IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 17, 2019, accepted April 2, 2019, date of publication April 9, 2019, date of current version May 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909969

An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure

LIAQAT ALI[1,2], AWAIS NIAMAT[3], JAVED ALI KHAN[4,5], NOORBAKHSH AMIRI GOLILARZ[6], XIONG XINGZHONG[3], ADEEB NOOR[7], REDHWAN NOUR[8], AND SYED AHMAD CHAN BUKHARI[9]

[1]School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China
[2]Department of Electrical Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan
[3]School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin 643000, China
[4]School of Software Engineering, Tsinghua University, Beijing 100084, China
[5]Department of Software Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan
[6]School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China
[7]Faculty of Computing and Information Technology (FCIT), King Abdulaziz University, Jeddah 21441, Saudi Arabia
[8]Department of Computer Science, Taibah University, Medina 42353, Saudi Arabia
[9]Department of Pathology, Yale School of Medicine, Yale University, New Haven, CT 06511, USA

Corresponding authors: Liaqat Ali (engr_liaqat183@yahoo.com) and Syed Ahmad Chan Bukhari (ahmad.chan@yale.edu)

ABSTRACT About half of the people who develop heart failure (HF) die within five years of diagnosis. Over the years, researchers have developed several machine learning-based models for the early prediction of HF and to help cardiologists to improve the diagnosis process. In this paper, we introduce an expert system that stacks two support vector machine (SVM) models for the effective prediction of HF. The first SVM model is linear and L_1 regularized. It has the capability to eliminate irrelevant features by shrinking their coefficients to zero. The second SVM model is L_2 regularized. It is used as a predictive model. To optimize the two models, we propose a hybrid grid search algorithm (HGSA) that is capable of optimizing the two models simultaneously. The effectiveness of the proposed method is evaluated using six different evaluation metrics: accuracy, sensitivity, specificity, the Matthews correlation coefficient (MCC), ROC charts, and area under the curve (AUC). The experimental results confirm that the proposed method improves the performance of a conventional SVM model by 3.3%. Moreover, the proposed method shows better performance compared to the ten previously proposed methods that achieved accuracies in the range of 57.85%–91.83%. In addition, the proposed method also shows better performance than the other state-of-the-art machine learning ensemble models.

INDEX TERMS Clinical expert system, feature selection, heart failure prediction, hybrid grid search algorithm, support vector machine.

I. INTRODUCTION

Heart failure (HF) is the failure of heart to pump sufficient amount of blood to meet the needs of the body. Narrowing or blockage of the coronary arteries is considered to be the main cause of HF. Coronary arteries are those arteries which are responsible for carrying blood to the heart itself [1]. The common symptoms of HF include shortness of breath, swollen feet and weakness of the body.

In literature, different risk factors that lead to heart disease have been reported. These risk factors are divided into two groups. The first group includes patient's family history, sex and age. These risk factors cannot be changed. However, the second group includes risk factors that are related to life style of the patient. Hence, these factors can be changed e.g., high cholesterol level, smoking, physical inactivity and high blood pressure [2].

Due to lack of adequate diagnostic tools and medical experts, effective diagnosis of heart failure is a challenge [3], [4]. Furthermore, conventional methods for diagnosis of HF are based on various medical tests recommended by physicians, analysis of relevant symptoms and evaluating patients' medical history [5]. Among them,

The associate editor coordinating the review of this manuscript and approving it for publication was Navanietha Krishnaraj Krishnaraj Rathinam.

angiography is considered a key tool for diagnosis of HF. It is a type of diagnosis used to confirm heart disease and is regarded as a promising method for the diagnosis of HF. However, it has some limitations such as the high cost and side effects associated to it. Moreover, it also requires high level of technical expertise [6], [7]. A machine learning based expert system can reduce the associated health risk of the medical test. At the same time, it can help to improve the diagnosis process.

In the literature, researchers have developed different expert systems based on k-nearest neighbor (KNN), decision tree, support vector machine (SVM), fuzzy logic, artificial neural network (ANN) and ensembles of ANN for HF disease classification [1], [2], [8]–[15]. Robert Detrano collected the Cleveland dataset and used logistic regression for HF risk prediction. His model could achieve classification accuracy of 77%. Newton Cheung used different predictive models including C4.5, Naive Bayes, BNND and BNNF algorithm. These algorithms correctly classified patients and healthy subjects with accuracies of 81.11%, 81.48%, 81.11% and 80.95%, respectively. Polat *et al.* [16] developed an expert system based on artificial immune system (AIS) and obtained 84.5% accuracy. Özşen and Güneş [17] proposed another similar system and achieved accuracy of 87%. Das *et al.* [2] designed a neural network ensemble model with an aim to improve the classification accuracy. His ensemble model could achieve classification accuracy of 89.01%. Recently, Samuel *et al.* [5] proposed diagnostic system based on ANN and Fuzzy_AHP and achieved prediction accuracy of 91.10%.

