IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 23, 2019, accepted October 8, 2019, date of publication November 7, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952107

An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection

ASHIR JAVEED[1], SHIJIE ZHOU[1], LIAO YONGJIAN[1], IQBAL QASIM[2], ADEEB NOOR[3], AND REDHWAN NOUR[4]

[1]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Computer Science, University of Science and Technology at Bannu, Bannu 28100, Pakistan
[3]Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 80221, Saudi Arabia
[4]Department of Computer Science, Taibah University, Medina 42353, Saudi Arabia

Corresponding authors: Shijie Zhou (sjzhou@uestc.edu.cn)

This work was supported in part by the Sichuan Science and Technology Program under Grant 2018GZ0085 and Grant 2019YFG0399 and in part by the Deanship of Science Research (DSR) at King Abdulaziz University, Jeddah, under Grant DF-608-611-1441.

ABSTRACT Heart failure is considered one of the leading cause of death around the world. The diagnosis of heart failure is a challenging task especially in under-developed and developing countries where there is a paucity of human experts and equipments. Hence, different researchers have developed different intelligent systems for automated detection of heart failure. However, most of these methods are facing the problem of overfitting i.e. the recently proposed methods improved heart failure detection accuracy on testing data while compromising heart failure detection accuracy on training data. Consequently, the constructed models overfit to the testing data. In order, to come up with an intelligent system that would show good performance on both training and testing data, in this paper we develop a novel diagnostic system. The proposed diagnostic system uses random search algorithm (RSA) for features selection and random forest model for heart failure prediction. The proposed diagnostic system is optimized using grid search algorithm. Two types of experiments are performed to evaluate the precision of the proposed method. In the first experiment, only random forest model is developed while in the second experiment the proposed RSA based random forest model is developed. Experiments are performed using an online heart failure database namely Cleveland dataset. The proposed method is efficient and less complex than conventional random forest model as it produces 3.3% higher accuracy than conventional random forest model while using only 7 features. Moreover, the proposed method shows better performance than five other state of the art machine learning models. In addition, the proposed method achieved classification accuracy of 93.33% while improving the training accuracy as well. Finally, the proposed method shows better performance than eleven recently proposed methods for heart failure detection.

INDEX TERMS Heart failure, hyperparameters optimization, feature selection, random search algorithm, grid search algorithm.

I. INTRODUCTION

Abnormal functionality of the heart due to any cause is known as heart disease. There are different kinds of heart disease. The most common types are heart failure (HF) and Coronary Artery Disease (CAD). The major cause of heart failure (HF) is due to the blockage or narrowing down of coronary arteries. Coronary arteries also supply blood to the heart. CAD is a prevalent kind of heart disease and well-known source of heart attacks in the world [1], [2]. HF is an expeditious healthcare problem [3] of the modern world and it has been reported that 26 million adults around the globe are suffering from HF [4]. It is reported that approximately 31% of the overall deaths around the world are due to HF [3]. HF disease was the leading cause of death globally in 2005, responsible for 17.5 million deaths, more than 80% of which occurred

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

in low and middle income countries such as China [5]. Furthermore, HF disease is expected to contribute to over 80% of the predicted future increase in mortality in developing nations [6]. Heart disease-related morbidity and mortality in women have also increased dramatically [7], and HF is the leading cause of death in China in general [8].

Different factors have been identified through medical and clinical research as the risk factors for heart failure and CAD. The risk factors are classified into two main classes. The first class includes risk factors that cannot be changes e.g., age, gender and family history. The second class contains risk factors that can be changed e.g., smoking, eating habit and high cholesterol. Hence, the risk factors belonging to second class can be eradicated or controlled by changing lifestyle and through medication [2].

In order to diagnose or predict HF disease, different methods are utilized by medical practitioners. Presently, Angiography is the most widely used practice for the diagnosis of CAD by physician that is believed to be the most precise method CAD detection or HF due to CAD [9]. However, higher cost and considerable side effects of angiography are the main limitations in the way of diagnosis of HF or CAD. Moreover, diagnose of CAD requires a thorough analysis of different factors which makes the physician's job tough. These problems inspire the development of non- intrusive mechanism for the detection of heart failure. Additionally, conventional approaches of heart failure diagnosis focused on the examination of patient's medical history, evaluation of various symptoms by a cardiologist and physical cross-examination report. Therefore, the conventional methods used for heart failure diagnosis are time consuming and may yield erroneous diagnostic results due to human errors [10], [11]. Thus, to avoid these problems we need to develop an automated learning system for the efficient and fast diagnosis of heart failure.

