

Received January 30, 2020, accepted February 16, 2020, date of publication February 20, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975377

A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease

JIKUO WANG^{ID1}, CHANGCHUN LIU¹, LIPING LI², WANG LI³, LIanke YAO¹, HAN LI¹, AND HUAN ZHANG¹

¹School of Control Science and Engineering, Shandong University, Jinan 250061, China

²School of Science and Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China

³School of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing 400054, China

Corresponding author: Changchun Liu (changchunliu@sdu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61471223.

ABSTRACT Coronary arteriography (CAG) is an accurate invasive technique for the diagnosis of coronary heart disease (CHD). However, its invasive procedure is not appropriate for the detection of CHD in the annual physical examination. With the successful application of machine learning (ML) in various fields, our goal is to perform selective integration of multiple ML algorithms and verify the validity of feature selection methods with personal clinical information commonly seen in the annual physical examination. In this study, a two level stacking based model is designed in which level 1 is base-level and level 2 is meta-level. The predictions of base-level classifiers is selected as the input of meta-level. The pearson correlation coefficient and maximum information coefficient are first calculated to find the classifier with the lowest correlation. Then enumeration algorithm is used to find the best combining classifiers which acquire the best result in the end. The Z-Alizadeh Sani CHD dataset which we use consists of 303 cases verified by CAG. Experimental results demonstrate that the proposed model obtains an accuracy, sensitivity and specificity of 95.43%, 95.84%, 94.44%, respectively for the detection of CHD. The proposed method can effectively aid clinicians to detect those with normal coronary arteries from those with CHD.

INDEX TERMS Coronary heart disease, machine learning, feature selection, stacking.

I. INTRODUCTION

Coronary heart disease (CHD) remains one of the leading causes of cardiovascular death globally. At present, the diagnostic methods of CHD can be divided into invasive and non-invasive ways. Coronary angiography (CAG) is a relatively safe and reliable invasive diagnostic technique, which has been widely used in clinical practice as the gold standard for the CHD diagnosis [1]. However, its invasive nature and relatively expensive operation cost makes it difficult to apply in the annual physical examination. Electrocardiogram (ECG) and echocardiography are non-invasive methods, but neither with reliable accuracy [2]. Therefore, it is necessary to find new non-invasive methods to detect CHD.

In clinical cardiology, machine learning (ML) has been proved an effective method for prediction of all-cause mortality in patients with suspected CHD [3]. In subclinical

The associate editor coordinating the review of this manuscript and approving it for publication was Chua Chin Heng Matthew^{ID}.

cardiovascular epidemiology, ML can provide better prediction than standard cardiovascular risk scores in conjunction with phenotypic data points [4]. ML methods are widely used in dealing with existing data in medicine. In recent years, a quantity of ML algorithms for diagnosing CHD has been developed. Feshki and Shijani improved CHD diagnosis by an evolutionary algorithm and a feedforward neural network [5]. Davari *et al.* extracted features from ECGs by frequency and nonlinear domain methods to identify CHD symptoms with support vector classifier (SVC) classifier [6]. Vernekar *et al.* extracted Markov features along with other statistical and frequency domain features from phonocardiogram (PCG) and used the set of artificial neural network and gradient enhancement tree for model training [7]. Kumar *et al.* also used ECG signals but with flexible analytic wavelet transform to characterize the CHD [8]. Verma *et al.* proposed a hybrid method which included risk factor identification using correlation-based feature subset selection with particle swarm optimization search method and K-means clustering algorithms [9]. Alizadehsani *et al.* used three

classifiers for detection of the stenosis of three coronary arteries, i.e., left anterior descending, left circumflex and right coronary artery to get higher accuracy for CHD diagnosis [2]. Davari *et al.* achieved 99.2% detection accuracy with the Long Term ST database, but the database they used for CHD patients is accompanied by various ST segment changes [6]. And in clinical practice, many CHD patients have normal ST segment. Therefore, using the databases of patients with coronary artery disease but normal ST segments maybe more helpful to the application of artificial intelligence-based CHD diagnosis model in clinical complex situations. Besides, previous research usually employed only one kind of ML classifier to automatically diagnose CHD. However, many ML researchers especially those participating in ML competitions have successfully used classifier combinations techniques to improve the accuracy of the classifiers [10], [11].

Techniques for combining predictions obtained from multiple base-level classifiers can be summarized into three combinatorial frameworks: voting (used in bagging and boosting), stacking and cascading [12]. For more complex data sets, the traditional classifier can be improved by various types of combination rules [13]–[16]. In stacking, the predictions of a collection of classifiers are given as inputs to the next-level learning algorithm [17]. The next-level of algorithm is trained to associate the model predictions optimally and to form the next-level of the final set of predictions. Coupling relationships always exist between the different levels before the final prediction. We analyze the relationships between models in the base-level and find the optimal combination of the model by an enumeration algorithm.

In summary, the main contribution of this work are summarized as follow:

- Eight feature selection methods are investigated to evaluate their performances for automated CHD diagnosis. We find that the RFECV machine-learning strategy achieved the highest predictive performance in repeated ten-fold cross-validation. Those features selected by the RFECV method are of high reference value to cardiologists in their clinical CHD diagnosis.

- A total of 10 classification methods are utilized. By analyzing the results, it is found that the model combination exhibiting the best performance cannot be determined by directly calculating pearson correlation coefficient (PCC) and maximum information coefficient (MIC). Therefore, a novel strategy of seeking the optimal combination is proposed, in which a model having the minimum correlation with other models is first selected and then the optimal combination is determined by enumerating any possible combination of the selected model with others. Our results show that the proposed strategy yield satisfactory performances.

- The optimal model combination for automated CHD diagnosis is determined. The application of the proposed model combination strategy on the other 3 data sets also shows satisfactory results, which demonstrates the generalization ability of our proposed model combination strategy.

The remainder of this paper is organized as follows. In section II, the data source and the preprocessing methods of the data are introduced. In section III, the technical details of our proposed two-level stacking based model are described. Experimental results are presented in Section IV followed by discussions in Section V.