Motivated by the development of different expert systems to lower down barriers in the diagnosis of heart disease and to improve the prediction accuracy, we attempt to develop an expert system based on stacked SVMs. Two SVM models have been used. The first model is linear and L_1 regularized while the second model is L_2 regularized and uses different kernels including linear and radial basis function i.e. RBF kernel. The first model has the capability to eliminate irrelevant features by shrinking their coefficients to zero. For different values of its hyperparameter C_1, different features are selected. Hence, we need to search the optimal value of C_1 which gives us optimal subset of features. The optimal subset of features are applied to the second SVM model which is used as a predictive model. The second model has its own hyperparameters i.e. kernel, C_2 and gamma denoted by G, which also need to be optimized. In this paper, we formulate the hyperparameters optimization problem of the two models as one hybrid optimization problem. To solve the hybrid optimization problem, we propose a hybrid grid search algorithm (HGSA).

The rest of the paper is organized as follows: In section 2, the dataset and the proposed methods are discussed. Section 3 deals with evaluation metrics and validation schemes. While section 4 is about experimental results and discussion. Finally, section 5 is about conclusion.

II. MATERIALS AND METHODS

A. DATASET DESCRIPTION

In this study, we collected a heart disease dataset known as Cleveland heart disease database from an online machine learning and data mining repository of the University of California, Irvine (UCI). The dataset was collected by Dr. Robert Detrano and was obtained from V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. The dataset consists of 303 subjects. However, the data of 6 subjects have missing values. Thus, the data of 297 subject is considered for experiments. Moreover, original dataset has 76 raw features per subject. But, most of the previous studies used only 13 of them. Hence, in this study the commonly used 13 HF features are considered. These 13 HF features are described and tabulated in Table 1. Moreover, two samples, one belonging to a patient and other belonging to a healthy subject, are depicted in Figure 1.

TABLE 1. Features Description of the HF Dataset.

Feature No	Features Code	Feature Description	Feature Abbreviations
1	F_1	Age (Years)	AGE
2	F_2	Sex	SEX
3	F_3	Chest Pain Type	CPT
4	F_4	Resting Blood Pressure	RBP
5	F_5	Serum Cholesterol	SCH
6	F_6	Fasting Blood Sugar	FBS
7	F_7	Resting Electrocardiographic Results	RES
8	F_8	Maximum Heart Rate achieved	MHR
9	F_9	Exercise Induced Angina	EIA
10	F_{10}	Old Peak	OPK
11	F_{11}	Peak Exercise Slope	PES
12	F_{12}	Number of Major Vessels Colored by Fluoroscopy	VCA
13	F_{13}	Thallium Scan	THA

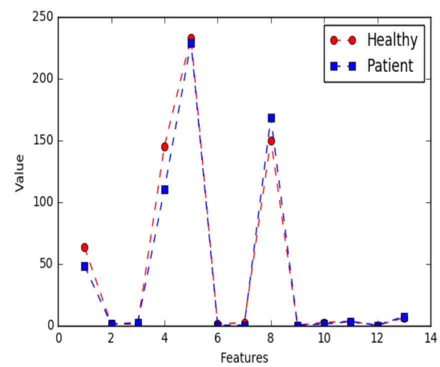


FIGURE 1. Samples of a patient and a healthy subject.