From the literature, it has been studied that various automatic diagnosis system have been proposed in the past by exploiting various machine learning techniques for the detection of heart failure e.g., diagnostic systems have been developed based on naive Bayes(NB), support vector machine (SVM), stacked and optimized SVMs, fuzzy logic, artificial neural network (ANN), deep neural network (DNN) and ensembles of ANN [2], [9], [12]–[19], [19]–[27]. Khemphila et al. proposed a classification approach based on Multi-Layer Perceptron (MLP) along with Back-propagation learning algorithm and biomedical test values to diagnose the heart failure through a feature selection algorithm [28]. Total number of thirteen features are reduced to eight features through a feature selection algorithm. The accuracy of training dataset was 89.56%, while for data validation the accuracy of 80.99% was reported. Paul et al. introduced a fuzzy decision support system (FDSS) for the detection of heart failure [29]. The proposed system obtained the accuracy of 80%. Verma et al. proposed a novel hybrid method for Coronary artery disease (CAD) diagnosis [30]. The accuracy of the proposed approach was 88.4%. The proposed model improved the efficiency of classification algorithms by 11.4% for Cleaveland dataset. Shah et al., proposed a technique based on the feature extraction for reducing features dimensions [31]. The proposed approach used Probabilistic Principal Component Analysis (PPCA). The accuracy of the proposed technique for Cleveland dataset was 82.18%. Dwivedi et al. tested the performance of different machine learning techniques for the prediction of heart failure [32]. The highest classification accuracy of 85% was reported based on logistic regression. Amin et al. evaluated different data mining techniques and identification of significant features for predicting heart failure [33]. From experimental results, it was observed that the best performance of the data mining technique for classification accuracy was 87.4% for the heart failure prediction.

Recently, Resul et al. achieved heart failure detection accuracy of 89.01%, sensitivity of 80.95% specificity of 95.91% by developing ensemble of neural network model [2]. Samuel et al. further improved the heart failure prediction accuracy to 91.10% by proposing a novel hybrid decision support by integrating ANN and fuzzy analytic hierarchy (Fuzzy-AHP) [34]. Furthermore, Ali et al. improved the heart failure prediction accuracy to 92.22% by developing a novel diagnostic system that hybridized two SVMs [9]. The first SVM was used for selection of important features and the second SVM was used for prediction purposes. Both the models were optimized by using a new searched algorithm. Most recently, Paul et al. further improved the heart failure prediction accuracy to 92.31% by proposing an adaptive weighted fuzzy system ensemble method [35]. However, the HF prediction accuracy still needs considerable amount of improvement.

From the literature survey, it is clear that different researchers exploited features preprocessing methods for improving the prediction capabilities of machine learning models. Motivated by this fact, in this paper, we try to further improve the HF detection accuracy. We develop a learning system that hybridizes two algorithms. The first algorithm is random search algorithm (RSA) that is used to search a subset of features having complementary information about the heart failure prediction. The second algorithm is a random forest (RF) algorithm. The two algorithms work in connection with each other thus constitute one hybrid learning system named RSA-RF. In the first iteration, the RSA generates a boolean mask vector (random boolean mask) with only one True value (at a location randomly generated). The size of boolean mask vector is equal to the size of feature vector. The boolean mask vector is logically anded with the feature vector as a result only that feature is extracted or selected whose corresponding location in boolean vector contains True value. In next iteration, another boolean mask is generated with two True values at two locations randomly chosen and the same process is repeated resulting in a subset of features with only two features. The same process is repeated until the number of True values (generated at random locations) approach the size of the feature vector. It is important to note that each time, the obtained subset of features is applied to RF algorithm

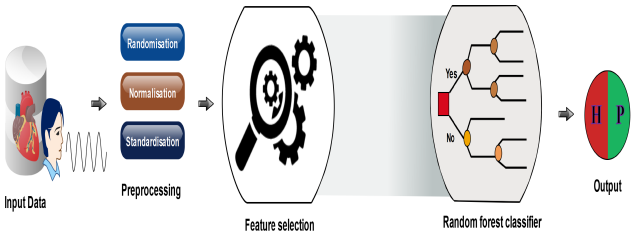


FIGURE 1. Block diagram of the proposed diagnostic system.

for classification and the optimal hyperparameter of RF are searched out using grid search algorithm. The working of the proposed RSA-RF method is more clearly depicted in FIGURE 1.

The main contributions of this study are summarized as follows:

1. An intelligent learning system RSA-RF is proposed for the automated detection of heart failure.

2. The proposed RSA-RF model was proposed and developed for the first time for the heart failure detection. Previously, RSA algorithm have shown successful applications in searching optimal hyperparameters of a model. This paper presents its application in searching optimal subset of features.

3. The developed learning system improves heart failure prediction of conventional random forest model by 3.3% and shows better performance than eleven recently proposed methods and other state of the art machine learning models for heart failure detection.

4. Moreover, the proposed method shows lower time complexity as it reduces the number of features.

The remaining of the paper is organized as follows: In section II, we give details about the data and methods. In section III, we describe the evaluation and validation methods used in this paper. The last two sections are about experiments, discussion and conclusion.

II. MATERIALS AND METHODS

A. DATASET DESCRIPTION

In this study, we utilized an online available database of heart failure, the dataset recognized as Cleveland heart failure dataset that is freely available on machine learning UCI repository. Total number of instances in the dataset are 303 among them 297 instance have complete attributes information while six instances have missing details. There are 76 raw features in the dataset in its genuine form. Although, published work in the past refer to only 13 features which are described in the TABLE 1. Moreover, in the past, published work utilized only those instances which have no missing attributes in the experiments. Additionally, a binary label has been used in this study where label value 0 means no HF disease and label value of 1 means, HF disease is present.

B. PROBLEM FORMULATION AND PROPOSED SOLUTION

In machine learning, different types of data mining algorithms also known as features selection algorithms which

TABLE 1. Description of features of the dataset.