II. MATERIAL

The Z-Alizadeh Sani dataset [<http://archive.ics.uci.edu/ml/datasets/extention+of+Z-Alizadeh+sani+dataset>] consists of 216 CHD patients and 87 healthy subjects represented by 54 different kinds of clinical and demographic features as shown in Table 1 [18]. The dataset exhibits a huge imbalance in the distribution of the target classes, for there are approximately 3 times more CHD patients than healthy subjects. In such case, the synthetic minority oversampling technique (SMOTE) is employed to solve the imbalance problem. The basic idea of the SMOTE method is to analyze minority classes and synthesize new minority classes by oversampling. The data of normal individuals are oversampled by SMOTE during cross validation and not prior to the cross validation process. Synthetic data are created only for the training set without affecting the test set. If a feature has a variance that is orders of magnitude larger than others, it might affect the objective function and makes the estimator fail to learn from other features correctly as expected [19]. Since the 54 features of the dataset include 23 numeric and 31 categorical data, the technique of maximum and minimum normalization is applied to standardize these features. Maximum and minimum normalization is a common method of data processing, which can be defined as (1). x is the input feature, max represents the maximum value, min represents the minimum value, and x^* represents the output value after normalization. In this study, we use this approach to scale the 23 features (“Weight”, “Length”, “BMI”, “BP”, “PR”, “Age”, “FBS”, “CR”, “TG”, “LDL”, “HDL”, “BUN”, “ESR”, “HB”, “K”, “Na”, “WBC”, “Lymph”, “Neut”, “PLT”, “EF-TTE”, “Region RWMA”, “VHD”). It’s helpful to find the potential importance relationships among the features.

$$x^* = \frac{x - \min}{\max - \min} \quad (1)$$

III. METHODS

A. FEATURE SELECTION

Feature selections are of great importance in dealing with the redundant features [20], [21]. Three common feature selection criteria consist of filter, wrapper and embedded. The filter methods calculate the relationship between the features and the label using the statistical tools including variance, mutual information and chi-square test (CHI2) [22], [23]. The wrapper methods are closely related to the classifier. The principle of the wrapper method is to select the best subset according to the classifier performances [24], [25]. What’s more, the recursive feature elimination with cross-validation (RFECV) can eliminate the influence of artificially setting

TABLE 1. Features of Z-Alizadeh Sani dataset.

Feature type	Feature name	Range
Demographic	Age	30-86
	Weight	48-120
	Length	140-188
	Sex	Male,Female
	BMI(Body Mass Index)	18.1-40.9
	DM (Diabetes Mellitus)	Yes,No
	HTN (Hyper Tension)	Yes,No
	Current Smoker	Yes,No
	Ex-Smoker	Yes,No
	FH (Family History)	Yes,No
	Obesity	Yes,No
	CRF (Chronic Renal Failure)	Yes,No
	CVA (Cerebrovascular Accident)	Yes,No
	Airway Disease	Yes,No
	Thyroid Disease	Yes,No
	CHF (Congestive Heart Failure)	Yes,No
	DLP (Dyslipidemia)	Yes,No
	BP (Blood Pressure: mmHg)	90-190
	PR (Pulse Rate) (ppm)	50-110
Clinical	Edema	Yes,No
	Weak peripheral pulse	Yes,No
	Lung Rales	Yes,No
	Systolic murmur	Yes,No
	Diastolic murmur	Yes,No
	Typical Chest Pain	Yes,No
	Dyspnea	Yes,No
	Function Class	1,2,3,4
	Atypical	Yes,No
	Nonanginal	Yes,No
	Exertional CP (Exertional Chest Pain)	Yes,No
	LowTH Ang (low Threshold angina)	Yes,No
	Rhythm	Yes,No
	Q Wave	0,1
	ST Elevation	0,1
	ST Depression	0,1
	T inversion	0,1
	LVH (Left Ventricular Hypertrophy)	Yes,No
	Poor R Progression (Poor R Wave Progression)	Yes,No
	FBS (Fasting Blood Sugar) (mg/dl)	62-400
	Cr (creatinine) (mg/dl)	0.5-2.2
	TG (Triglyceride) (mg/dl)	37-1050
	LDL (Low density lipoprotein) (mg/dl)	18-232
	HDL (High density lipoprotein) (mg/dl)	15.9-111
	BUN (Blood Urea Nitrogen) (mg/dl)	6-52
	ESR (Erythrocyte Sedimentation rate) (mm/h)	1-90
	HB (Hemoglobin) (g/dl)	8.9-17.6
	K (Potassium) (mEq/lit)	3.0-6.6
	Na (Sodium) (mEq/lit)	128-156
	WBC (White Blood Cell) (cells/ml)	3700-18000
	Lymph (Lymphocyte) (%)	7-60
	Neut (Neutrophil) (%)	32-89
	PLT (Platelet) (1000/ml)	25-742
	EF (Ejection Fraction) (%)	15-60
	Region with RWMA a (Regional Wall Motion Abnormality)	0,1,2,3,4
	VHD (Valvular Heart Disease)	Normal, Mild, Moderate, Severe

of the features number remaining in the feature set. The embedded methods are integrated with the process of model training to select features automatically. Extreme gradient enhancement (XGB) has been widely used as an embedded feature selection method due to its high efficiency [26].

B. MODEL BUILDING

The proposed model mainly consists of two levels, in which the level 1 is the base-level and the level 2 is meta-level.

The predictions of base-level classifiers are selected as the input of meta-level. The base-level contains 10 models from scikit-learn, including random forest (RF), extra trees (ET), adaBoost (ADB), SVC, multi-layer perceptron (MLP), XGB, gaussian process classification (GPC), gaussian naive bayes (GNB), logistic regression (LR), gradient boosting (GB) [27]–[36]. The performance of the stacking schemes is affected by the number of base-level classifiers [37]. Generally, the base-level classifiers with weakly correlated

predictions yield good performance [37]. The PCC and MIC can be used as a measure of quantifying the relevance and redundancy among features [38]–[40], with being closer to 0 indicating weaker correlation. Then, we use the enumeration algorithm to search for the best combining classifiers.

We summarize two algorithms that can illustrate the process of the stacking and enumeration. The dataset is first shuffled randomly and split into 10 folds. For each fold, one fold is treated as a test data (S) and the remaining folds are taken as R . The whole process is repeated 10 times. R and S are the input to Alg.1. The Alg.1 mainly contains two loops. The first loop is the process of building the ten base-level models, and the second loop is the process of 10-fold cross-validation to produce training and test data. R are also split into 10 folds. One fold is taken as the validation set (R_{kv}) and the remaining folds are treated as training data set (R_{kt}). R_{kt} is entered into the base-level model used to train the base-level model (ξ_l). R_{kv} is used to produce the $train_l$. Later, S is entered into the base-level model ξ_l to generate $test_l$. Since the loop repeats 10 times, the $train_l$ is exactly equal to the sum of the ten folds, and the test data set needs to be averaged. Finally, the union set of training and test generated by 10 different basic models is taken as output.

