the first feature of the heart disease dataset, which is age, and continues until the last feature, thallium.
2. The next step is to compute the fitness value for each feature. Equation 1 is used for this purpose [12]:

$$f^i(t) = (1-p)(t-1) + \text{`}\Upsilon^i(x^i(t)) \tag{1}$$

   Here t represents time, f and i represent feature level and feature position in the list, respectively. Two constants are used, which are represented as p and `Υ, and represents the decay constant and luciferin constant. The search space is denoted by $x^i(t)$.
3. Check if all the features now have a fitness value computed and stored. If there are still some features left, then continue the computation process; otherwise, store and arrange the features with respect to their fitness values.
4. Calculate the cumulative sum, select features with maximum fitness values, and store as a feature subset.
5. Start the searching process again by adding the selected feature again in the list, but this time the searching direction will be from right to left.
6. Pick the maximum fitness value j and check if it satisfies the condition $L_j < L_k$ where L_j is the cumulative sum, and L_k is the newly formed subsequence set.
7. If the condition is not satisfied, then start the searching process by adding the feature at the end of the list, and repeat all the steps again. If the condition is satisfied, then this feature and repeat the process until we have the optimal subset of features.

B. DEEP BELIEF NETWORK

DBN is based on a Restricted Boltzmann Machine (RBM) and consists of ANN in which a RBM structure is composed of hidden layers and visible layers [17], [23]. The connectivity between neurons in the same layer is weak, but it is strong and fully connected between two layers, which means that there are connections only between visible and hidden layers. If there are visible and hidden states, then the RBM

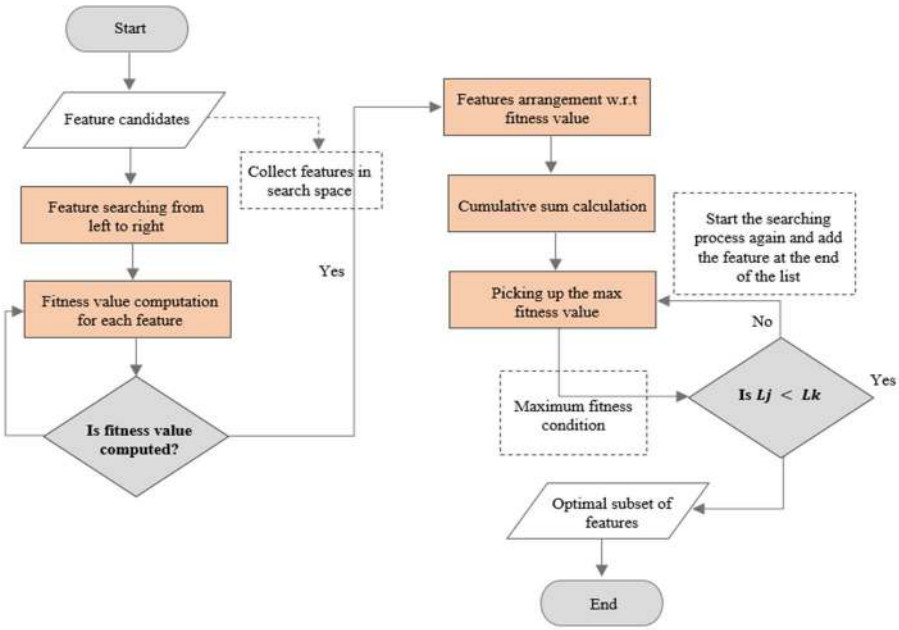

FIGURE 1. Ruzzo-Tompa-based feature selection process.

energy function can be shown as in Equation 2.

$$E(v, h) = \sum_{i=1}^{n} a^i v^i - \sum_{j=1}^{m} b^j h^j - \sum_{i=1}^{n} \sum_{i=1}^{m} v^i w^{ij} h^j \tag{2}$$

In Equation 2, the offset vector of the visible layer is denoted by $a = (\mathbf{a^1}, \mathbf{a^2}, \mathbf{a^3}, \ldots\ldots, \mathbf{a^n})$ while a hidden layer bias vector is denoted by $b = (\mathbf{b^1}, \mathbf{b^2}, \mathbf{b^3}, \ldots\ldots, \mathbf{b^m})$. The state vector for a visible layer is denoted by $v = (\mathbf{v^1}, \mathbf{v^2}, \mathbf{v^3}, \ldots\ldots, v^n)$ while the hidden layer state vector is denoted by $h = (\mathbf{h^1}, \mathbf{h^2}, \mathbf{h^3}, \ldots\ldots, \mathbf{h^m})$. The weight matrix between the visible layers i and the hidden layers j is denoted by $\mathbf{w^{ij}}$. To make predictions, the goal of the training model is to generate a hypothesis, which means that the model analyzes the training data, and a fitting function is learned [25]. In DBN, two stages are followed to build a heart disease prediction system. The first stage consists of training upwards, and the second stage consists of downward adjustments. The first stage is the training stage, in which RBM is trained by a greedy layer-by-layer approach [17], [24]. The initial weights are set up by continuously performing unsupervised training. At this time, the backpropagation algorithm is not performed for error analyses, but once the initial weights are determined, then it uses the backpropagation algorithm for tuning the parameters and optimizing the performance. Although the output after these two stages is the DBN, after fine-tuning the parameters, the optimization problem for the network remains as to how many layers and nodes to use and the hyper-parameters needed for best performance. The improvement in performance results for any technique is deeply linked with the number of layers and nodes, but finding an optimal value is a major problem because it is dependent on dataset type and dataset characteristics. If random values are provided to the DBN, then the model may show good performance, but one cannot say that this is the optimal value for resolving the optimization problem of the network. To solve this optimization problem, we use two stacked genetic algorithms, known as SGAs.