Feature No	Feature Description	Feature Abbreviations	Features Code
1	Age (Years)	AGE	A_1
2	Sex	SEX	A_2
3	Chest Pain Type	CPT	A_3
4	Resting Blood Pressure	RBP	A_4
5	Serum Cholesterol	SCH	A_5
6	Fasting Blood Sugar	FBS	A_6
7	Resting Electrocardiographic Results	RES	A_7
8	Maximum Heart Rate achieved	MHR	A_8
9	Exercise Induced Angina	EIA	A_9
10	Old Peak	OPK	A_{10}
11	Peak Exercise Slope	PES	A_{11}
12	Number of Major Vessels Colored by Fluoroscopy	VCA	A_{12}
13	Thallium Scan	THA	A_{13}

are used to improve the performance of machine learning models as well as to reduce their training time. Thus, making the diagnosis process fast and accurate. One way to select a subset of features that will have more discriminatory information about the two classes (in case of binary classification), is to evaluate all possible combinations of features (i.e., using exhaustive search methodology). But, it is impractical because it will take too much time to evaluate all the combinations of features by training the model and testing it with each subset. For examples, if a dataset has 30 features, we will have $2^{30} = 1073741824$ possible combinations or subsets of features for evaluation. If our machine learning model takes just one minute in training and testing phases, then it will take 1073741824 minutes or 17895697.0667 hours or 745654.04 days. Hence, it is an impractical solution. To avoid this problem, we propose random search method for searching subset of features. The method is faster and controllable. That is if we could not find desirable results, we can repeat the process for as many times as we want. It is important to discuss that the use of random search algorithm (RSA) for features selection got motivation from the study conducted by Bergstra et al. in [36]. In the study they proposed random search algorithm for hyperparameters optimization of machine learning models. However, we propose the random search for searching out optimal subset of features.

In the first iteration, the RSA generates a boolean mask vector (random boolean mask) with only one True value (at a location randomly generated). The size of boolean mask vector is equal to the size of feature vector. The boolean mask vector is logically anded with the feature vector as a result only that feature is extracted or selected whose corresponding location in boolean vector contains True value. In next iteration, another boolean mask is generated with two True values at two location randomly chosen. The boolean mask vector is anded with the feature vector and a subset of features is generated having only two features. The same process is repeated for N-1 iteration resulting in N-1 subsets of features,

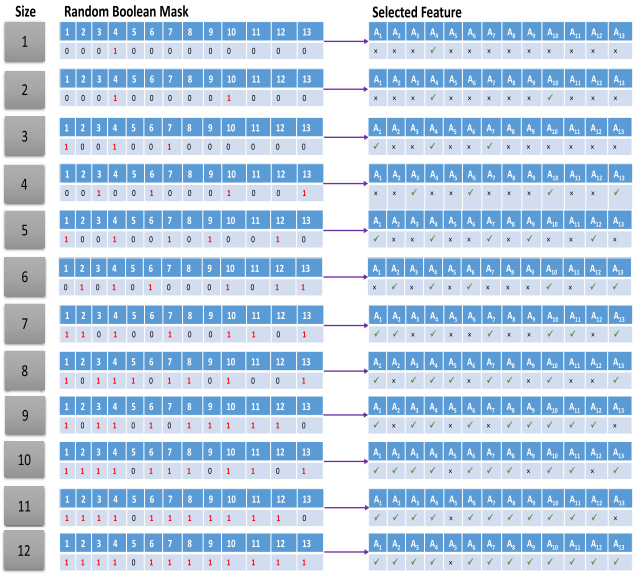


FIGURE 2. Block diagram of the proposed RSA based feature selection.

where N denotes the size of full features in the original feature space. In order to better understand the working of RSA, the process is depicted in FIGURE 2. It is important to note that each of the produced subset of features by RSA is applied to RF algorithm for classification. The formulation and working of RF algorithm is briefly discussed as follows:

Random forest is an ensemble model or classifier $p(m, n)$ where n are identically scattered random vectors and every tree participate for the selection of most prominent class at input m. For Q input sample where Q is mostly identical to the training set samples, each is to derive q samples, after that based on column sampling f features are derived from F features. N times was chosen randomly and n training sets will be attained as $S_1, S_2, \ldots, S_k$. The decision tree T_1, $T_2, .., T_k$ will be obtained from the corresponding training sets. Every tree in the forest is entirely grown apart from shearing.

During classification, large number of decision trees take part in the formation of random forest. Two hyperparameters are very decisive for classification that is the number of decision tree i.e., E, the number of trees forming the forest and D, depth of each tree [37]. In this paper, we utilized exhaustive search strategy to search out optimal E and D that would ensure better performance by random forest model.

Furthermore, a random forest model is constructed and a new sample is insert into the model. Moreover, decision tree will analysis and figure out the class of new sample.The final classification of a sample can be determined based on the overall decision trees vote in the forest. The random forest formation trees are assembled from resampling by returning to training data and known as bootstrap. Bootstrap is a simple and practical technique used for model integration with replacement method [38]. Fixed number of samples from training set are extracted through bootstrap random sampling. After sampling, the number of samples are return back to the training set. The extracted samples forms a new batch of bootstrap sample. Furthermore, there is a probability of the collected samples are to be sampled after return to the training set. Therefore, collected samples in the past are liked to be sampled after putting back.