Algorithm 1 The Process of Building Base-Level Model

```

Input:  $R$ , nine folds
Input:  $S$ , one fold(test data set)
 $R_{kt}$ , training data set
 $R_{kv}$ , validation data set
the model of base-level  $\xi_l$ , where
 $l = \{RF, ET, ADB, SVC, MLP, XGB, GPC, GNB, LR, GB\}$ 
//the model which can chose
forall the
 $l = \{RF, ET, ADB, SVC, MLP, XGB, GPC, GNB, LR, GB\}$  do
  forall the  $k = 1, 2, \dots, 10$  do
     $\xi_l \leftarrow R_{kt}$ 
    //use  $R_{kt}$  to train  $\xi_l$ 
     $train_l \leftarrow \xi_l \leftarrow R_{kv}$ 
    //use  $R_{kv}$  to get  $train_l$ 
     $test_l \leftarrow \xi_l \leftarrow S$ 
    //input  $S$  to  $\xi_l$  to predict  $test_l$ 
  end
   $train_l = (train_1 + train_2 + \dots + train_k)$ 
   $test_l = (test_1 + test_2 + \dots + test_k)/10$ 
  //calculate the mean of test data set
end
 $train = [train_{RF}, train_{ET}, \dots, train_{GB}]$ 
 $test = [test_{RF}, test_{ET}, \dots, test_{GB}]$ 
Output:  $train, test$ 
```

The output of Alg. 1 is considered as the new features of the meta-level. Since it is unwise to directly use all the new features without filtering, the Alg. 2 is employed to search for the optimal combination. The Alg. 2 mainly contains two loops. In the first loop, there are 10 kinds of possible combinations, including $C_{10}^1, C_{10}^2, C_{10}^3, C_{10}^4, C_{10}^5, C_{10}^6, C_{10}^7, C_{10}^8, C_{10}^9, C_{10}^{10}$ as the input of the second loop. All possible combinations (without repeating them) are enumerated rather than putting them all into the next loop. In the second loop,

the input of the *train* is used to train the model H_m . The LR model is applied to reduce the complexity of the model [37]. The *test* is then imported into the trained model (H_m) to evaluate the performance of the model on the test data set. Finally, the model combination with the highest accuracy is determined.

Algorithm 2 The Process of Searching the Best Combination

```

Input:  $train$ , training data set
Input:  $test$ , test data set
// $H_m$ , the second model
forall the
 $l = \{RF, ET, ADB, SVC, MLP, XGB, GPC, GNB, LR, GB\}$  do
  forall the  $m = \{LR\}$  do
     $train \leftarrow C_{10}^l$ 
     $test \leftarrow C_{10}^l$ 
     $H_m \leftarrow train$ 
     $result_m \leftarrow H_m \leftarrow test$ 
    //evaluate the performance of the model on the test
    //data set
  end
end
Output:  $result_m$ 
```

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The programming language is python and the version is 3.7.2. Jupyter Notebook is a friendly environment IDE for programmers, which provides smart code completion, code inspections and especially services for interactive computing across lots of programming languages. In this study, our project is implemented in this open-source software. According to the knowledge of medical diagnosis, the accuracy(A_{cc}), sensitivity(S_e), specificity(S_p), F1, and AUC area are used to evaluate our method. Since 10-fold cross-validation is used, the results are expressed in the form of mean \pm standard deviation.

A. RESULTS OF DATA PRE-PROCESSING

After the preprocessing depicted in MATERIAL section, the raw data are standardized and the ‘Processed data’ have a range of 0 to 1. As shown in Table 2, the results of ‘Processed data’ show better performances than that of ‘Raw data’ through different classifiers. For ‘Processed data’, the LR and ADB models have higher accuracy than others. But the XGB and GNB acquire better scores than other models in sensitivity and specificity respectively. Therefore, these differences are produced by the heterogeneity of the model, which reflects the foundation of stacking.

B. RESULTS FOR FEATURE SELECTION

Table 3 shows the features selected by three different typical feature selection methods. The selected features will help doctors improve their understanding of the different importance of selected features. Furthermore, various therapeutic interventions can be specifically conducted to reduce or even eliminate the harmful influence of some selected features.

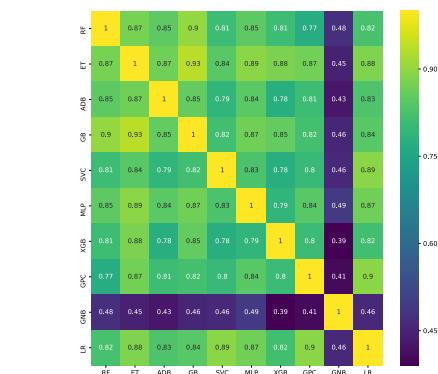
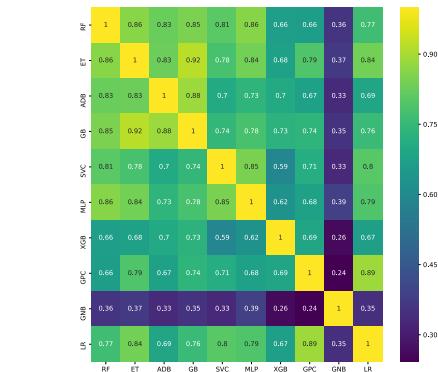
TABLE 2. Performance comparison of data precessing.