C. A STACKED GENETIC ALGORITHM FOR OPTIMIZATION

A genetic algorithm works on the basis of crossover and mutation to reproduce artificial chromosomes from the initial population or set of parameters. Those chromosomes that have high fitness values survive, while the others fall out. These high fitness values are given to the next generation, and the crossover process is followed. The mutation step is performed at the end, in which the global search is maximized, and the optimum value is found. The chromosome represents the selected feature subset x, and to make sure that a feature subset will satisfy the subset size requirement, it is being forced with size value 'd' as a constraint. Any selected feature subsets that break this constraint will be given a penalty. Therefore, the fitness of the chromosome can be defined by Equation 3.

$$C = J(Xc)penalty\ (Xc) \tag{3}$$

Here Xc denotes the corresponding feature subset for C, and the penalty is separately defined as Equation 4.

$$\mathbf{Penalty\ (Xc)} = \mathbf{w} * ||\mathbf{Xc}| - \mathbf{d}| \tag{4}$$

where w is the penalty coefficient.

To solve the optimization problem, we considered simple genetic algorithm, although they can perform better in search

space exploration when properly controlled. However, they are weak to fine-tune the parameters near local optimum points [13]. Therefore, we used an SGA for improving the fine-tuning capability of the simple genetic algorithm. In simple genetic algorithm after getting the optimal value from a mutation step, we embed a local search operation after selection and sent the results back to compute the fitness value again for the feature subset. In SGA, when there is a need for replacement, the feature subset gets a chance to improve itself locally, which helps to gain its fitness value when it sends back for computation. This step is performed regularly until the SGA achieves an optimal value and decides how many layers and nodes to use in DBN. So that the optimal values can solve the optimization problem and can improve the system performance.

The proposed OCI-DBN approach and all the phases involved in it are shown in Figure 2. The system is divided into three main phases. The first phase gets the dataset and assigns the feature types as discrete or continuous. Six features fall under the discrete type, while seven features fall under the continuous type. The next task is to normalize the dataset by min-max scaling; all the features having different ranges are transformed to a range of 0-1 for better understanding by the learning algorithm [26]. In the second phase, the main step is to use feature selection because there is a possibility that due to the noisy nature of some of the features, the performance may be affected. Overfitting is another problem caused when the model learns too much from the training dataset, which makes it unable to produce the desired performance and results on the test set. This is also solved if features are selected based on their relevance and contribution to improving the performance of the system [22], [27]. In this study, we used Ruzzo-Tompa for feature selection, which selects relevant features by following a series of steps that find the optimal subset of features and discard the irrelevant features. The third and the last phase is to find the optimal network configuration for the DBN so that the DBN can be improved and can provide better predictive results. To get the optimal network configuration values, like the number of layers, nodes, and hyperparameters, we use SGA. The optimal values are then ready to be provided to the DBN model once the RBM is constructed and DBN is configured. To configure the DBN for building heart disease prediction system, a two-step process is followed; first step is to train upwards and the second step is the downward adjustment by utilizing the generative weights using fine tuning process. As shown in the Figure 2, RBM is constructed by dividing the multilayer DBN in simpler models because the training of shallow network is easier than the deep network. At this phase, pre-training is required because the optimization can be done in a much better way by initializing all the layers weights. The pre-training in the first step is done by greedy learning approach in which the initial weights are set up by continuously performing unsupervised training. The use of greedy learning approach also helps in learning the parameters for each layer of RBM. The initial weights are responsible for determining the dependency of variables from one layer to the other. The first RBM layer i.e. the visible layer received the training data and created a V1 state. The initial weight W1 is responsible for generating a hidden layer state and for reconstructing the V1 state by hidden layer state. In this way, the first RBM training is completed. In the same way, RBM2 and RBM3 are trained using greedy learning approach. Once these initial weights are determined, then the backpropagation algorithm is used for fine tuning the parameters and performance optimization. Fine tuning helps to adjust the weights regularly by modifying the features slightly. After performing these two stages, the final output is the trained DBN but the optimization problem for the network remains as to how many layers and nodes to use and the hyperparameters needed for best performance. The improvement in performance results for any technique is deeply linked with the number of layers and nodes, but finding an optimal value is a major problem because it is dependent on dataset type and dataset characteristics. As the optimal values are already found out by SGA, the values are provided to the configured DBN model for improving the performance and the accuracy of the model.

IV. DATASET, VALIDATION SCHEME AND EVALUATION METRICS

In this section, the dataset used for heart disease prediction is briefly described, and then the validation scheme and evaluation metrics are presented.