B. PROPOSED METHOD

The proposed diagnostic system has two sequential stages as shown in Figure 2. The first stage uses a linear and L_1

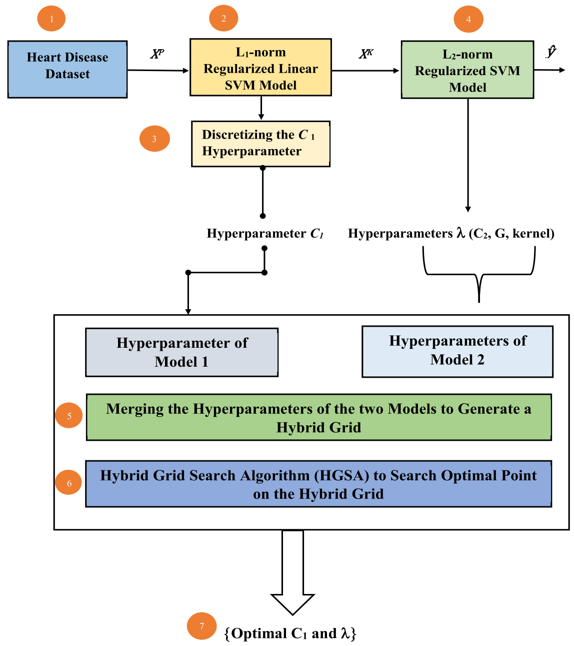


FIGURE 2. **Block Diagram of the newly proposed method. X^P: Original set of HF features, X^K: Optimal subset of features, $\hat{y}$: predicted label, C: Hyperparameter of the first linear SVM model which acts are feature selector, and λ: Contains the C, G, and kernel hyperparameters of the second SVM model which acts as predictive model.**

regularized SVM while the second stage uses L_2 regularized SVM with different kernels including linear and RBF. The first model has the capability to eliminate irrelevant features by shrinking their coefficients to zero. For different values of its hyperparameter i.e. C_1, different features are eliminated resulting in different subsets of features. To find the set of discrete values of C_1 that would yield different subsets of features, we manually tune C_1 with distinct values. After searching these discrete values of C_1, we declare the hyperparameter space for C_1 with these discrete values. Next, we need to search the optimal value of C_1 out of the declared finite discrete values of C_1 which would result in optimal subset of features. The optimal subset of features is applied to the second SVM model which is used as a predictive model. The second model has its own hyperparameters i.e. kernel, C_2 and gamma G which also need to be optimized. To understand the impact of L_1 regularization and L_2 regularization on the functionality of SVM, and how the two models perform the feature selection and classification tasks, it is important to discuss the formulation of the two models. The formulation of L_2 regularized SVM model is as follows:

1) L_2 SUPPORT VECTOR MACHINE

Support Vector Machines (SVMs) have been widely used as powerful machine learning method in different classification problems including bioinformatics. The model tries to search an optimal hyperplane which will maximize the distance from the nearest training data points of any class. SVM models are widely used in classification problems owing to their powerful capability of generalization to new unseen data objects, absence of local minima, flexible non-linear decision boundary, and their dependence on very few hyperparameters [18].

Considering a dataset S with k instances: $S = \{(x_i, y_i)|x_i \in R^P, y_i \in \{-1, 1\}\}_{i=1}^k$ where x_i denotes i^{th} instance and P denotes the dimension of each instance or feature vector. Moreover, the class label is denoted by y_i. The class label may be -1 or 1 for HF disease binary classification problem. The SVM model learns hyper-plane given by $f(x) = w^T * x + b$, where b is the bias and w is the weight vector. The hyperplane of the SVM model maximizes the margin while minimizes the classification error. The margin is computed as the sum of the distances to one of the closest positive and one of the closest negative instances. That is the hyperplane maximizes the margin distance $\frac{2}{\|w\|_2^2}$.

By introducing a set of slack variables $\xi_i, i = 1, \ldots, k$ and a penalty parameter i.e., C, the SVM model tries to balance the minimization of $\| w \|_2^2$ and the minimization of the misclassification errors. This is clear from the formulation given below:

$$\min_{w,b,\xi} \underbrace{\frac{1}{2} \| w \|_2^2}_{\text{Regularizer}} + C \underbrace{\sum_{i=1}^{k} \xi_i}_{\text{Error or Loss}}$$
$$s.t \begin{cases} y_i(wx_i + b) \geq 1 - \xi_i, \\ \xi_i \geq 0, i = 1, \ldots, k \end{cases} \tag{1}$$

where L_2-norm is the regularizer term and ξ is slack variable which measures the degree of misclassification.

