Hindawi
Computational and Mathematical Methods in Medicine
Volume 2017, Article ID 8272091, 11 pages
https://doi.org/10.1155/2017/8272091

Research Article

A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method

Xiao Liu,[1] Xiaoli Wang,[1] Qiang Su,[1] Mo Zhang,[2] Yanhong Zhu,[3] Qiugen Wang,[4] and Qian Wang[4]

[1]*School of Economics and Management, Tongji University, Shanghai, China*
[2]*School of Economics and Management, Shanghai Maritime University, Shanghai, China*
[3]*Department of Scientific Research, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China*
[4]*Trauma Center, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China*

Correspondence should be addressed to Xiaoli Wang; xiaoli-wang@tongji.edu.cn

Received 19 March 2016; Revised 12 July 2016; Accepted 1 August 2016; Published 3 January 2017

Academic Editor: Xiaopeng Zhao

Heart disease is one of the most common diseases in the world. The objective of this study is to aid the diagnosis of heart disease using a hybrid classification system based on the ReliefF and Rough Set (RFRS) method. The proposed system contains two subsystems: the RFRS feature selection system and a classification system with an ensemble classifier. The first system includes three stages: (i) data discretization, (ii) feature extraction using the ReliefF algorithm, and (iii) feature reduction using the heuristic Rough Set reduction algorithm that we developed. In the second system, an ensemble classifier is proposed based on the C4.5 classifier. The Statlog (Heart) dataset, obtained from the UCI database, was used for experiments. A maximum classification accuracy of 92.59% was achieved according to a jackknife cross-validation scheme. The results demonstrate that the performance of the proposed system is superior to the performances of previously reported classification techniques.

1. Introduction

Cardiovascular disease (CVD) is a primary cause of death. An estimated 17.5 million people died from CVD in 2012, representing 31% of all global deaths (http://www.who.int/mediacentre/factsheets/fs317/en/). In the United States, heart disease kills one person every 34 seconds [1].

Numerous factors are involved in the diagnosis of heart disease, which complicates a physician's task. To help physicians make quick decisions and minimize errors in diagnosis, classification systems enable physicians to rapidly examine medical data in considerable detail [2]. These systems are implemented by developing a model that can classify existing records using sample data. Various classification algorithms have been developed and used as classifiers to assist doctors in diagnosing heart disease patients.

The performances obtained using the Statlog (Heart) dataset [3] from the UCI machine learning database are compared in this context. Lee [4] proposed a novel supervised feature selection method based on the bounded sum of weighted fuzzy membership functions (BSWFM) and Euclidean distances and obtained an accuracy of 87.4%. Tomar and Agarwal [5] used the *F*-score feature selection method and the Least Square Twin Support Vector Machine (LSTSVM) to diagnose heart diseases, obtaining an average classification accuracy of 85.59%. Buscema et al. [6] used the Training with Input Selection and Testing (TWIST) algorithm to classify patterns, obtaining an accuracy of 84.14%. The Extreme Learning Machine (ELM) has also been used as a classifier, obtaining a reported classification accuracy of 87.5% [7]. The genetic algorithm with the Naïve Bayes classifier has been shown to have a classification accuracy of 85.87% [8]. Srinivas et al. [9] obtained an 83.70% classification accuracy using Naïve Bayes. Polat and Güneş [10] used the RBF kernel *F*-score feature selection method to detect heart disease. The LS-SVM classifier was used, obtaining a classification accuracy of 83.70%. In [11], the GA-AWAIS method was used for heart disease detection, with

a classification accuracy of 87.43%. The Algebraic Sigmoid Method has also been proposed to classify heart disease, with a reported accuracy of 85.24% [12]. Wang et al. [13] used linear kernel SVM classifiers for heart disease detection and obtained an accuracy of 83.37%. In [14], three distance criteria were applied in simple AIS, and the accuracy obtained on the Statlog (Heart) dataset was 83.95%. In [15], a hybrid neural network method was proposed, and the reported accuracy was 86.8%. Yan et al. [16] achieved an 83.75% classification accuracy using ICA and SVM classifiers. Şahan et al. [17] proposed a new artificial immune system named the Attribute Weighted Artificial Immune System (AWAIS) and obtained an accuracy of 82.59% using the k-fold cross-validation method. In [18], the k-NN, k-NN with Manhattan, feature space mapping (FSM), and separability split value (SSV) algorithms were used for heart disease detection, and the highest classification accuracy (85.6%) was obtained by k-NN.

From these works, it can be observed that feature selection methods can effectively increase the performance of single classifier algorithms in diagnosing heart disease [19]. Noisy features and dependency relationships in the heart disease dataset can influence the diagnosis process. Typically, there are numerous records of accompanied syndromes in the original datasets as well as a large number of redundant symptoms. Consequently, it is necessary to reduce the dimensions of the original feature set by a feature selection method that can remove the irrelevant and redundant features.