A. DATASET DESCRIPTION

The Cleveland heart disease dataset is used to carry out experiments described in this paper, and it is publicly available online at the UCI machine-learning repository. A total number of 76 features are included in the dataset; however, existing studies mostly use 13 features. The dataset contains a total of 303 instances with missing values for 6 instances, which are discarded for experimental purposes. As different features of the heart disease dataset have different values and orders of magnitude, we use min-max scaling to adjust the values to the range of 0-1 so that the training model can easily interpret it. The names, short code, data type, and description of each of these feature sets are shown in Table 1.

B. VALIDATION SCHEME AND EVALUATION METRICS

Datasets are most commonly divided into three subsets, namely training, test, and validation. To train the dataset, the training subset is used. To optimize the parameters of classifiers, a validation subset is used, and to evaluate the performance of the classifier, the test subset is used [28]. Cross-validation is a statistical technique in which the dataset is divided into some chunks, one chunk for testing, and the remaining for training. Each chunk is selected for testing in one of the folds, ensuring the use of all the chunks in both training and testing [18],

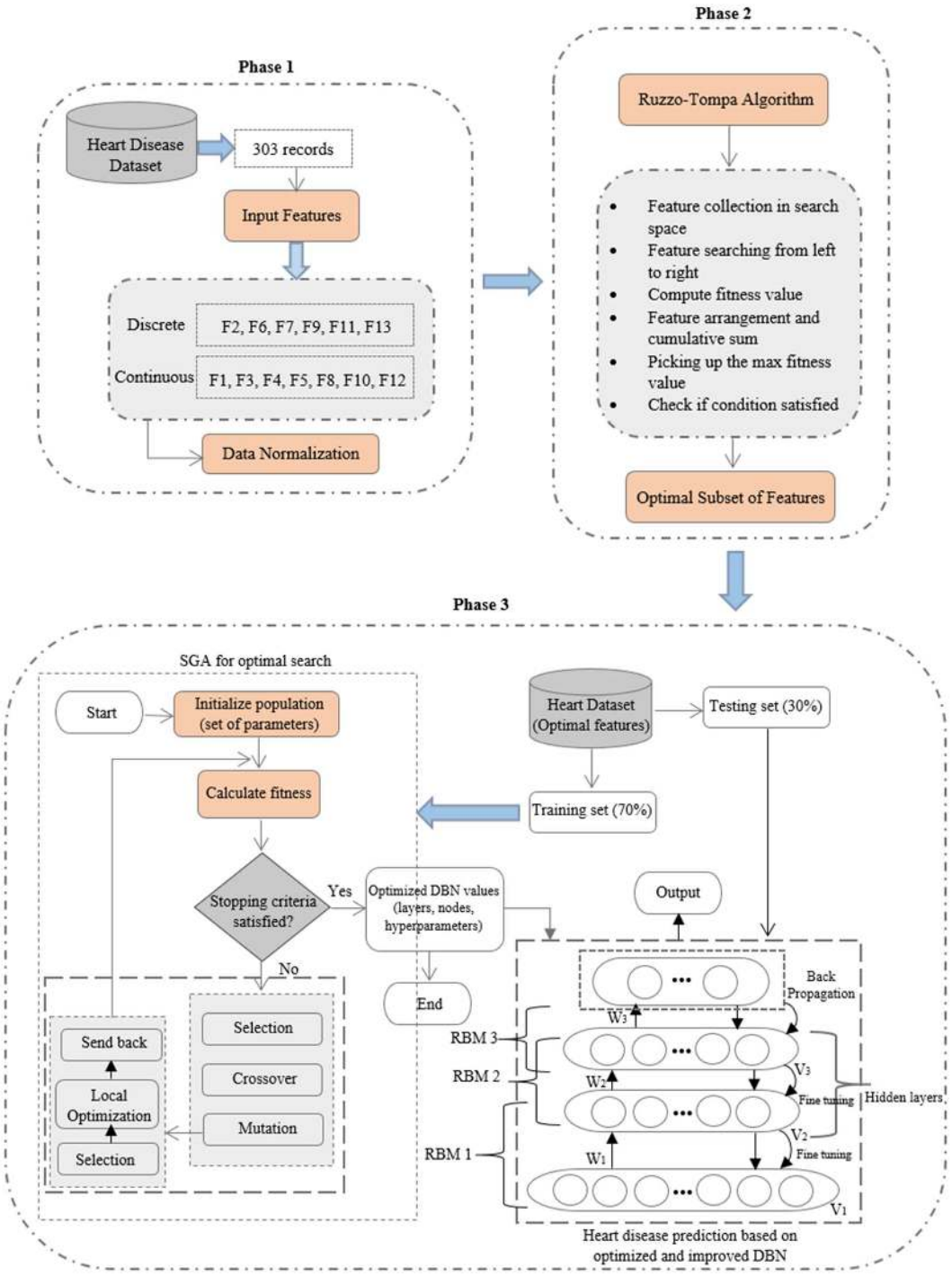

FIGURE 2. Proposed OCI-DBN using Ruzzo-Tompa and SGA.