Method	Raw data			Processed data		
	A_c	S_e	S_p	A_c	S_e	S_p
LR	0.854(± 0.04)	0.893(± 0.04)	0.760(± 0.15)	0.881(± 0.06)	0.917(± 0.07)	0.792(± 0.11)
RF	0.842(± 0.03)	0.917(± 0.04)	0.657(± 0.11)	0.839(± 0.07)	0.894(± 0.07)	0.706(± 0.20)
GNB	0.734(± 0.09)	0.679(± 0.12)	0.872(± 0.11)	0.428(± 0.06)	0.217(± 0.09)	0.953(± 0.08)
SVC	0.814(± 0.06)	0.888(± 0.06)	0.633(± 0.11)	0.845(± 0.04)	0.884(± 0.06)	0.749(± 0.13)
DT	0.813(± 0.05)	0.870(± 0.09)	0.669(± 0.16)	0.786(± 0.06)	0.851(± 0.07)	0.622(± 0.18)
KNN	0.681(± 0.05)	0.875(± 0.05)	0.197(± 0.12)	0.805(± 0.06)	0.879(± 0.06)	0.622(± 0.16)
ADB	0.874(± 0.07)	0.930(± 0.06)	0.733(± 0.18)	0.882(± 0.06)	0.927(± 0.06)	0.769(± 0.11)
GB	0.859(± 0.05)	0.921(± 0.05)	0.704(± 0.16)	0.862(± 0.05)	0.907(± 0.04)	0.749(± 0.15)
ET	0.825(± 0.05)	0.875(± 0.06)	0.704(± 0.16)	0.852(± 0.06)	0.898(± 0.06)	0.736(± 0.15)
MLP	0.547(± 0.18)	0.617(± 0.43)	0.357(± 0.43)	0.852(± 0.05)	0.898(± 0.04)	0.735(± 0.09)
XGB	0.872(± 0.04)	0.935(± 0.04)	0.717(± 0.16)	0.862(± 0.05)	0.935(± 0.05)	0.681(± 0.16)

TABLE 3. Features selected by three feature selection methods.

CHI2	RFEcv	XGB
Age	Age	Age
DM	DM	DM
HTN	HTN	HTN
CRF	FH	FH
Diastolic Murmur	Sex	Atypical
Typical Chest Pain	Typical Chest Pain	Typical Chest Pain
Dyspnea	Nonanginal	Lymph
Function Class	Q Wave	Nonanginal
Atypical	St Elevation	St Depression
Nonanginal	T inversion	TG
Q Wave	HB	VHD
St Elevation	TG	PR
St Depression	Poor R Progression	T inversion
T inversion	EF-TTE	FBS
Poor R Progression	Current Smoker	ESR
FBS	DLP	EF-TTE
Airway disease	Lung rales	Na
Weak Peripheral Pulse	Dyspnea	BMI
BP	PR	LDL
Region RWMA	Region RWMA	Region RWMA
		WBC
		K
		LDL
		HB
		Neut
		HDL
		Length
		Weight
		CR
		PLT
		BP
		Obesity

Table 4 shows the accuracy of CHD diagnosis for different feature selection methods including CHI2, mutual information, variance, RFE, SVC [41], and LR, with different k values. The classification algorithm uses SVC ($C = 1.0$, kernel = ‘linear’). The accuracy increases first and then decrease with the increase of k value. The values ($k = 15, 17, 20$ and 22) are highlighted when the accuracy of the model higher than 90%. Table 4 shows that when $k = 15$ an accuracy of 91.1% is obtained by the LR.

**FIGURE 1.** Pearson correlation coefficient of the data (7 : 3).**FIGURE 2.** Pearson correlation coefficient of the data (4 : 6).

As shown in Table 5, the results of two representative feature selection methods including the wrapper and embedded approaches are compared. Table 5 shows that the best performance is achieved by the RFECV. And the results of RFECV have a smaller standard deviation. Therefore, the RFECV is decided as our feature selection method.

C. RESULTS FOR THE PROPOSED METHOD AND OTHER METHODS

The data are split into training data set and testing data set with a ratio of 7:3 and 4:6, respectively. Then, the training data set is used to train the model and calculate the PCC and MIC between each model pairwise. Fig. 1 and Fig. 2

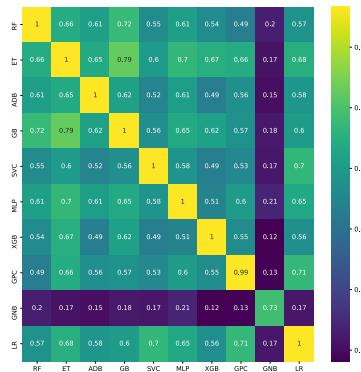
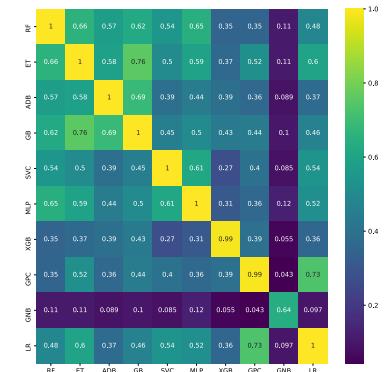
TABLE 4. Accuracy of CHD diagnosis for different feature selection methods with different k values.

K value	Feature selection methods					
	CHI2	Mutual information	Variance	RFE	SVC	LR
5	0.835(±0.04)	0.852(±0.07)	0.819(±0.05)	0.838(±0.04)	0.848(±0.04)	0.848(±0.04)
10	0.852(±0.07)	0.825(±0.07)	0.838(±0.05)	0.888(±0.06)	0.881(±0.06)	0.858(±0.07)
15	0.851(±0.08)	0.825(±0.06)	0.852(±0.07)	0.898(±0.07)	0.891(±0.06)	0.911(±0.06)
20	0.868(±0.07)	0.865(±0.06)	0.852(±0.07)	0.904(±0.06)	0.888(±0.06)	0.878(±0.07)
25	0.868(±0.07)	0.871(±0.07)	0.852(±0.06)	0.898(±0.05)	0.895(±0.05)	0.871(±0.06)
30	0.858(±0.07)	0.875(±0.07)	0.881(±0.06)	0.898(±0.06)	0.897(±0.05)	0.898(±0.07)
35	0.851(±0.07)	0.861(±0.05)	0.868(±0.05)	0.891(±0.06)	0.898(±0.05)	0.882(±0.08)
40	0.848(±0.07)	0.838(±0.08)	0.868(±0.08)	0.891(±0.06)	0.881(±0.07)	0.875(±0.07)
45	0.852(±0.06)	0.851(±0.09)	0.872(±0.06)	0.884(±0.07)	0.884(±0.07)	0.878(±0.07)
50	0.865(±0.07)	0.865(±0.08)	0.855(±0.08)	0.855(±0.08)	0.855(±0.08)	0.861(±0.08)
17	0.865(±0.08)	0.858(±0.07)	0.839(±0.07)	0.908(±0.06)	0.891(±0.05)	0.888(±0.06)
22	0.861(±0.07)	0.845(±0.06)	0.865(±0.05)	0.901(±0.05)	0.898(±0.05)	0.868(±0.07)

The error of K value is 1 because of the threshold of some selection method like Variance

TABLE 5. Performance of different testing methods with the RFECV and XGB feature selection method.