For example, a random sample with r samples. We can calculate the probability of the sample whether it is collected every time or not through $\frac{1}{r}$ and $(1 - \frac{1}{r})$. If random sampling is performed R time, then the probability of sample whether it is selected or not is given as $(1 - \frac{1}{r})R$ where R convergent to ∞, $(1 - \frac{1}{r})R$ converge to $\frac{1}{e} = 0.368$. Moreover, there will exist mirror sample. Furthermore, $(\frac{1}{3})$ instance will escape into new samples. The data that is ignored during extraction is known as Out-of-Bag instances. Error associated with out of bag is termed as OOB error. Mathematically this error can be express as:

$$OE = \frac{X}{NB} \tag{1}$$

In Equation 1, X illustrates the value of error for testing NB Number of Data whereas NB show the number of OOB and acknowledged as a class of each data.

A decision tree is formed from Gini Index that elaborates the impurity level of the model through CART algorithm. A lower value of Gini index shows minor impurity. A smaller Gini index means a lower impurity. For classification issue, the probability of the N-th category is p_n for N classes, the Gini index formula is illustrated as follows:

$$Gini(D) = \sum_{n=1}^{n} Pn(1 - Pn) = 1 - \sum_{n=1}^{n} Pn^2 \tag{2}$$

In decision tree for feature selection Gini index is used. Equation (3) shown the formula:

$$\Delta Gini(A) = Gini(D) - Gini_A(D') \tag{3}$$

The maximum value of the Gini index will be preferred for split attribute and node for split condition.

Decision tree will resemble as over-fitting. Over-fitting rate can be reduced through the pre-pruning and post-pruning methods. The growth of decision trees may premature by the pre-pruning while post-pruning can form a superior result. Furthermore, decision is grown completely without pruning.

In order to select a subset of features as optimal subset out of the N-1 features vectors (or subsets) produced by RSA, we apply a subset of feature to the RF algorithm for classification and the optimal hyperparameters of RF are searched out using grid search algorithm for that subset of features. Next, another subset of features is applied to RF algorithm and again the optimized hyperparameters of RF are searched by grid searched algorithm. The same process is repeated for each of the produced subset of features. Finally, that subset of features is selected and reported which produces the highest heart failure prediction accuracy.

Algorithm 1 RSA-RF

Input: {N_F: Full features set, E, D Hyperparameters of RF}

Output: {F_{sub}: Optimal Subset of features, Optimal E and D }

1. Initialize counter to 1 and Best_Accuracy to 0.
2. Generate a random boolean mask vector with size $N_F - 1$ and number of true values equal to counter (generated at random locations)
3. Construct a subset of features (F_{sub}) with only those features present at location of boolean mask vector with true value.
4. Initialize
5. **for** $D = 1 : D_{max}$
6. **for** $E = 1 : E_{max}$
7. Evaluate Accuracy for each combination of E and D of random forest.
8. if ($Accuracy > Best\_Accuracy$)
   Begin if
   $Best\_Accuracy = Accuracy$
   Report F_{sub}, E, D
   End if
9. Increment counter
10. Repeat sept 2 until size of F_{sub} approaches N_F.

C. VALIDATION SCHEMES

In data mining and machine learning, different types of validation methods are exploited to measure the performance of a developed method. Among them train-test holdout is one of the most commonly used method. However, different types of data partitioning schemes are usually utilized during train-test holdout methodology. Following the approach of most of the previous studies like Das *et al.* [2], Ali *et al.* [25], Paul *et al.* [35] and Ali *et al.* [9], we utilized 70-30% data partitioning method. That is we holdout 30% of the dataset for testing purposes and 70% for the training of the proposed supervised learning system. The main motive behind choosing exactly same data portioning protocol was to better compare our proposed method with the recently proposed method.

D. EVALUATION METRICS

In this study, accuracy, sensitivity, specificity and MCC are the evaluation metrics exploited to measure the performance of the proposed model. Accuracy can be described as the correctly classified subjects in the test dataset. The percentage of precisely classified patients in the data are known as sensitivity. Furthermore, specificity elaborates the accurate classification of the healthy subjects.

Mathematically evaluation metrics derive as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP stand for number of true positives, FP denotes number of false positives, TN express as number of true negatives and FN show number of false negatives.

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

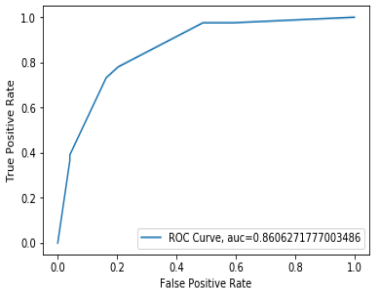
$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

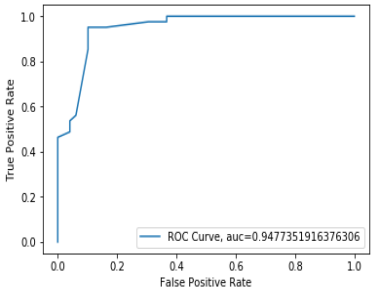
Statistical analysis is the key element for performance measurement of any proposed predictive model. We used Matthews correlation coefficient (MCC) in statistical analysis of binary classification. Test's accuracy is determined through (MCC) which has a values ranging between -1 and 1. Where 1 denotes the precise predictions and -1 indicate inferior predictions.