2) L_1 SUPPORT VECTOR MACHINE

In 1998, Bradley and Mangasarian proposed a variation of SVM, reducing the model's complexity by using L_1-norm as the penalty function or regularizer instead of the Euclidean norm i.e., L_2 norm [19]. The L_1-norm SVM can be used for feature selection due to its capability of suppressing irrelevant or noisy features automatically. It shrinks components of the vector w that correspond to the features that would be eliminated. It can be formulated as follows:

$$\min_{w,b,\xi} \underbrace{\| w \|_1}_{\text{Regularizer}} + C \underbrace{\sum_{i=1}^{k} \xi_i}_{\text{Error or Loss}}$$
$$\text{s.t} \begin{cases} y_i(wx_i + b) \geq 1 - \xi_i, \\ \xi_i \geq 0, i = 1, \ldots, k \end{cases} \tag{2}$$

For sufficiently small C, some of the fitted coefficients i.e., components of w in 2 will be exactly zero i.e. sparse solutions. This property of L_1 regularized linear SVM model makes it capable of feature selection. Additionally, if we change value of the C hyperparameter, different fitted coefficients will be made zero. As a result, different subsets of features will be obtained [20]. Thus, we need to search the optimal value of the hyperparameter C that will yield optimal subset of features.

3) FORMULATING THE TWO OPTIMIZATION PROBLEMS AS ONE OPTIMIZATION PROBLEM BY MERGING THEM

From the above discussion, it is evident that we are dealing with two models stacked together. As discussed above, both the models have their hyperparameters. In this paper, to differentiate between the two models, the C hyperparameter of the L_1 regularized linear SVM model, which acts as a feature selection model, is denoted by C_1 and the hyperparameter of the second model i.e., the L_2 regularized SVM which acts as predictive model or classifier is denoted by C_2. The second model also has another hyperparameter i.e., type of kernel. If the type of kernel used is linear, then the second model will have only one hyperparameter i.e., C_2. On the other hand, if the type of kernel used is RBF kernel, then the second SVM model will have another hyperparameter i.e., G. In any case, the hyperparameters of both the models need to be optimized. Thus, we are dealing with two optimization problems i.e. optimization of hyperparameter of the first model and optimization of hyperparameters of the second model. The optimization of C_1 will generate optimal subset of features while the optimization of the second model will yield optimized predictive model.

In this paper, we merge the hyperparameters of the two models, as a result a hybrid grid is produced. That is the first coordinate of each point on the hybrid gird will be the hyperparameter of the first model i.e., C_1 while the second and third coordinates will be the hyperparameters of the second model i.e. C_2 and G. Hence, each point on the hybrid grid can be denoted by (C_1, C_2, G). The optimal point on the hybrid gird will yield the optimized versions of the two models simultaneously. In other words, the optimal point on the hybrid grid corresponds to the optimal subset of features and the optimized predictive model which will show good performance on the optimal subset of features. To search the optimal point on the hybrid grid, we propose to use a hybrid grid search algorithm (HGSA).

III. VALIDATION SCHEME AND EVALUATION METRICS

A. VALIDATION SCHEME

In previous studies, holdout validation schemes have been used for evaluating the performance of the developed diagnostic systems. Different studies have used different train-test split percentage for data portioning. Most of these studies like Das *et al.* in [2], Anooj P.K in [9] and Paul et al. in [21] have used holdout validation with 70-30 split. That is 70% of the dataset is used for training the proposed model while 30% is used for testing purpose. In this paper, we used the same approach with the same percentage for train-test split during data portioning.

B. EVALUATION METRICS

To evaluate the effectiveness of the newly proposed method, different evaluation metrics including accuracy, sensitivity, specificity and Matthews correlation coefficient (MCC) have been used. Accuracy is the percentage of correctly classified subjects. Sensitivity is the percentage of correctly classified patients while specificity is the correctly classified healthy subjects.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP denotes number of true positives, TN denotes number of true negatives, FP denotes number of false positives, and FN denotes number of false negatives.

$$Sensitivity = \frac{TP}{TP+FN} \tag{4}$$

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

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

In machine learning and statistics, the quality of binary classification is measured using MCC. Its value can be between -1 and 1. MCC value of -1 indicates total disagreement between prediction and observation, 1 indicates a prefect prediction and 0 means the classification is no better than a random prediction.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, three types of experiments are performed to evaluate the effectiveness of the proposed model. In the first experiment, L_1 regularized linear SVM model is stacked with L_2 regularized linear SVM model. In the second experiment, L_1 regularized linear SVM model is stacked with L_2 regularized SVM model with RBF kernel. Finally, to compare the performance of the proposed model with other machine learning models, third experiment is performed. All computations are performed on Intel (R) Core (TM) i3-2330M CPU @2.20GHz with 64bit windows 7 as operating system. Moreover, Python programming software package is used to simulate the experiments.