Method	RFECV			XGB		
	A_c	S_e	S_p	A_c	S_e	S_p
LR	0.901(±0.06)	0.930(±0.05)	0.826(±0.14)	0.868(±0.06)	0.921(±0.06)	0.736(±0.15)
RF	0.832(±0.04)	0.862(±0.07)	0.760(±0.14)	0.849(±0.05)	0.898(±0.05)	0.725(±0.15)
GNB	0.440(±0.12)	0.220(±0.15)	0.989(±0.03)	0.861(±0.08)	0.916(±0.05)	0.724(±0.22)
SVC	0.911(±0.05)	0.926(±0.04)	0.875(±0.12)	0.872(±0.06)	0.907(±0.06)	0.783(±0.13)
DT	0.799(±0.07)	0.865(±0.07)	0.633(±0.18)	0.789(±0.06)	0.856(±0.07)	0.622(±0.18)
KNN	0.855(±0.05)	0.870(±0.05)	0.814(±0.12)	0.855(±0.05)	0.907(±0.05)	0.728(±0.12)
ADB	0.881(±0.07)	0.926(±0.05)	0.768(±0.16)	0.859(±0.05)	0.903(±0.05)	0.747(±0.13)
GB	0.876(±0.05)	0.922(±0.05)	0.761(±0.16)	0.859(±0.06)	0.917(±0.05)	0.715(±0.13)
ET	0.829(±0.03)	0.871(±0.03)	0.726(±0.08)	0.833(±0.07)	0.875(±0.05)	0.729(±0.19)
MLP	0.908(±0.07)	0.945(±0.05)	0.815(±0.18)	0.868(±0.05)	0.912(±0.06)	0.760(±0.12)
XGB	0.868(±0.05)	0.944(±0.03)	0.681(±0.18)	0.865(±0.06)	0.930(±0.04)	0.703(±0.17)

**FIGURE 3.** MIC of the data (7 : 3).**FIGURE 4.** MIC of the data (4 : 6).

show the PCC of two different proportions of the same data. Fig. 3 and Fig. 4 show the MIC of two different proportions of the same data. In different proportions, the PCC and MIC of GNB always get the minimum values. So the model of GNB can be selected as one of the best combining classifiers. Seven optimal models (GNB, GB, RF, ET, ADB, MLP, XGB) are selected as the base-level model by Alg. 1 and Alg. 2. Table 6 shows the comparative results of our method and other different methods include the one proposed by the publisher of the dataset. It is shown that our proposed stacking

based model acquire significant improvements in nearly all measures of results. Our method achieves an accuracy, sensitivity, specificity and F1-score of 95.43%, 95.84%, 94.44%, 96.77%, respectively for the detection of CHD. The model parameters we use are given in Table 7.

D. RECEIVER OPERATING CHARACTERISTIC CURVES

As shown in Fig. 5, the receiver operating characteristic (ROC) is also used to evaluate the proposed method. The ROC curve is shown to be a simple yet complete recognition

TABLE 6. Performance comparison of different testing methods.

Method	A_c	S_e	S_p	$F1 - score$
LR	0.901(± 0.06)	0.930(± 0.05)	0.826(± 0.14)	0.931(± 0.04)
RF	0.833(± 0.04)	0.862(± 0.07)	0.760(± 0.14)	0.879(± 0.03)
GNB	0.440(± 0.12)	0.220(± 0.15)	0.989(± 0.03)	0.336(± 0.19)
SVC	0.911(± 0.05)	0.926(± 0.04)	0.875(± 0.12)	0.937(± 0.03)
DT	0.7987(± 0.07)	0.865(± 0.07)	0.633(± 0.18)	0.857(± 0.04)
KNN	0.855(± 0.05)	0.870(± 0.05)	0.814(± 0.12)	0.895(± 0.04)
ADB	0.881(± 0.07)	0.926(± 0.05)	0.768(± 0.16)	0.918(± 0.05)
GB	0.876(± 0.05)	0.922(± 0.05)	0.761(± 0.16)	0.914(± 0.04)
ET	0.829(± 0.03)	0.871(± 0.03)	0.726(± 0.08)	0.897(± 0.02)
MLP	0.908(± 0.07)	0.945(± 0.05)	0.815(± 0.18)	0.936(± 0.04)
XGB	0.868(± 0.05)	0.944(± 0.03)	0.681(± 0.18)	0.912(± 0.03)
Alizadehsani	0.941(± 0.0514)	0.963(None)	0.885(None)	None
Our method	0.9543(± 0.029)	0.9584(± 0.032)	0.9444(± 0.102)	0.9677(± 0.020)

TABLE 7. The parameters of the model.

Model name	Model parameters
RF	$n_estimators = 10, max_depth = None, min_samples_split = 2, random_state = 0$
GNB	$priors = None, var_smoothing = 1e - 09$
ADB	$n_estimators = 50, learning_rate = 1.0, algorithm = SAMME.R$
ET	$n_estimators = 100, criterion = gini, min_samples_split = 2$
GB	$loss = deviance, learning_rate = 0.1, n_estimators = 100$
MLP	$hidden_layer_sizes = (100,), activation = relu, solver = adam, alpha = 0.0001$
XGB	$random_state = 1, learning_rate = 0.05, n_estimators = 7, max_depth = 5$ $eta = 0.05, objective = binary : logistic$
LR	$C = 100, solver = newtoncg$

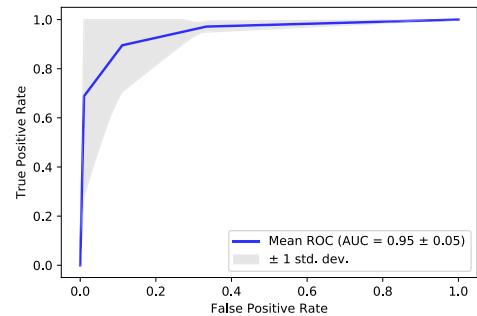
TABLE 8. Comparative results for the proposed method and other methods on the Statlog dataset.