[29]. The results are averaged over 10 folds. The evaluation metrics, like sensitivity, f-score, specificity, precision, MCC, and accuracy, are used to determine the performance of a proposed OCI-DBN approach. Sensitivity, specificity, and precision are described in Equation 5, 6, and 7, respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{x} 100 \tag{5}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \text{x} 100 \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \text{x} 100 \tag{7}$$

TABLE 1. Thirteen commonly used features from the Cleveland heart disease dataset.

No.	Code	Abbreviation	Type	Description
1	F1	AGE	Continuous	Patient age (in years)
2	F2	SEX	Discrete	Patient gender
3	F3	CPT	Continuous	Chest pain type
4	F4	RBP	Continuous	Resting blood pressure
5	F5	SCH	Continuous	Serum cholesterol
6	F6	FBS	Discrete	Fasting blood sugar
7	F7	RES	Discrete	Resting electro cardio graphic
8	F8	MHR	Continuous	Maximum heart rate
9	F9	EIA	Discrete	Exercise induced angina
10	F10	OPK	Continuous	ST depression
11	F11	PES	Discrete	The slope of the peak exercise ST Segment
12	F12	VCA	Continuous	Number of major vessels
13	F13	THA	Discrete	Thalassemia

MCC and prediction accuracy are given in Equation 8 and 9:

$$\text{Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \text{x100} \tag{8}$$

$$\text{MCC} = x = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP}+\text{FP})\,(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{9}$$

The performance and efficiency of the model are shown by the F-measure, which is another very important statistical index and is defined in Equation 10 below.

$$\text{F-Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{10}$$

In these equations, the patients with heart disease that are correctly classified are considered as True Positive (TP), the patients without any heart disease that are correctly classified are denoted by True Negative (TN), the patients without any heart disease but misclassified as having heart disease are represented by False Positive (FP), while the patients having heart disease but misclassified as not having heart disease are denoted by False Negative (FN).

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the results of our experiments on the heart disease dataset are described in detail. All the experiments are performed in Python on a PC with an Intel®Core™ i7-4600M @2.90 GHz CPU. In our proposed OCI-DBN approach, the first step is to use the feature selection technique. This step is necessary, as selecting relevant features from the dataset will help to improve model performance. For this purpose, an efficient evolutionary feature selection technique, Ruzzo-Tompa, is used. The idea is to select those features that contribute more to improving the performance of the system because, if all the features are given to the model, the noisy nature of some features might affect the performance of the model [21]. Additionally, the results will be best on the training set but not on the testing set [22]. Ruzzo-Tompa serves in this case by selecting the most relevant features, thereby also avoiding the overfitting problem due to less relevant features. Seven features are selected as optimal features, and the feature codes are as follows: F3, F7, F8, F9, F10, F12, and F13.

After selecting the optimal subset of features, three experiments are performed so that the effectiveness of the proposed OCI-DBN approach can be evaluated. In the first experiment, OCI-DBN approach is applied to the selected features, and the results are noted down. In the second experiment, the same set of optimal features is given to the conventional DNN, and the conventional ANN and the results are noted down. In the third experiment, there are two parts; firstly, all the features from the heart disease dataset are given to the proposed model, the conventional DNN, and the conventional ANN to make the comparative analysis stronger and to show that the optimal subset of features produces better results. Secondly, to conclude that the proposed feature selection technique is suitable for OCI-DBN, we have used other well-known feature selection techniques and applied the selected features from them on the trained OCI-DBN model. The experimental results from these different feature selection techniques are then compared for making comparative analysis much stronger and to prove that the proposed OCI-DBN based on Ruzzo-Tompa and SGA gives the best performance and best prediction results for heart disease. Finally, the proposed OCI-DBN approach is compared to other conventional machine learning methods and state-of-the-art techniques that have been used in recent years for heart disease prediction to show the effectiveness of the proposed approach.

A. EXPERIMENTAL RESULTS OF PROPOSED OCI-DBN APPROACH

In the first experiment, the 297 records of the heart disease dataset are divided into two sets, training and testing. 70% of records are used as a training set and 30% are used for testing the proposed model. Less training data tends to cause overfitting so in order to overcome the problem of overfitting there are several remedies taken.. Firstly, the main cause of overfitting is fewer training samples with a large number of features. The feature selection algorithm Ruzzo-Tompa is utilized to select the most relevant and important features so that only important features can be used for the instances chosen as training samples. This tends to resolve the problem of overfitting. Secondly, the deep learning methods such as the DBN has a dropout and drop connect technique. In this study, during the training process, the neurons are randomly deactivated knows as dropout and connections are randomly deactivated known as drop connect. This deactivation forces the network to become redundant and independent and the network does not have to depend on specific neurons or connections to extract specific features. As soon as the training part is done then all the neurons and connections are restored. This helps to avoid the overfitting problem. To train the proposed model on the training set, it is necessary to configure the network so that the model can perform well on both training and testing sets.