TABLE 2. Classification accuracy at different values of hyperparameters of a conventional random forest. E: number of estimators. D: Depth hyperparameter. Sen: Sensitivity. Acc: Accuracy. Spec: Specificity.

E	D	Acc(%)	Sen(%)	Spec(%)	Mcc
17	2	87.77	82.92	91.83	0.753
27	2	88.88	80.48	95.91	0.780
27	12	84.44	82.92	85.71	0.686
27	22	84.44	82.92	85.71	0.686
97	**2**	**90.0**	**82.92**	**95.91**	**0.801**



(a) ROC plot of simple Random Forest

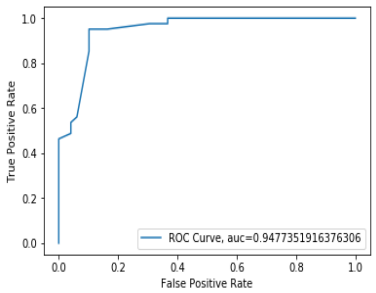


(b) ROC plot of the proposed RSA-RF method

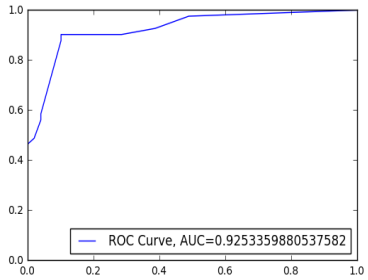
FIGURE 3. Comparative ROC plots.

III. EXPERIMENTAL RESULTS AND DISCUSSION

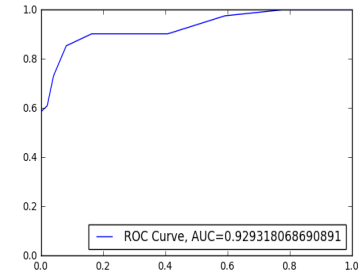
In order to evaluate the effectiveness of the proposed method rigorously, three different types of experiments are performed



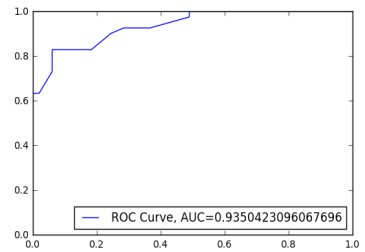
(a) ROC chart of the proposed model



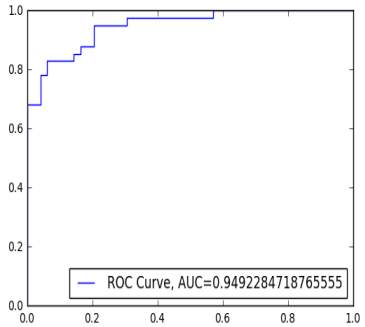
(b) ROC chart of Adaboost ensemble model



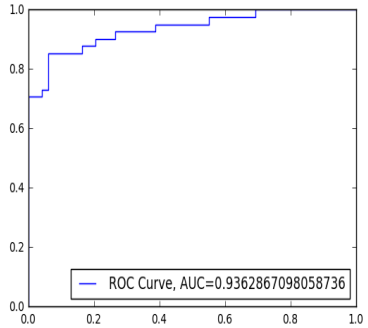
(c) ROC chart of extra tree ensemble model



(d) ROC chart of random forest



(e) ROC chart of linear SVM model



(f) ROC chart of SVM (RBF) model

FIGURE 4. ROC charts of the proposed model and optimized SVM and ensemble models.

TABLE 3. Classification accuracy of different subset of features at optimal hyperparameters of random forest. E: number of estimators. D: Depth hyperparameter. Sen: Sensitivity. Acc: Accuracy. Spec: Specificity. N: number of features returned by RSA.

N	E	D	Acc(%)	Sen(%)	Spec(%)	Mcc
1	1	2	78.88	75.60	81.63	0.573
2	2	2	80.00	85.36	75.51	0.606
3	8	2	82.22	80.48	83.67	0.641
4	8	2	78.88	85.36	73.46	0.587
5	92	2	76.66	75.60	77.55	0.530
6	92	12	76.66	75.60	77.55	0.530
7	**73**	**4**	**93.33**	**95.12**	**89.79**	**0.846**
8	95	2	88.88	80.48	95.91	0.780
9	95	2	85.55	80.48	89.79	0.708
10	95	2	88.88	80.48	95.91	0.780
11	96	12	84.44	78.04	89.79	0.686
12	96	12	84.44	78.04	89.79	0.686

on the heart heart failure dataset. In the first experiment, conventional random forest model is developed and optimized using grid search algorithm. In the second experiment, the proposed RSA-RF method is developed while in the third experiment other state of the art machine learning models are developed using the same heart heart failure data. All computations were performed on Intel (R) Core (TM) i3-2330M CPU @2.20GHz with 64bit windows 7 as the operating system. All the experiments are performed using Python software package.