Method	A_c	S_e	S_p	$F1 - score$
LSTSVM [42]	0.8559	None	None	None
NSABC [43]	0.92815	0.64511	None	None
FELM [44]	0.9444	0.9565	0.9355	None
KPCA-IBA-LSSVM [45]	0.9074	None	None	None
Our method	0.933(± 0.06)	0.892(± 0.10)	0.967(± 0.06)	0.921(± 0.07)

of all possible combinations of the relative frequencies of the correct and incorrect decisions [50]. A series of sensitivity and specificity are calculated. Then, the sensitivity is used as the ordinate and (1-specificity) as the abscissa to draw the curve. The larger the area under the curve are, the higher the diagnostic accuracy. On the ROC, the point closest to the upper left of the coordinate graph is the critical value with high sensitivity and specificity. Our proposed method has a high mean area under curve (AUC) up to 0.95, as shown in Fig. 5.

E. APPLYING OUR PROPOSED METHOD ON OTHER DATASETS

We test our method on the other three data sets to show that our proposed method does not depend on a particular data set. The first data set is the Statlog heart disease data set consisting of 270 subjects [51]. Each subject is presented with 14 features including age, sex, chestpaintype, restbloodpressure, serumcholesterol, fastingbloodsugar, reselectrocardiographic, maxheartrate, exerciseinduced, oldpeak, slope,

**FIGURE 5.** ROC of the 10-fold cross-validation with our method.

majorvessels, thal. Each of the subjects is classified into two categories: normal and abnormal. Table 8 shows the comparative results of our method and other different methods on the Statlog dataset. Our proposed method achieves an accuracy, sensitivity and specificity of 90.7%, 85.8%, 94.7%, respectively for the detection of CHD on the Statlog dataset.

TABLE 9. Comparative results for the proposed method and other methods on the big dataset.

Method	A_c	S_e	S_p	$F1 - score$
LR	0.680($\pm 0.35\%$)	0.570($\pm 0.73\%$)	0.788($\pm 0.80\%$)	0.638($\pm 0.45\%$)
RF	0.697($\pm 0.50\%$)	0.646($\pm 0.87\%$)	0.746($\pm 0.73\%$)	0.678($\pm 0.60\%$)
GNB	0.636($\pm 0.60\%$)	0.499($\pm 0.59\%$)	0.771($\pm 0.83\%$)	0.576($\pm 0.65\%$)
SVC	0.698($\pm 0.39\%$)	0.633($\pm 0.75\%$)	0.761($\pm 0.74\%$)	0.675($\pm 0.47\%$)
DT	0.682($\pm 0.5\%$)	0.538($\pm 3.2\%$)	0.823($\pm 2.8\%$)	0.626($\pm 1.5\%$)
KNN	0.693($\pm 0.74\%$)	0.665($\pm 0.98\%$)	0.721($\pm 0.84\%$)	0.682($\pm 0.81\%$)
ADB	0.719($\pm 0.36\%$)	0.683($\pm 0.60\%$)	0.755($\pm 0.76\%$)	0.707($\pm 0.37\%$)
GB	0.731($\pm 0.52\%$)	0.685($\pm 0.77\%$)	0.775($\pm 0.59\%$)	0.716($\pm 0.59\%$)
ET	0.689($\pm 0.55\%$)	0.640($\pm 0.85\%$)	0.737($\pm 0.56\%$)	0.671($\pm 0.66\%$)
MLP	0.714($\pm 0.41\%$)	0.661($\pm 4.76\%$)	0.765($\pm 4.57\%$)	0.695($\pm 1.63\%$)
XGB	0.726($\pm 0.53\%$)	0.680($\pm 0.87\%$)	0.771($\pm 0.58\%$)	0.711($\pm 0.63\%$)
Our method	0.732($\pm 0.58\%$)	0.693($\pm 0.67\%$)	0.770($\pm 0.77\%$)	0.719($\pm 0.58\%$)

TABLE 10. Comparative results for the proposed method and other methods on the SPECTF dataset.

Method	A_c	S_e	S_p	$F1 - score$
PCA-DHMM [46]	0.790	0.790	0.800	None
LMSVM [47]	0.8204	0.60	None	None
weighted vote-based classifier [48]	0.727	0.727	0.733	0.73
IT2FLS [49]	0.872	0.942	0.68	None
Our method	0.963($\pm 0.04\%$)	1.000($\pm 0.00\%$)	0.817($\pm 0.18\%$)	0.978($\pm 0.02\%$)

The second data set is the SPECTF heart data set which describes CHD diagnosis with cardiac single proton emission computed tomography (SPECT) images [51]. The data set contains 44 different types of continuous features derived from a total of 267 SPECT images that can be classified into 2 categories: normal and abnormal groups. Table 10 shows the comparative results of our method and other different methods on the SPECTF dataset. Our proposed method achieves an accuracy, sensitivity and specificity of 92.2%, 98.2%, 69.0%, respectively for the detection of CHD on the SPECTF dataset.

The third data set is the cardiovascular disease database (<https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>) of the Kaggle competition platform. It contains 70000 records of subjects with or without cardiovascular diseases. Each patient is expressed with 11 features which can be categorized into 3 types, namely, the subjective, objective, and examination features. The objective features provide factual information that consists of age, height, weight and gender. The examination features are results of medical examinations containing the systolic and diastolic blood pressures as well as the concentrations of cholesterol and glucose. The subjective features are information given by the subjects including the status of smoking, alcohol take, and physical activity. Our proposed method achieves an accuracy, sensitivity and specificity of 73.2%, 69.3%, 77.0%, respectively for the detection of CHD on the big dataset.

As shown in Table 8, Table 9 and Table 10, our proposed fusion model can significantly improve the performance. All algorithms are used with their default parameter settings. The three data sets we used have different features and have different kinds of relationships with each other. But, the performance exhibits the robustness of our algorithm,

indicating that the results achieved on Alizadehsani dataset are not stochastic.

V. DISCUSSION

In our proposed method, the technologies of stacking with all combinations searching are used for CHD diagnosis. The performance for the detection of CHD is higher than the known approaches in the literature. Additionally, the standard deviation of the results presented by our method is also minimum, indicating a better stability. It is undoubtedly more suitable for clinical application to make the model more stable under the premise of improving accuracy.

The data are normalized in order to compare the features more reasonably. Feature evaluation by filter, wrapper and embedded approaches is applied to select several relatively important features for the construction of our proposed model. The eight feature selection methods on the experimental results have important reference value for other researches in this field. When selecting the best combination in base-level, the PCC and MIC are calculated to find the classifier with the lowest correlation. This step greatly accelerates the training process of the model. Then, an enumeration process is developed to determine the other models in the optimal combination. In other words, with regard to the models included in the ultimately optimal combination, one of the models is determined via the PCC and MIC, and the other models are selected by the proposed enumeration process. This enumeration procedure is depicted as pseudocodes (Alg. 1 and Alg. 2) in the manuscript.