A. EXPERIMENT NO 1: L_1 REGULARIZED LINEAR SVM STACKED WITH L_2 REGULARIZED LINEAR SVM

In this experiment, at first stage L_1 regularized linear SVM is implied while at second stage L_2 regularized linear SVM is used. The first model eliminates noisy and irrelevant features while the second model is used as a predictive model. The best accuracy of 91.11% is obtained using only 11 and 12 features i.e., for $K = 11$ and $K = 12$, respectively. The optimal subset of features for $K = 11$ includes $F_2, F_3, F_4, F_6, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}$ and F_{13} while the optimal subset obtained for $K = 12$ includes $F_1, F_2, F_3, F_4, F_6, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}$ and F_{13}. The last row shows a case where all features are used i.e. only conventional linear SVM is used for classification and no feature selection is carried out. Hence, the proposed method is evidently effective as it gives better performance with less number of features. The performance at 11 features is the best as it has highest training accuracy as well. The results

TABLE 2. Simulation results of L_1-linear SVM model stacked with L_2 linear SVM model. C_1: Hyperparameter of the L_1-linear SVM model, C_2: Hyperparameter of the L_2-linear SVM model, K: Size of selected subset of features, $\mathrm{Acc_{test}}$: Accuracy of testing dataset, and $\mathrm{Acc_{train}}(\%)$: Accuracy of training dataset, Sens[itivity], and Spec[ificity].

C_1	K	C_2	$\mathrm{Acc_{test}}$	$\mathrm{Acc_{train}}(\%)$	Spec(%)	Sens(%)	MCC
0.010	1	0.500	78.88	75.36	81.63	75.60	0.573
0.011	2	90.00	82.22	76.81	85.71	78.04	0.640
0.015	3	0.0005	83.33	75.84	81.63	85.36	0.667
0.018	4	100.0	88.88	81.64	93.87	82.92	0.777
0.045	5	80.00	88.88	82.12	93.87	82.92	0.777
0.050	6	0.095	88.88	83.09	93.87	82.92	0.777
0.055	7	0.055	90.00	82.60	95.91	82.92	0.801
0.060	8	90.00	90.00	83.09	95.91	82.92	0.801
0.080	9	0.100	88.88	83.57	91.83	85.36	0.775
0.090	10	0.500	90.00	83.09	93.87	85.36	0.799
0.200	**11**	**0.500**	**91.11**	**84.05**	**93.87**	**87.80**	**0.820**
0.450	12	0.500	91.11	83.09	93.87	87.80	0.820
—	13	0.055	90.00	84.05	93.87	85.36	0.799

at different subsets of features and different hyperparameters are tabulated in Table 2.

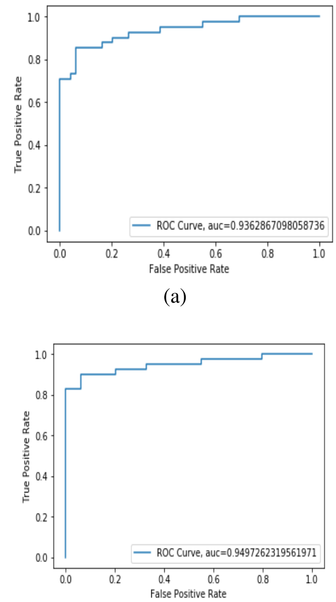
B. EXPERIMENT NO 2: L_1 REGULARIZED LINEAR SVM STACKED WITH L_2 REGULARIZED SVM USING RBF KERNEL

In this experiment, at first stage L_1 regularized linear SVM is used for selecting the most relevant features while at second stage L_2 regularized SVM with RBF kernel is used. The second model with RBF kernel acts as a predictive model. The best accuracy of 92.22% is obtained using only 8 features i.e., for $K = 8$. The optimal subset of features includes $F_2, F_3, F_7, F_8, F_9, F_{11}, F_{12}$ and F_{13}. The optimal subset of features not only improves the potential of the predictive model but also reduces the time complexity of the predictive model i.e., the training time of the predictive model is also reduced. The results at different subsets of features at different hyperparameters are tabulated in Table 3. In the table, the last row represents a case when only the second SVM model i.e., the L_2 regularized SVM model is used. This case represents the conventional SVM model. Thus, from the experimental results it clear that the proposed method improves the performance of a conventional SVM model by 3.3%.