TABLE 4. Classification accuracy of other machine learning model on the HF data.

Classifier	Parameters	Acc(%)	Spec(%)	Sen(%)	MCC
RF	E_s =50	88	93.87	82.92	0.777
ET	E_s =11	88	89.79	87.80	0.777
Adaboost	E_s =4	88	89.79	87.80	0.776
SVM(RBF)	C=5, G	90	93.87	85.36	0.799
SVM(Lin)	C=0.055	90	93.87	85.36	0.799

A. RANDOM FOREST MODEL DEVELOPED FOR HEART HEART FAILURE PREDICTION

In this experiment, we develop only random forest model which is implemented in Python programming package.

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

Study (Year)	Method	Accuracy(%)
Proposed Method (2019)	Random Search Algorithm + Random Forest	93.33
Dwivedi (2018) [32]	Naive Bayes, logistic regression, SVM etc	85
Paul et al. (2018) [29]	Adaptive Weighted Fuzzy system ensemble	92.31
Amin et al. (2019) [33]	feature selection and voting method	87.4
Ali et al. (2019) [9]	Stacked SVMs	92.22%
Senthil Kumar (2012) [39]	Fuzzy resolution mechanism	91.83
Samuel et al. (2017) [34]	ANN-Fuzzy-AHP	91.10
Shah et al. (2017) [31]	Probabilistic principle component analysis + SVM	82.18
Senthil Kumar (2011) [40]	ANFIS	91.18
Khemphila (2011) et al. [28]	Information gain + neural network	80.99
Jankowski and Kadirkamanathan (1997) [41]	IncNet	90.00
Resul et al. (2009) [2]	Neural network ensembles	89.01
Kahramanli and Allahverdi (2008) [42]	Hybrid neural network system	86.80
K. Polat et al. (2006) [44]	Fuzzy-AIRS-KNN based system	87.00
Ozsen et al. (2005) [43]	Kernel functions with AIS	85.93
K. Polat et al. (2005) [47]	AIRS	84.50
Ster and Dobnikar [45]	Naive Bayes	82.50-83.40
Ster and Dobnikar [45]	Linear discriminant Analysis	84.50
Ster and Dobnikar [45]	Fisher discriminant Analysis	84.20
WEKA, RA [46]	Naive Bayes	83.60
Newton Cheung [47]	BNNF	80.96
Newton Cheung [47]	BNND	81.11
Newton Cheung [47]	Naive Bayes	81.48
Newton Cheung [47]	C4.5	81.11
Robert Detrano [46]	Logistic regression	77.00
WEKA, RA [46]	K*	76.70
WEKA, RA [46]	IB1c	74.00
WEKA, RA [46]	1R	71.40
WEKA, RA [46]	T2	68.10
ToolDiag, RA [46]	MLP + BP	65.60
WEKA, RA [46]	FOIL	64.00
ToolDiag, RA [46]	RBF	60.00
WEKA, RA [46]	InductH	58.50
ToolDiag, RA [46]	IB1-4	50.00

The model hyperparameters are tuned using grid search algorithm. The heart failure prediction accuracy, sensitivity, specificity and MCC at tuned hyperparameters and untuned hyperparameters are given in Table 2. It can be seen that best heart failure detection accuracy of 90% is obtained using tuned hyperparameters. The obtained accuracy is comparable to the accuracy achieved by optimized SVM (see Table 4).

B. THE PROPOSED RSA-RF DEVELOPED FOR HEART FAILURE PREDICTION

In this subsection, we discuss the development of the proposed RSA-RF method. The dataset is initially supplied to the random search algorithm (RSA), which produced different subsets of features of different sizes from 1 to N-1, where N denotes the size of full features in the dataset. For each subset of features, the optimized version of random forest model is obtained by exploiting exhaustive grid search algorithm. For each subset of features, the optimized random forest and the obtained performance is reported in Table 3. In the table, it can be seen that the best heart failure detection of 93.33% is obtained using a subset of features having only 7 features. From comparison of Table 2 and Table 3, it is proved that the RSA-RF method shows 3.3% of higher heart failure detection accuracy than conventional random forest model.

In this experiment, we also depicted the ROC charts for the two types of experimental settings i.e. first setting simple random forest model and in second experimental setting we developed the proposed RSA-RF method. The important parameter in the ROC plot is area under the curve (AUC), that is a model having more AUC will be considered better one. Figure 3 (a) shows the ROC plot of simple random forest model while Figure 3 (b) shows the ROC plot of the proposed RSA-RF. The AUC for simple random forest is 0.860 while the AUC for the RSA-RF is 0.947. Thus, the proposed RSA-RF is an improved version of simple random forest.

C. OTHER STATE OF THE ART MACHINE LEARNING MODELS FOR HEART FAILURE PREDICTION

In this subsection, we consolidate the effectiveness of the developed RSA-RF by conducting its comparison with other state of art machine learning models namely adaptive boosting or Adaboost model, extra tree ensemble model, randomized tree model, support vector machine with linear

kernel function and support vector machine with RBF kernel function. All these models were optimized using the same exhaustive grid search methodology. The performance of each of these models at optimized hyperparameters is tabulated in Table 4. The main reason for comparing the performance of the proposed method with these models is that these models have shown state of the art performance on many disease detection problems. It can be seen from the table that the best performance is achieved by support vector machine (both with linear and RBF kernels). They achieved 90% of accuracy while the proposed method accuracy is 2.2% better than the performance of the SVM. Similarly, the performance of extra tree classifier and Adaboost classifier is 88%. Hence, the performance of the proposed method is 4.2% higher than the performance of these ensemble classifiers.