As far as we know, the high sensitivity and low specificity mean that more patients without CHD will be misdiagnosed. The proposed model acquires both high sensitivity and specificity which is clinically significant. The development of the

non-invasive measurements will be helpful for the people who suspected to have CHD. There will be no need for these people to suffer from CAG at the beginning. They can be tested with our method first. Later, the doctor can make a better decision on the issue whether the patient needs to undergo CAG.

The decisions of 7 classifiers are combined in order to produce accurate recognition results. Our proposed method exhibits the advantages of 7 ML algorithms. Due to the complementarity among multiple models, our proposed method in this study can be implemented as reference in other applications. We also acquire good results on other three datasets. The performance exhibits the robustness of our algorithm.

Our proposed method has two limitations. First, the model parameters we use are not optimal. We mainly focus on the way of searching model based on ten-fold cross-validation. The change of model parameters will also have great impact on the final results. Second, the training of multiple models in each level costs a lot of time and the method cannot narrow the search results quickly. Due to the 10-fold cross-validation and the using of multiple models, the experiments consume lots of time to train all model in each fold. However, when the model is trained, it will be convenient to test. Since the data set is too small, we add cross validation at the beginning to prevent overfitting. It complicates the framework. Nevertheless, doing so might have persuasive results.

The size of the dataset and useful features are the key issue in the field of ML. In future, a goal to meet is to build partnerships with hospitals to enlarge the data set of the CHD and extract more features from physiological signals.

CONFLICT OF INTEREST

There is no conflict of interest in this work.

ACKNOWLEDGMENT

The authors would like to thank Prof. C. Liu and Dr. L. Li for the enlightening discussions. They also thank Roohallah and so on the provider of Z-Alizadeh Sani dataset.

REFERENCES

- [1] J. Ouzan, J. C. Liehn, E. Beruben, J. Elaerts, J. Valeyre, and A. Bajoleit, "Diagnostic value of stress radionuclide angiography in coronary artery disease: A comparison of different interpretation criteria," *Eur. Heart J.*, vol. 9, no. 1, pp. 68–72, Jan. 1988.
- [2] R. Alizadehsani, M. J. Hosseini, A. Khosravi, F. Khozeimeh, M. Roshanzamir, N. Sarrafzadegan, and S. Nahavandi, "Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries," *Comput. Methods Programs Biomed.*, vol. 162, pp. 119–127, Aug. 2018.
- [3] M. Motwani et al., "Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis," *Eur. Heart J.*, vol. 38, no. 7, pp. 500–507, 2017, doi: [10.1093/eurheartj/ehw188](https://doi.org/10.1093/eurheartj/ehw188).
- [4] B. Ambale-Venkatesh, X. Yang, C. O. Wu, K. Liu, W. G. Hundley, R. McClelland, A. S. Gomes, A. R. Folsom, S. Shea, and E. Guallar, "Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis," *Circulat. Res.*, vol. 121, no. 9, pp. 1092–1101, 2017.
- [5] M. G. Feshki and O. S. Shijani, "Improving the heart disease diagnosis by evolutionary algorithm of PSO and feed forward neural network," in *Proc. Artif. Intell. Robot. (IRANOPEN)*, Apr. 2016, pp. 48–53.
- [6] A. Davari 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.
- [7] S. Vermekar, S. Nair, D. Vijayasan, and R. Ranjan, "A novel approach for classification of Normal/Abnormal phonocardiogram recordings using temporal signal analysis and machine learning," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 1141–1144.
- [8] M. Kumar, R. B. Pachori, and U. R. Acharya, "Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals," *Biomed. Signal Process. Control*, vol. 31, pp. 301–308, Jan. 2017.
- [9] L. Verma, S. Srivastava, and P. C. 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, Jun. 2016.
- [10] J. Sill, G. Takacs, L. Mackey, and D. Lin, "Feature-weighted linear stacking," 2009, *arXiv:0911.0460*. [Online]. Available: <https://arxiv.org/abs/0911.0460>
- [11] D. Yao, V. D. Calhoun, Z. Fu, Y. Du, and J. Sui, "An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment," *J. Neurosci. Methods*, vol. 302, pp. 75–81, May 2018.
- [12] L. Todorovski and S. Dzeroski, "A comparison of stacking with meta decision trees to other combining methods," in *Proc. IEEE Int. Conf. Data Mining*, Nov./Dec. 2001, pp. 144–147.
- [13] S. Avidan, "Ensemble tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 261–271, Feb. 2007.
- [14] B. Liu, S. Wang, R. Long, and K.-C. Chou, "IRSot-EL: Identify recombination spots with an ensemble learning approach," *Bioinformatics*, vol. 33, no. 1, pp. 35–41, Aug. 2016.
- [15] Q. Zhao and Y. Jiang, "LibEDM: A platform for ensemble based data mining," in *Proc. IEEE Int. Conf. Data Mining Workshop*, Dec. 2014, pp. 1250–1253.
- [16] D. Huang, C.-D. Wang, and J.-H. Lai, "Locally weighted ensemble clustering," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1460–1473, May 2018.
- [17] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.
- [18] A. Roohallah, H. Jafar, H. Mohammad Javad, M. Hoda, B. Reihane, G. Asma, B. Behdad, and S. Z. Alizadeh, "A data mining approach for diagnosis of coronary artery disease," *Comput. Methods Programs Biomed.*, vol. 111, no. 1, pp. 52–61, 2013.
- [19] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, no. 6, pp. 1157–1182, Jan. 2003.
- [20] H. Liu and H. Motoda, *Feature Extraction, Construction Selection*. Alphen aan den Rijn, The Netherlands: Kluwer, 1998.
- [21] A. Ulner and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *Eur. J. Oper. Res.*, vol. 206, no. 3, pp. 528–539, Nov. 2010.
- [22] C. Jie and R. Greiner, "Comparing Bayesian network classifiers," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2002–2012, 2013.
- [23] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.
- [24] E. Yu and S. Cho, "GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification," in *Proc. Int. Joint Conf. Neural Netw.*, 2003, pp. 2253–2257.
- [25] Q. Lin, H. Zhang, and Q. Li, "Wrapper feature selection algorithm based on MA-LSSVM," *J. Nanjing Univ. Sci. Technol.*, vol. 40, no. 1, pp. 10–16, 2016.
- [26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2016, pp. 785–794.
- [27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [28] A. K. Waljee, B. Liu, K. Sauder, J. Zhu, S. M. Govani, R. W. Stidham, and P. D. R. Higgins, "Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis," *Alimentary Pharmacol. Therapeutics*, vol. 47, no. 6, pp. 763–772, Jan. 2018.
- [29] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.
- [30] G. Rätsch, T. Onoda, and K.-R. Müller, "Soft margins for AdaBoost," *Mach. Learn.*, vol. 42, no. 3, pp. 287–320, 2001.
- [31] J. Chang, X. Zhao, E. Wong, Y. Wang, S. Lymeris, and S. Formenti, "A support vector machine (SVM) classifier enables prediction of optimal setup, prone versus supine, in left breast cancer patients," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 75, no. 3, pp. S218–S219, Nov. 2009.