To further investigate the performance of the proposed method, another evaluation metric i.e., ROC chart is used. The ROC chart is the plot of true positive rate (TPR) versus the false positive rate (FPR) for various thresholds. An ROC chart with more area under the curve (AUC) is considered best. An ideal ROC chart has AUC=100, such chart means that the model is capable of performing with 100% sensitivity and 100% specificity. Figure 3 and Figure 4 show ROC chart for the HF binary prediction problem using SVM linear model and SVM RBF model as predictive models, respectively. From the Figure 3, it can be seen that AUC=0.936 for the optimized conventional SVM RBF model while AUC=0.949 for the proposed optimized stacked SVM models in which the first model is SVM linear while the second model is SVM RBF. Hence, it is clear that the proposed method improves the strength of SVM RBF predictive model. Similarly, from Figure 4, it can be observed that AUC=0.949 for optimized conventional SVM linear model which is used as predictive model while AUC=0.952 for the proposed stacked SVM models where both the stacked models are SVM linear models. Thus, it is evident that the proposed method also improves the strength of SVM linear predictive model.

TABLE 3. Simulation results of L_1-linear SVM model cascaded with L_2 SVM model with RBF kernel. C_1: Hyperparameter of the L_1-linear SVM model, C_2: Hyperparameter of the L_2-SVM model with RBF kernel, G: Hyperparameter of the L_2-SVM model with RBF kernel, K: Size of selected subset of features, $\mathrm{Acc_{test}}$: Accuracy of testing dataset, and $\mathrm{Acc_{train}}(\%)$: Accuracy of training dataset, Sens[itivity], and Spec[ificity].

C_1	K	C_2	G	$\mathrm{Acc_{test}}$	$\mathrm{Acc_{train}}(\%)$	Spec(%)	Sens(%)	MCC
0.010	1	0.400	0.600	78.88	75.36	81.63	75.60	0.573
0.011	2	300.0	10.00	81.11	77.29	83.67	78.04	0.618
0.015	3	0.095	200.0	82.22	79.22	81.63	82.92	0.643
0.018	4	70.00	0.015	88.88	82.60	95.91	80.48	0.780
0.045	5	0.400	5.000	88.88	83.09	95.51	80.48	0.780
0.050	6	0.500	5.000	88.88	86.47	89.79	87.80	0.776
0.055	7	30.00	0.055	90.00	85.02	97.95	80.48	0.805
0.060	**8**	**400.0**	**0.015**	**92.22**	**85.02**	**100.0**	**82.92**	**0.851**
0.080	9	100.0	0.003	90.00	83.09	95.91	82.92	0.801
0.090	10	5.000	0.005	90.00	83.09	95.91	82.92	0.801
0.200	11	10.00	0.010	90.00	84.05	93.87	85.36	0.799
0.450	12	20.00	0.005	90.00	84.05	93.87	85.36	0.799
—	13	5.000	0.200	90.00	87.43	93.87	85.36	0.799

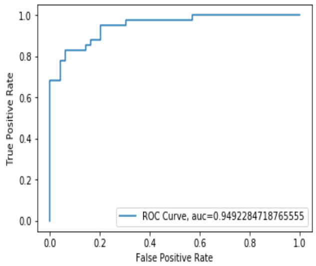


(a)
(b)

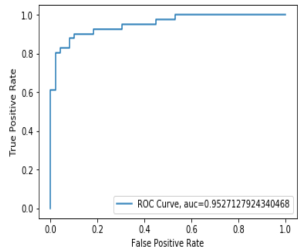
FIGURE 3. ROC charts using SVM RBF model as predictive model. (a) ROC chart of the optimized conventional SVM RBF Model. (b) ROC chart of the proposed optimized stacked models. The first model is L_1 linear SVM and the second model is L_2 SVM model with RBF kernel.

C. COMPARATIVE STUDY

In this subsection, experimental results of the proposed method are compared with other machine learning models and previously proposed methods. First, the performance



(a)



(b)

FIGURE 4. ROC charts using SVM linear model as predictive model. (a) ROC chart of the optimized conventional SVM Linear Model. (b) ROC chart of the proposed optimized stacked models. The first model is L_1 linear SVM and the second model is L_2 SVM model with linear kernel.

of the proposed stacked SVM models is compared with other state of the art machine learning models. In this case, the performance comparison is done using accuracy, ROC chart and AUC evaluation metrics. Second, the performance of the proposed method or model is compared with other methods presented in literature. In this case, the performance comparison is done using classification accuracy.