After validating the performance of the proposed method from classification accuracy aspect, we further analyze its performance by utilizing ROC plots and area under the curve (AUC). The ROC plots of all the models are depicted in Figure 4. It can be seen in the figure that the AUC of the ROC of the proposed method is 0.947. The AUC is 0.925, 0.949, 0.936, and 0.929 for Adaboost, SVM linear, SVM RBF and extra tree classifiers, respectively. Thus, the performance improvement of the proposed method is also validated from ROC plots and AUC aspects. It can be noticed that SVM linear and the proposed method have similar AUC values but the classification of the proposed method is 2.2% better than SVM linear, hence, the proposed is a better candidate for HF detection than SVM linear model.

D. PREVIOUSLY PROPOSED METHOD FOR HEART FAILURE DETECTION

In this subsection, we report some well known machine learning based methods proposed for heart failure detection. A brief description of these methods is tabulated in TABLE 5. It is important to note the the performance of our proposed method is better than the performance of the recently proposed method by Ali et al. in [9] in two ways. First, our proposed method shows 93.33% accuracy on the testing dataset. And secondly, the method reported by Ali *et al.* has achieved very low rate of training accuracy i.e. 85%, however our method also improves the training accuracy to 88% thus, the proposed RSA-RF method has better generalization than the method proposed in [9]. Additionally, Ali *et al.* achieved the classification accuracy of 92.22 using 9 features while our method utilizes only 7 features. Hence, the proposed method is also less complex than the method of Ali *et al.*

IV. CONCLUSION

In this study, we highlighted the problem of overfitting in the recently proposed methods for heart failure prediction and proposed a novel learning system to facilitate the heart failure prediction. The learning system hybridizes two algorithms. The first algorithm is a random search algorithm which is used to search out subset of features having complementary information about heart failure. The second algorithm was random forest which is used to predict heart failure based on the selected subset of features. It was shown that the proposed RSA-RF learning system improves the performance of random forest model by 3.3%. Additionally, the proposed learning system shows better performance than eleven recently proposed methods for heart failure detection and other well known machine learning models. It was also observed that the proposed system reduces the time complexity of the machine learning models by reducing the number of features. From the experimental results, we can conclude that the proposed learning system can help the physicians to improve the quality of heart failure detection.

NOMENCLATURE

1. Random Search Algorithm (RSA)
2. Cardiovascular Disease (CVD)
3. Coronary Artery Disease (CAD)
4. Heart Failure (HF)
5. Naive Bayes(NB)
6. Support Vector Machine (SVM)
7. Artificial Neural Network (ANN)
8. Deep Neural Network (DNN)
9. Random Forest (RF)
10. Random Search Algorithm- Random force (RSA-RF)
11. University of California, Irvine (UCI)
12. Out of Bag(OOB)
13. Matthews Correlation Coefficient (MCC)
14. Receiver Operating Characteristic Curve (ROC)
15. Accuracy (AUC)
16. Radial Basis Function (RBF)

ACKNOWLEDGMENT

The authors gratefully acknowledge Sichuan Science and Technology Program and DSR technical and financial support.

REFERENCES

[1] R. O. Bonow, D. L. Mann, D. P. Zipes, and P. Libby, *Braunwald's Heart Disease E-Book: A Textbook of Cardiovascular Medicine*. Amsterdam, The Netherlands: Elsevier, 2011.
[2] 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, 2009.
[3] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," *Nature Rev. Cardiol.*, vol. 8, no. 1, pp. 30–41, 2011.
[4] A. P. Ambrosy, G. C. Fonarow, J. Butler, O. Chioncel, S. J. Greene, M. Vaduganathan, S. Nodari, C. P. S. Lam, N. Sato, A. N. Shah, and M. Gheorghiade, "The global health and economic burden of hospitalizations for heart failure lessons learned from hospitalized heart failure registries," *J. Amer. College Cardiol.*, vol. 63, no. 12, pp. 1123–1133, Apr. 2014.
[5] R. Joshi, S. Jan, Y. Wu, and S. MacMahon, "Global inequalities in access to cardiovascular health care: Our greatest challenge," *J. Amer. College Cardiol.*, vol. 52, no. 23, pp. 1817–1825, Dec. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0735109708031173
[6] R. Ariely, K. Evans, and T. Mills, "Heart failure in china: A review of the literature," *Drugs*, vol. 73, no. 7, pp. 689–701, May 2013.
[7] B. M. Popkin, "Will China's nutrition transition overwhelm its health care system and slow economic growth?" *Health Affairs*, vol. 27, no. 4, pp. 1064–1076, Jul./Aug. 2008.

[8] J. He, D. Gu, X. Wu, K. Reynolds, X. Duan, C. Yao, J. Wang, C.-S. Chen, J. Chen, R. P. Wildman, M. J. Klag, M. P. H, and Paul K. Whelton, "Major causes of death among men and women in China," *New England J. Med.*, vol. 353, no. 11, pp. 1124–1134, Sep. 2005.