- [32] N. Stamatis, D. Parthimos, and T. M. Griffith, "Forecasting chaotic cardiovascular time series with an adaptive slope multilayer perceptron neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 12, pp. 1441–1453, Dec. 1999.
- [33] M. Seeger, "Gaussian processes for machine learning," *Int. J. Neural Syst.*, vol. 14, no. 2, pp. 69–106, 2004.
- [34] X. Chen, F. Pereira, W. Lee, S. Strother, and T. Mitchell, "Exploring predictive and reproducible modeling with the single-subject FIAC dataset," *Hum. Brain Mapping*, vol. 27, no. 5, pp. 452–461, 2006.
- [35] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.
- [36] F. Niu, B. Recht, C. Re, and S. J. Wright, "Hogwild: A lock-free approach to parallelizing stochastic gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 693–701.
- [37] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?" *Mach. Learn.*, vol. 54, no. 3, pp. 255–273, Mar. 2004.
- [38] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.
- [39] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, no. 3, pp. 1531–1555, Nov. 2004.
- [40] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [41] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.
- [42] D. Tomar and S. Agarwal, "Feature selection based least square twin support vector machine for diagnosis of heart disease," *Int. J. Bio-Sci. Bio-Technol.*, vol. 6, no. 2, pp. 69–82, Apr. 2014.
- [43] B. Ragothaman and B. Sarojini, "A multi-objective non-dominated sorted artificial bee colony feature selection algorithm for medical datasets," *Indian J. Sci. Technol.*, vol. 9, no. 45, pp. 1–5, Dec. 2016.
- [44] K. B. Nahato, K. H. Nehemiah, and A. Kannan, "Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets," *Inform. Med. Unlocked*, vol. 2, pp. 1–11, 2016.
- [45] J. L. Jiang, S. Y. Li, M. L. Liao, and Y. Jiang, "Application in disease classification based on KPCA-IBA-LSSVM," *Procedia Comput. Sci.*, vol. 154, pp. 109–116, Jan. 2019.
- [46] R. Saraoğlu, "Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction," *Eng. Appl. Artif. Intell.*, vol. 25, no. 7, pp. 1523–1528, Oct. 2012.
- [47] Y. Rizk, N. Mitrí, and M. Awad, "A local mixture based SVM for an efficient supervised binary classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–8.
- [48] S. Bashir, U. Qamar, and F. H. Khan, "A multicriteria weighted vote-based classifier ensemble for heart disease prediction," *Comput. Intell.*, vol. 32, no. 4, pp. 615–645, Jun. 2015.
- [49] 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.
- [50] C. E. Metz, "Basic principles of ROC analysis," *Seminars Nucl. Med.*, vol. 8, no. 4, pp. 283–298, Oct. 1978.
- [51] D. Dua and C. Graff. *UCI Machine Learning Repository*. Accessed: 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>



CHANGCHUN LIU received the B.S. and M.S. degrees in automation from Shandong Industrial University (now merged into Shandong University), Jinan, China, in 1982 and 1987, respectively.

Since 2000, he has been a Professor in biomedical engineering with the School of Control Science and Engineering, Shandong University, where he was the Head of the Research Group of Non-invasive Evaluation of Cardiovascular Function. He has authored more than 100 articles. He also holds more than 15 Chinese invention patents. His research interests include novel solution for noninvasive detection of cardiovascular function, biomedical measurements, and biomedical devices.

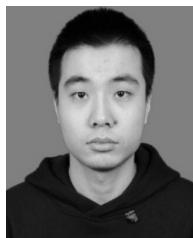
Prof. Liu was a recipient of the First Prize of the Science and Technology Award in technical invention awarded by the Ministry of Education, China, in 2003, and the Second Prize of the Science and Technology Award in invention awarded by Shandong Province, China, in 2016.



LIPING LI received the B.S. and Ph.D. degrees in biomedical engineering from Shandong University, China, in 2006 and 2011, respectively. She is currently a Lecturer with the School of Science and Engineering, Shandong University of Traditional Chinese Medicine. Her research interest includes physiological signal processing analysis.



WANG LI received the bachelor's degree from the Harbin Institute of Technology, in 2009. He has participated a successive postgraduate and doctoral programs of study with Chongqing University, from 2011 to 2016, where he majored in biomedical engineering. He is currently working with the Chongqing University of Technology. His research interests include noninvasive disease detection technology and device, and e-nose systems for disease detection.



LIANKE YAO received the B.S. degree in biomedical engineering from the Chongqing University of Technology, Chongqing, China, in 2012. He is currently pursuing the Ph.D. degree with Shandong University.

His research interests include biomedical signal processing QT interval variability and machine learning.



HAN LI received the B.S. degree in automation from Shandong University, Jinan, China, in 2014, where she is currently pursuing the Ph.D. degree with the School of Control Science and Engineering.

Her current research interests include the application of computational intelligence in the detection of cardiovascular diseases and biomedical signal processing.



HUAN ZHANG received the B.S. degree in electronic science and technology from Shanxi University, Taiyuan, China, in 2013. She is currently pursuing the Ph.D. degree with the School of Control Science and Engineering, Shandong University, China.

Her research interests include biomedical signal processing, machine learning, and early detection of coronary artery disease.



JIKUO WANG received the B.S. degree in electrical engineering and its automation and the M.S. degree in control science and engineering from Qufu Normal University, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree in biomedical engineering with Shandong University. His current research interests include deep learning and medical image analysis.