TABLE 4. Performance of other well known machine learning models after optimization.

Model	Hyperparameters	$\mathrm{Acc_{test}}$	Spec.	Sens.	MCC
Adaboost	$N_e = 4$	88.00	89.79	87.80	0.776
Random Forest	$N_e = 50$	88.00	93.87	82.92	0.777
Extra Tree	$N_e = 11$	88.00	89.79	87.80	0.776
Proposed	$\mathbf{n = 8}$	**92.22**	**100.0**	**82.92**	**0.851**

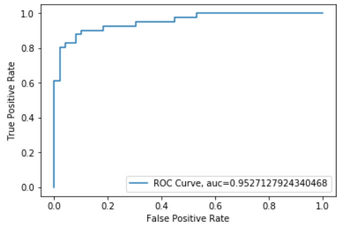
1) COMPARATIVE STUDY WITH OTHER STATE OF THE ART MACHINE LEARNING MODELS

To validate the effectiveness of the proposed model, a comparative study is performed with other state of the art machine learning models. These models include random forest (RF), Adaboost and extra tree also known as randomized decision tree. The hyperparameters of all these models are optimized using exhaustive search strategy. The performance of these models is reported in Table 4. In the table, for Adaboost model, the hyperparameter N_e denotes the maximum number of estimators at which boosting is terminated. For random forest model, the hyperparameter N_e denotes the number of trees in the forest and for extra tree ensemble model the hyperparameter N_e denotes the number of trees used by the ensemble model. From the table, it is evidently clear that the proposed model show better performance than the ensemble machine learning models.

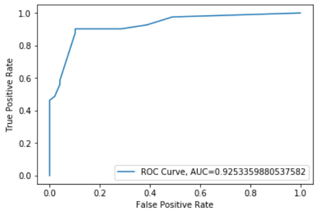
To further validate the effectiveness of the proposed optimized stacked model, we also compare it's performance with other models based on ROC charts and AUC evaluation metrics. The ROC charts for the proposed model, Adaboost ensemble model, random forest ensemble model and extra

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

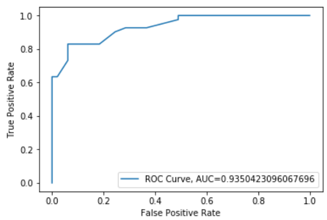
Study (Year)	Method	Accuracy(%)
ToolDiag, RA [22]	IB1-4	50.00
WEKA, RA [22]	InductH	58.50
ToolDiag, RA [22]	RBF	60.00
WEKA, RA [22]	FOIL	64.00
ToolDiag, RA [22]	MLP + BP	65.60
WEKA, RA [22]	T2	68.10
WEKA, RA [22]	1R	71.40
WEKA, RA [22]	IB1c	74.00
WEKA, RA [22]	$K*$	76.70
Robert Detrano [22]	Logistic regression	77.00
Newton Cheung (2001) [23]	C4.5	81.11
Newton Cheung (2001) [23]	Naive Bayes	81.48
Newton Cheung (2001) [23]	BNND	81.11
Newton Cheung (2001) [23]	BNNF	80.96
WEKA, RA [22]	Naive Bayes	83.60
Ster and Dobnikar [24]	Fisher discriminant Analysis	84.2
Ster and Dobnikar [24]	Linear discriminant Analysis	84.5
Ster and Dobnikar [24]	Naive Bayes	82.5-83.4
K. Polat et al. (2005) [16]	AIRS	84.50
Resul et al. (2009) [2]	Neural network ensembles	89.01
Jankowski and Kadirkamanathan [25]	IncNet	90.00
Senthil Kumar (2011) [26]	ANFIS	91.18
Anooj P.K (2012) [9]	Weighted fuzzy rules	62.35
Samuel et al. (2017) [5]	ANN-Fuzzy-AHP	91.10
Senthil Kumar (2012) [27]	Fuzzy resolution mechanism	91.83
Proposed Method (2019)	$\mathbf{L_1}$ **Linear SVM +** $\mathbf{L_2}$ **Linear & RBF SVM**	**92.22**



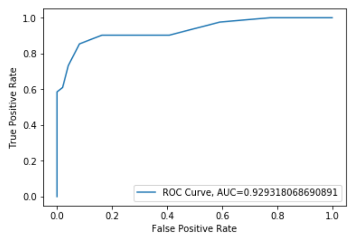
(a)



(b)



(c)



(d)