ReliefF is one of the most popular and successful feature estimation algorithms. It can accurately estimate the quality of features with strong dependencies and is not affected by their relations [20]. There are two advantages to using the ReliefF algorithm: (i) it follows the filter approach and does not employ domain specific knowledge to set feature weights [21, 22], and (ii) it is a feature weighting (FW) engineering technique. ReliefF assigns a weight to each feature that represents the usefulness of that feature for distinguishing pattern classes. First, the weight vector can be used to improve the performance of the lazy algorithms [21]. Furthermore, the weight vector can also be used as a method for ranking features to guide the search for the best subset of features [22–26]. The ReliefF algorithm has proved its usefulness in FS [20, 23], feature ranking [27], and building tree-based models [22], with an association rules-based classifier [28], in improving the efficiencies of the genetic algorithms [29] and with lazy classifiers [21].

ReliefF has excellent performance in both supervised and unsupervised learning. However, it does not help identify redundant features [30–32]. ReliefF algorithm estimates the quality of each feature according to its weight. When most of the given features are relevant to the concept, this algorithm will select most of them even though only some fraction is necessary for concept description [32]. Furthermore, the ReliefF algorithm does not attempt to determine the useful subsets of these weakly relevant features [33].

Redundant features increase dimensionality unnecessarily [34] and adversely affect learning performance when faced with shortage of data. It has also been empirically shown that removing redundant features can result in significant performance improvement [35]. Rough Set (RS) theory is a new mathematical approach to data analysis and data mining that has been applied successfully to many real-life problems in medicine, pharmacology, engineering, banking, financial and market analysis, and others [36]. The RS reduction algorithm can reduce all redundant features of datasets and seek the minimum subset of features to attain a satisfactory classification [37].

There are three advantages to combining ReliefF and RS (RFRS) approach as an integrated feature selection system for heart disease diagnosis.

(i) The RFRS method can remove superfluous and redundant features more effectively. The ReliefF algorithm can select relevant features for disease diagnosis; however, redundant features may still exist in the selected relevant features. In such cases, the RS reduction algorithm can remove remaining redundant features to offset this limitation of the ReliefF algorithm.

(ii) The RFRS method helps to accelerate the RS reduction process and guide the search of the reducts. Finding a minimal reduct of a given information system is an NP-hard problem, as was demonstrated in [38]. The complexity of computing all reducts in an information system is rather high [39]. On one hand, as a data preprocessing tool, the features revealed by the ReliefF method can accelerate the operation process by serving as the input for the RS reduction algorithm. On the other hand, the weight vector obtained by the ReliefF algorithm can act as a heuristic to guide the search for the reducts [25, 26], thus helping to improve the performance of the heuristic algorithm [21].

(iii) The RFRS method can reduce the number and improve the quality of reducts. Usually, more than one reduct exists in the dataset; and larger numbers of features result in larger numbers of reducts [40]. The number of reducts will decrease if superfluous features are removed using the ReliefF algorithm. When unnecessary features are removed, more important features can be extracted, which will also improve the quality of reducts.

It is obvious that the choice of an efficient feature selection method and an excellent classifier is extremely important for the heart disease diagnosis problem [41]. Most of the common classifiers from the machine learning community have been used for heart disease diagnosis. It is now recognized that no single model exists that is superior for all pattern recognition problems, and no single technique is applicable to all problems [42]. One solution to overcome the limitations of a single classifier is to use an ensemble model. An ensemble model is a multiclassifier combination model that results in more precise decisions because the same problem is solved by several different trained classifiers, which reduces the variance of error estimation [43]. In recent years, ensemble learning has been employed to increase classification accuracies beyond the level that can be achieved by individual classifiers [44, 45]. In this paper, we used an ensemble classifier to evaluate the feature selection model.

To improve the efficiency and effectiveness of the classification performance for the diagnosis of heart disease, we propose a hybrid classification system based on the ReliefF and RS (RFRS) approach in handling relevant and redundant

features. The system contains two subsystems: the RFRS feature selection subsystem and a classification subsystem. In the RFRS feature selection subsystem, we use a two-stage hybrid modeling procedure by integrating ReliefF with the RS (RFRS) method. First, the proposed method adopts the ReliefF algorithm to obtain feature weights and select more relevant and important features from heart disease datasets. Then, the feature estimation obtained from the first phase is used as the input for the RS reduction algorithm and guide the initialization of the necessary parameters for the genetic algorithm. We use a GA-based search engine to find satisfactory reducts. In the classification subsystem, the resulting reducts serve as the input for the chosen classifiers. Finally, the optimal reduct and performance can be obtained.