[9] 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.

[10] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron-based medical decision support system for heart disease diagnosis," *Expert Syst. Appl.*, vol. 30, no. 2, pp. 272–281, 2006.

[11] K. Vanisree 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, Apr. 2011.

[12] O. W. Samuel, M. O. Omisore, and B. A. Ojokoh, "A Web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4164–4171, 2013.

[13] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 24, no. 1, pp. 27–40, 2012.

[14] I. 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 Syst. Appl.*, vol. 37, no. 4, pp. 3177–3183, Apr. 2010.

[15] S. B. Patil and Y. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," *Eur. J. Sci. Res.*, vol. 31, no. 4, pp. 642–656, 2009.

[16] 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 Syst. Appl.*, vol. 32, no. 2, pp. 625–631, Feb. 2007.

[17] 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, 2015.

[18] N.-S. Tomov and S. Tomov, "On deep neural networks for detecting heart disease," Aug. 2018, *arXiv:1808.07168*. [Online]. Available: https://arxiv.org/abs/1808.07168

[19] 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, 2018.

[20] A. D. Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," *Comput. Methods Programs Biomed.*, vol. 138, pp. 117–126, Jan. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260716306605

[21] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, "Reliable Parkinson's disease detection by analyzing handwritten drawings: Construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model," *IEEE Access*, vol. 7, pp. 116480–116489, 2019.

[22] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741741930452X

[23] Y. Chen, T. Zhang, Z. Luo, and K. Sun, "A novel rolling bearing fault diagnosis and severity analysis method," *Appl. Sciences*, vol. 9, no. 11, p. 2356, Jun. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/11/2356

[24] 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.

[25] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on X^2 statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34938–34945, 2019.

[26] I. Paschalidis, "How machine learning is helping us predict heart disease and diabetes," *Harvard Bus. Rev.*, May 2017.

[27] M. A. Abushariah, A. A. Alqudah, O. Y. Adwan, and R. M. Yousef, "Automatic heart disease diagnosis system based on artificial neural network (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) approaches," *J. Softw. Eng. Appl.*, vol. 7, no. 12, p. 1055, 2014.

[28] A. Khemphila and V. Boonjing, "Heart disease classification using neural network and feature selection," in *Proc. 21st Int. Conf. Syst. Eng.*, Aug. 2011, pp. 406–409.

[29] 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 *Proc. 5th Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2016, pp. 145–150.

[30] L. Verma, S. Srivastava, and P. Negi, "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data," *J. Med. Syst.*, vol. 40, no. 7, p. 178, Jul. 2016.

[31] 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," *Phys. A, Stat. Mech. Appl.*, vol. 482, pp. 796–807, Sep. 2017.

[32] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," *Neural Comput. Appl.*, vol. 29, no. 10, pp. 685–693, 2018.

[33] M. S. Amin, Y. K. Chiam, K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics Inform.*, vol. 36, pp. 82–93, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0736585318308876

[34] 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.

[35] 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," *Appl. Intell.*, vol. 48, no. 7, pp. 1739–1756, Jul. 2018.

[36] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.

[37] M. Liu, R. Lang, and Y. Cao, "Number of trees in random forest," *Comput. Eng. Appl.*, vol. 51, no. 5, pp. 126–131, 2015.

[38] U. Aprilliani and Z. Rustam, "Osteoarthritis disease prediction based on random forest," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Oct. 2018, pp. 237–240.

[39] A. S. Kurnar, "Diagnosis of heart disease using fuzzy resolution mechanism," *J. Artif. Intell.*, pp. 1–9, 2012.

[40] A. S. Kumar, "Adaptive neuro-fuzzy inference system for heart disease diagnosis," in *Proc. Int. Conf. Inf. Syst., Comput. Eng. Appl. (ICISCEA)*, Singapore, 2011, pp. 91–99.

[41] N. Jankowski and V. Kadirkamanathan, "Statistical control of RBF-like networks for classification," in *Proc. Int. Conf. Artif. Neural Netw.* Springer, 1997, pp. 385–390.

[42] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Syst. Appl.*, vol. 35, nos. 1–2, pp. 82–89, 2008.

[43] S. Ozsen, S. Gunes, S. Kara, and F. Latifoglu, "Use of kernel functions in artificial immune systems for the nonlinear classification problems," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 621–628, Jul. 2009.

[44] K. Polat, S. Sahan, H. Kodaz, and S. Güneş, "A new classification method to diagnosis heart disease: Supervised artificial immune system (AIRS)," in *Proc. Turkish Symp. Artif. Intell. Neural Netw. (TAINN)*, 2005.

[45] B. Šter and A. Dobnikar, "Neural networks in medical diagnosis: Comparison with other methods," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 1996, pp. 427–430.

[46] (2007). *Datasets Used for Classification: Comparison of Results*. [Online]. Available: http://www.is.umk.pl/projects/datasets.html

[47] N. Cheung, "Machine learning techniques for medical analysis. school of information technology and electrical engineering," Ph.D. dissertation, University of Queenland, Lucia QLD, Australia, 2001.