Simple genetic algorithm has shown great results and performance in network optimization in a recent study [34]. However, in our study, we used SGA to find the optimal network configurations for our proposed model because simple genetic algorithm achieves lower performance than the SGA. The information about the parameters of the SGA are as follows: the initial population is set randomly at 200, number of generations used are 100 with crossover and mutation probability of 0.5 and 0.001, respectively. In the beginning, the error rate is high, but as soon as the generations increase, the error rate starts to decrease. At the end of this process, we have the optimal configuration network, which is two hidden layers and eight and three neurons in the first and second layers, respectively. The experimental results revealed that the proposed OCI-DBN model achieved 94.61% predictive accuracy when the optimized configuration of 2 hidden layers with 8 and 3 neurons in the first and second hidden layers are used respectively. The accuracy achieved by proposed model using SGA has improved by 1.68% in comparison with the accuracy achieved by using simple genetic algorithm. The proposed OCI-DBN model also achieved 96.03% sensitivity, 93.15% specificity, 93.55% precision, 0.947 F1-Score and 0.892 MCC score. We have performed experiments by using different number of hidden layers to compare and analyze the results between optimized DBN configuration and standard DBN configuration. For each RBM, the training period is set and the initial values are provided to network i.e. number of hidden layers, neurons in each layers and hyperparameters. These values are provided to standard DBN configuration because we have already found out an optimized network configuration from SGA for proposed OCI-DBN model which is 2 hidden layers. The results are shown in Table 2.

TABLE 2. DBN accuracy with different network depth.

Network Configuration	Network Depth / No of Hidden Layers	Accuracy (%)
Optimized (Proposed OCI-DBN model)	2	94.61
Standard	3	91.58
Standard	4	88.22
Standard	5	86.53

The analysis of Table 2 shows that the optimal network configuration obtained by SGA with 2 hidden layers and 8 and 3 neurons in the first and second hidden layer achieves the best accuracy while the standard DBN configuration with other number of hidden layers i.e. 3, 4 and 5 hidden layers have achieved lower accuracy. This is due to the fact that a local search operation is performed after getting the optimal value from the simple genetic algorithm and the results are sent back for fine-tuning. The process is repeated, and SGA finds an optimal number of layers and nodes to be used in DBN which achieves better performance than the other network configurations.

TABLE 3. An optimal subset of features on OCI-DBN, DNN, and ANN.

Performance metrics	ANN (%)	DNN (%)	OCI-DBN (%)
Sensitivity	87.92	91.95	96.03
Specificity	84.46	89.19	93.15
Precision	85.06	89.54	93.55
F1-Score	86.47	90.73	94.77
MCC	72.43	81.17	89.25
Accuracy	86.20	90.57	94.61

B. EXPERIMENTAL RESULTS OF DNN AND ANN

After getting the experimental results for the proposed model, the next step is to use conventional learning methods like DNN and ANN by giving them the optimized features. The data was divided in a ratio of 70:30, to be used as training and testing sets, respectively [30]. For DNN, the optimized network configuration is found by the SGA, and for ANN, there is no optimization. We cannot add any layer because it contains only one hidden layer. To evaluate the performance of these two learning methods, six evaluation metrics were again used. In order to compare all the three models i.e. proposed model, DNN model and ANN model with optimal subset of features and to get an idea of how much difference in there in the performance metrics, we have presented the comparisons in Table 3.

The comparative analysis presented in Table 3 reveals the fact that there is much difference in the performance of the proposed model and the other two models i.e. conventional ANN and the DNN. The sensitivity, specificity, precision and f1 score for OCI-DBN is better than the other two models with an optimum subset of features. It can also be seen that the sensitivity, specificity, precision, f1 score, MCC and accuracy of applying the DNN model on an optimal subset of features achieves better results than the conventional ANN. This is because the conventional DNN was optimized by using the same strategy of stacked genetic algorithm and that is why it gives better performance and predictive accuracy than the conventional ANN because it cannot be optimized to improve the performance of the model.

C. EXPERIMENTAL RESULTS OF OCI-DBN, DNN, AND ANN ON FULL FEATURE SET

This experiment was conducted to evaluate the effectiveness of the proposed Ruzzo-Tompa algorithm for feature selection. After getting the results for the proposed model OCI-DBN, DNN with the same network optimization strategy, and conventional ANN, the next step is to test their performance on a full feature set so that a comparative analysis can be made to show that Ruzzo-Tompa selected the best features, improve the performance by removing features that have a noisy behavior or features that are not contributing enough

TABLE 4. Classification accuracy achieved with other well-known feature selection techniques on proposed OCI-DBN, DNN, and ANN.

Common feature selection algorithms	Selected Features (code)	The accuracy achieved in ANN (%)	The accuracy achieved in DNN (%)	The accuracy achieved in the proposed OCI-DBN (%)
Info gain	12 (F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9, F_{10}, F_{12}, F_{13})	76.09	80.13	84.11
Relief	6 (F_3, F_7, F_8, F_9, F_{10}, F_{13})	78.45	76.77	84.24
Recursive feature elimination (RFE)	10 (F_1, F_2, F_3, F_4, F_7, F_8, F_{10}, F_{11}, F_{12}, F_{13})	83.50	78.79	85.73
Correlation-based feature selection (CFS)	8 (F_3, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}, F_{13})	80.81	81.82	87.32
Chi-squared	11 (F_1, F_2, F_3, F_4, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}, F_{13})	85.52	87.88	90.17
Lease absolute shrinkage and selection operator (LASSO)	8 (F_2, F_3, F_5, F_8, F_9, F_{11}, F_{12}, F_{13})	82.15	84.51	91.46
Genetic algorithm	8 (F_3, F_4, F_6, F_7, F_8, F_9, F_{10}, F_{13})	80.13	81.14	92.62
Ruzzo-Tompa	7 (F_3, F_7, F_8, F_9, F_{10}, F_{12}, F_{13})	86.20	90.57	94.61