FIGURE 5. ROC charts of the proposed model and other machine learning models. (a) ROC chart of the proposed optimized stacked SVM models. (b) ROC chart of Adaboost ensemble model. (c) ROC chart of random forest ensemble model. (d) ROC chart of extra tree ensemble model.

tree ensemble model are shown in Figure 5 (a), Figure 5 (b), Figure 5 (c) and Figure 5 (d), respectively. From the figure, it is evidently clear that the AUC of the ROC chart for the proposed model is 0.952 and the AUC of the ROC chart for the Adaboost ensemble model is 0.925. Similarly, the AUC of the ROC chart for the random forest ensemble model is 0.935 and the AUC of the ROC chart for the extra tree ensemble model is 0.929. Thus, it is evidently clear that the proposed model shows better performance than the ensemble machine learning models from all the three evaluation aspects i.e., accuracy, ROC chart and AUC. Hence, the effectiveness of the proposed method is validated.

2) COMPARATIVE STUDY WITH PREVIOUS METHODS

To further validate the improved performance of the proposed model, comparative study is conducted with previously proposed methods applied to the Cleveland heart disease dataset. The comparative study is conducted in terms of classification accuracy. The previously proposed methods and their accuracies achieved are tabulated in Table 5.

V. CONCLUSION

In this paper, an expert system based on stacked SVMs was proposed to facilitate the diagnosis of heart failure. The first SVM model was used to eliminate irrelevant features while the second model was used as predictive model. Both the models were optimized using a hybrid grid search algorithm. It was shown that the proposed method outperformed ten renowned existing methods in literature and other state of the art machine learning models. It was also observed that the proposed model improves the strength of conventional SVM model by 3.3%. Moreover, the proposed method is capable of showing better results with a few features. Thus, the proposed method is efficient in terms of time complexity as well. Because it reduces the training time of the predictive model. Hence, from the experimental results achieved on the heart failure dataset, it is concluded that the proposed expert system can improve the decision making process of the physicians during diagnosis of heart failure.

CONFLICT OF INTEREST

The authors declare no competing interests.

REFERENCES

[1] 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, 2007.

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

[4] L. A. Allen *et al.*, "Decision making in advanced heart failure: A scientific statement from the american heart association," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.

[5] 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.

[6] R. Alizadehsani, M. J. Hosseini, Z. A. Sani, A. Ghandeharioun, and R. Boghrati, "Diagnosis of coronary artery disease using cost-sensitive algorithms," in *Proc. IEEE 12th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2012, pp. 9–16.

[7] 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.

[8] 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.

[9] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1319157811000346

[10] 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, 2010.

[11] 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.
[12] 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.
[13] I. Paschalidis, "How machine learning is helping us predict heart disease and diabetes," *Harvard Bus. Rev.*, 2017.
[14] 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.
[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, 2018.
[16] K. Polat, S. Sahan, H. Kodaz, and S. Günes, "A new classification method to diagnosis heart disease: Supervised artificial immune system (AIRS)," in *Proc. Turkish Symp. Artif. Intell. Neural Netw. (TAINN)*, 2005.
[17] S. Özşen and S. Güneş, "Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 386–392, 2009.
[18] 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, 2014.
[19] P. S. Bradley and O. L. Mangasarian, "Feature selection via concave minimization and support vector machines," in *Proc. ICML*, Feb. 1999, pp. 82–90.
[20] J. Zhu and H. Zou, "Variable selection for the linear support vector machine," in *Trends in Neural Computation*. Berlin, Germany: Springer, 2007, pp. 35–59.
[21] 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, 2017.
[22] (2007). *Datasets Used for Classification: Comparison of Results*. [Online]. Available: http://www.is.umk.pl/projects/datasets.html
[23] N. Cheung, "Machine learning techniques for medical analysis," Ph.D. dissertation, School Inf. Technol. Elect. Eng., Univ. Queenland, Brisbane, QLD, Australia, 2001.
[24] 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.
[25] N. Jankowski and V. Kadirkamanathan, "Statistical control of RBF-like networks for classification," in *Proc. Int. Conf. Artif. Neural Netw.* Berlin, Germany: Springer, 1997, pp. 385–390.
[26] A. V. 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.
[27] A. V. S. Kumar, "Diagnosis of heart disease using fuzzy resolution mechanism," *J. Artif. Intell.*, vol. 5, no. 1, pp. 47–55, 2012.

Authors' photographs and biographies not available at the time of publication.