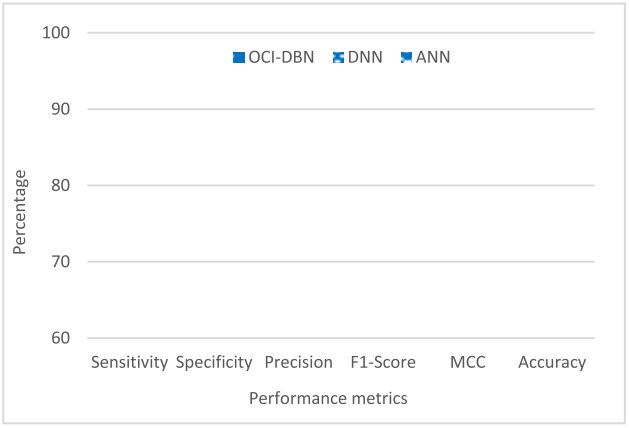

FIGURE 3. Experimental results of OCI-DBN, DNN, and ANN with a full feature set.

to improve the heart disease prediction performance. The experimental results of applying these three models on the full feature set are shown in Figure 3.

From Figure 3, the proposed OCI-DBN still achieves better accuracy and has better evaluation metrics than the other two models. A classification accuracy of 90.91% is achieved when applying OCI-DBN on full features, which is 3.37% higher than the DNN and 8.08% higher than the conventional ANN. If we carefully analyzed the experimental results achieved by OCI-DBN, ANN and DNN on an optimal subset of features and full feature set then it can be seen that the performance increase with the optimal subset of features for all the three models. In the case of the proposed OCI-DBN, the improvement in classification accuracy is 3.7% when using Ruzzo-Tompa-based feature selection to get the optimal subset of features. Similarly, the classification accuracy for DNN and ANN are also increased by 3.03% and 3.67% respectively when using optimal features. Hence, it is clear that the proposed feature selection method improves the performance of the model. To conclude this evidence, we show the results of other well-known feature selection techniques on the proposed OCI-DBN in Table 4. Feature selection techniques, like chi-squared, LASSO, and genetic algorithm, achieve good classification accuracies of around 90%. Chi-squared statistical model [31] has often used in different studies for feature selection. The difference between Ruzzo-Tompa and chi-squared feature selection is that the aim of Ruzzo-Tompa is to compute the fitness value for each feature in the dataset using the fitness function equation while the aim of chi-squared is to perform a test that measures the dependency between the features and the class. Another difference between the two is that Ruzzo-Tompa selects those features which have high fitness values while the chi-squared selects those features which are more dependent on the class by computing chi-squared statistics between each non-negative feature and class. In the proposed Ruzzo-Tompa algorithm, there is an improvement of 1.99% in classification accuracy, which shows that the proposed approach for feature selection produces the best results and improves the performance of the system.

D. COMPARATIVE ANALYSIS OF PROPOSED OCI-DBN WITH OTHER STATE-OF-THE-ART TECHNIQUES

In this section, the experimental results and the predictive accuracy achieved by other models and techniques presented by other researchers are compared with the proposed OCI-DBN. The comparative study is conducted to further validate the performance of the proposed OCI-DBN. We compared our proposed OCI-DBN approach with existing approaches including Chi-squared-CNN [6], ANN-Fuzzy AHP [9], Adaptive weighted fuzzy ensemble [11], Stacked SVM [16], ANFIS [22], Gaussian Naive Bayes Chi-squared-GBN [32], cluster-based DT learning (CDTL) [33] in heart disease prediction. This comparative analysis is presented in Figure 4. The recently proposed approaches for heart disease

FIGURE 4. Comparative analysis of proposed OCI-DBN and other state-of-the-art techniques.

prediction achieved much better performance in terms of predictive accuracy and that makes the comparison with the proposed OCI-DBN much stronger because Cleveland heart disease dataset is used to carry out experiments in all these studies. We can see that the predictive accuracy achieved by the proposed OCI-DBN approach is 94.61% which is better than all the other recently proposed approaches. This is due to the fact that the overfitting and underfitting problem and the network optimization problem is resolved in the proposed approach.

VI. CONCLUSION AND FUTURE WORK

This study provides a novel approach for heart disease prediction. The proposed OCI-DBN uses Ruzzo-Tompa for finding the optimal subset of features and SGA for finding the optimal configuration of the DBN. The proposed OCI-DBN for heart disease prediction resolves three main problems: overfitting, underfitting, and finding an optimal network configuration. The proposed method is compared with the conventional DNN using the same optimization strategy and compared with the conventional ANN method. An analysis is also performed by comparing the results of optimized network configuration with other network configurations. The comparative analysis shows that the proposed approach outperforms the other approaches and state-of-the-art machine learning techniques. The comparison of different feature selection methods with OCI-DBN also shows that Ruzzo-Tompa-based feature selection improves the performance and outperforms the other approaches. From these experimental and comparative results, we can say that the proposed OCI-DBN improves heart disease prediction with a classification accuracy of 94.61% and can help medical practitioners to make effective decisions.

In the future, the time complexity of the proposed OCI-DBN will be investigated, which is also one of the most critical factors in healthcare. Additionally, the proposed approach will be tested on some other datasets with more instances to evaluate the effectiveness of the approach.

REFERENCES

[1] K. Polat, S. ahan, and S. Güne , "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 Syst. Appl.*, vol. 32, no. 2, pp. 625–631, Feb. 2007.
[2] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network," *Knowl.-Based Syst.*, vol. 132, pp. 62–71, Sep. 2017.
[3] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7675–7680, May 2009.
[4] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8221–8231, Nov. 2015.
[5] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the risk of heart failure with EHR sequential data modeling," *IEEE Access*, vol. 6, pp. 9256–9261, 2018.
[6] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34938–34945, 2019.
[7] J. Zhang, R. L. Lafta, X. Tao, Y. Li, F. Chen, Y. Luo, and X. Zhu, "Coupling a fast Fourier transformation with a machine learning ensemble model to support recommendations for heart disease patients in a telehealth environment," *IEEE Access*, vol. 5, pp. 10674–10685, 2017.
[8] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
[9] 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 Syst. Appl.*, vol. 68, pp. 163–172, Feb. 2017.
[10] N. Gupta, N. Ahuja, S. Malhotra, A. Bala, and G. Kaur, "Intelligent heart disease prediction in cloud environment through ensembling," *Expert Syst.*, vol. 34, no. 3, Jun. 2017, Art. no. e12207.
[11] 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," *Int. J. Speech Technol.*, vol. 48, no. 7, pp. 1739–1756, Jul. 2018.

[12] J. Vijayashree and H. P. Sultana, "Heart disease classification using hybridized ruzzo-tompa memetic based deep trained neocognitron neural network," *Health Technol.*, vol. 10, no. 1, pp. 207–216, Jan. 2020.

[13] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, Apr. 2017.

[14] J. K. Kim and S. Kang, "Neural network-based coronary heart disease risk prediction using feature correlation analysis," *J. Healthcare Eng.*, vol. 2017, pp. 1–13, Sep. 2017.

[15] G. Manogaran, R. Varatharajan, and M. K. Priyan, "Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4379–4399, Feb. 2018.

[16] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019.

[17] P. Lu, S. Guo, H. Zhang, Q. Li, Y. Wang, Y. Wang, and L. Qi, "Research on improved depth belief network-based prediction of cardiovascular diseases," *J. Healthcare Eng.*, vol. 2018, pp. 1–9, May 2018.

[18] R. Alizadehsani, M. H. Zangooei, M. J. Hosseini, J. Habibi, A. Khosravi, M. Roshanzamir, F. Khozeimeh, N. Sarrafzadegan, and S. Nahavandi, "Coronary artery disease detection using computational intelligence methods," *Knowl.-Based Syst.*, vol. 109, pp. 187–197, Oct. 2016.

[19] B. Raza, A. Sher, S. Afzal, A. K. Malik, A. Anjum, Y. J. Kumar, and M. Faheem, "Autonomic workload performance tuning in large-scale data repositories," *Knowl. Inf. Syst.*, vol. 61, no. 1, pp. 27–63, Oct. 2019.

[20] B. Raza, Y. J. Kumar, A. K. Malik, A. Anjum, and M. Faheem, "Performance prediction and adaptation for database management system workload using case-based reasoning approach," *Inf. Syst.*, vol. 76, pp. 46–58, Jul. 2018.

[21] H. Yang and J. M. Garibaldi, "A hybrid model for automatic identification of risk factors for heart disease," *J. Biomed. Informat.*, vol. 58, pp. S171–S182, Dec. 2015.

[22] A. V. S. Kumar, "Diagnosis of heart disease using fuzzy resolution mechanism," *J. Artif. Intell.*, vol. 5, no. 1, pp. 47–55, Jan. 2012.

[23] N. Liu and J.-M. Kan, "Improved deep belief networks and multi-feature fusion for leaf identification," *Neurocomputing*, vol. 216, pp. 460–467, Dec. 2016.

[24] A. Fischer and C. Igel, "Training restricted Boltzmann machines: An introduction," *Pattern Recognit.*, vol. 47, no. 1, pp. 25–39, Jan. 2014.

[25] R. O. Bonow, D. L. Mann, D. P. Zipes, and P. Libby, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. Amsterdam, The Netherlands: Elsevier, 2011.

[26] V. K and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," *Int. J. Comput. Appl.*, vol. 19, no. 6, pp. 6–12, 2011.

[27] Z. K. Pourtaheri and S. H. Zahiri, "Ensemble classifiers with improved overfitting," in *Proc. 1st Conf. Swarm Intell. Evol. Comput. (CSIEC)*, Mar. 2016, pp. 93–97.

[28] S. Maldonado, J. Pérez, R. Weber, and M. Labbé, "Feature selection for support vector machines via mixed integer linear programming," *Inf. Sci.*, vol. 279, pp. 163–175, Sep. 2014.

[29] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart diseases diagnosis using neural networks arbitration," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 12, p. 72, Nov. 2015.

[30] L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, "Decision making in advanced heart failure: A scientific statement from the American heart association," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.

[31] 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 Syst. Appl.*, vol. 137, pp. 22–28, Dec. 2019.

[32] L. Ali, S. U. Khan, N. A. Golilarz, I. Yakubu, I. Qasim, A. Noor, and R. Nour, "A feature-driven decision support system for heart failure prediction based on X^2 statistical model and Gaussian naive bayes," *Comput. Math. Methods Med.*, vol. 2019, pp. 1–8, Nov. 2019.

[33] G. Magesh and P. Swarnalatha, "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction," *Evol. Intell.*, pp. 1–11, Jan. 2020, doi: 10.1007/s12065-019-00336-0.

[34] 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 J. Transl. Eng. Health Med.*, vol. 7, 2019, Art. no. 2000410.

SYED ARSLAN ALI received the bachelor's degree from COMSATS University, Pakistan, where he is currently pursuing the master's degree in research and Science. His research interests include big data, the IoT, deep learning, and cloud computing.

BASIT RAZA received the master's degree in computer science from the University of Central Punjab, Lahore, Pakistan, and the Ph.D. degree in computer science from International Islamic University Islamabad (IIU), Islamabad, Pakistan, in 2014, and conducted his Ph.D. research from the Faculty of Computing, Universiti Technology Malaysia (UTM), Malaysia. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad. He has authored several articles in refereed journals. His research interests are ad hoc wireless and sensor networks, database management systems, the IoT, security and privacy, data mining, data warehousing, machine learning, and artificial intelligence. He has been serving as a Reviewer for prestigious journals, such as the *Journal of Network and Computer Applications*, *Physical Communication*, *Networks*, *Applied Soft Computing*, *Swarm and Evolutionary Computation*, *Swarm Intelligence*, *Applied Intelligence*, and *Future Generation Computer Systems*.

AHMAD KAMRAN MALIK received the M.S. degree in computer science from Muhammad Ali Jinnah University, Islamabad, and the Ph.D. degree from the Vienna University of Technology (TU-Wien), Austria. He has been teaching with Quaid-I-Azam University. He has also been teaching and supervising computer science students at the bachelor's and master's levels, since 1999. He is currently working as an Assistant Professor with COMSATS, Islamabad. His current research interests include social network analysis, access control, and collaborative systems.

AHMAD RAZA SHAHID received the Ph.D. degree in computer science from York, U.K., in 2012. During his Ph.D. studies, he worked on automatically building a WordNet for four languages: English, German, French, and Greek. Since his Ph.D. degree, he has been working in the areas of computer vision and pattern recognition, machine learning, and natural language processing. A few of the problems that he has worked on include cancer detection, pedestrian detection, driver fatigue detection, and data mining. He is currently working as an Assistant Professor with the COMSATS Institute of Information Technology, Islamabad, Pakistan.

MUHAMMAD FAHEEM received the B.Sc. degree in computer engineering from the Department of Computer Engineering, University College of Engineering and Technology, Bahauddin Zakariya University Multan, Pakistan, in 2010, and the M.S. degree in computer science from the Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Malaysia, in 2012, where he is currently pursuing the Ph.D. degree with the Faculty of Computing, School of Computer Science. In the past, he has served as a Lecturer with the Comsats Institute of Information and Technology, Pakistan, from 2011 to 2012. Since 2013, he has been working as a Lecturer with the Computer Engineering Department, Abdullah Gul University, Kayseri, Turkey. He has authored several articles in refereed journals His research interests include smart grid communications, energy harvesting, underwater acoustic communications, cognitive radio sensor networks, and information storage and retrieval architecture in sensor networks. He has been serving as a reviewer for numerous journals, such as the *Journal of Network and Computer Applications*, *Ad-hoc Networks*, *Australian Journal of Electrical Engineering*, the *International Journal of Computer, Communication, and Control*, *Computer Standards and Interfaces*, IEEE Access, the IEEE Transactions on Vehicular Technology, the *IEEE Communications Magazine*, and *Future Generation Computer Systems*.

HANI ALQUHAYZ received the bachelor's in computer science and the master's degree in information systems management from King Saud University and the Ph.D. degree in computer science from De Montfort University, U.K. He is currently an Assistant Professor with the Computer Science Department, College of Science, Majmaah University, Saudi Arabia. He has authored several articles in high-impact journals, such as the IEEE Access, *Sensors*, and *Wireless Communications and Mobile Computing*. His research interests are in wireless security, scheduling, image processing, the IoT, security and privacy, data mining, machine learning, and artificial intelligence.

YOGAN JAYA KUMAR received the bachelor's and master's degrees in mathematical science from Universiti Sains Malaysia (USM), in 2003 and 2005, respectively, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia, in 2014. His current research involves in the field of text mining, information extraction, and AI applications.