To evaluate the performance of the proposed hybrid method, a confusion matrix, sensitivity, specificity, accuracy, and ROC were used. The experimental results show that the proposed method achieves very promising results using the jack knife test.

The main contributions of this paper are summarized as follows.

(i) We propose a feature selection system to integrate the ReliefF approach with the RS method (RFRS) to detect heart disease in an efficient and effective way. The idea is to use the feature estimation from the ReliefF phase as the input and heuristics for the RS reduction phase.

(ii) In the classification system, we propose an ensemble classifier using C4.5 as the base classifier. Ensemble learning can achieve better performance at the cost of computation than single classifiers. The experimental results show that the ensemble classifier in this paper is superior to three common classifiers.

(iii) Compared with three classifiers and previous studies, the proposed diagnostic system achieved excellent classification results. On the Statlog (Heart) dataset from the UCI machine learning database [3], the resulting classification accuracy was 92.59%, which is higher than that achieved by other studies.

The rest of the paper is organized as follows. Section 2 offers brief background information concerning the ReliefF algorithm and RS theory. The details of the diagnosis system implementation are presented in Section 3. Section 4 describes the experimental results and discusses the proposed method. Finally, conclusions and recommendations for future work are summarized in Section 5.

2. Theoretical Background

2.1. Basic Concepts of Rough Set Theory. Rough Set (RS) theory, which was proposed by Pawlak, in the early 1980s, is a new mathematical approach to addressing vagueness and uncertainty [46]. RS theory has been applied in many domains, including classification system analysis, pattern reorganization, and data mining [47]. RS-based classification algorithms are based on equivalence relations and have been used as classifiers in medical diagnosis [37, 46]. In this paper, we primarily focus on the RS reduction algorithm, which can reduce all redundant features of datasets and seek the minimum subset of features necessary to attain a satisfactory classification [37]. A few basic concepts of RS theory are defined [46, 47] as follows.

Definition 1. U is a certain set that is referred to as the universe; R is an equivalence relation in U. The pair $A = (U, R)$ is referred to as an approximation space.

Definition 2. $P \subset R$, $\cap P$ (the intersection of all equivalence relations in P) is an equivalence relation, which is referred to as the R-indiscernibility relation, and it is represented by $\mathrm{Ind}(R)$.

Definition 3. Let X be a certain subset of U. The least composed set in R that contains X is referred to as the best upper approximation of X in R and represented by $R^{-}(X)$; the greatest composed set in R contained in X is referred to as the best lower approximation of X in R, and it is represented by $R\_(X)$.

$$\begin{aligned} R\_(X) &= \{x \in U : [x]_R \subset X\}, \\ R^{-}(X) &= \{x \in U : [x]_R \cap X \neq \phi\}. \end{aligned} \tag{1}$$

Definition 4. An information system is denoted as

$$S = (U, A, V, F), \tag{2}$$

where U is the universe that consists of a finite set of n objects, $A = \{C \cup D\}$, in which C is a set of condition attributes and D is a set of decision attributes, V is the set of domains of attributes, and F is the information function for each $a \in A$, $x \in U$, $F(x, a) \in V_a$.

Definition 5. In an information system, C and D are sets of attributes in U. $X \in U/\mathrm{ind}(Q)$, and $\mathrm{pos}_p(Q)$, which is referred to as a positive region, is defined as

$$\mathrm{pos}_p(Q) = \cup P\_(X). \tag{3}$$

Definition 6. P and Q are sets of attributes in U, $P, Q \subseteq A$, and the dependency $r_p(Q)$ is defined as

$$r_p(Q) = \frac{\mathrm{card}\left(\mathrm{pos}_p(Q)\right)}{\mathrm{card}(U)}. \tag{4}$$

Card (X) denotes the cardinality of X. $0 \leq r_p(Q) \leq 1$.

Definition 7. P and Q are sets of attributes in U, $P, Q \subseteq A$, and the significance of a_i is defined as

$$\mathrm{sig}(\{a_i\}) = r_p(Q) - r_{p-a_i}(Q). \tag{5}$$

2.2. ReliefF Algorithm. Many feature selection algorithms have been developed; ReliefF is one of the most widely used and effective algorithms [48]. ReliefF is a simple yet efficient procedure for estimating the quality of features in problems with dependencies between features [20]. The pseudocode of ReliefF algorithm is listed in Algorithm 1.

ReliefF algorithm
Input: A decision table $S = (U, P, Q)$
Output: the vector W of estimations of the qualities of features
(1) set all weights $W[A] := 0.0$;
(2) for $i := 1$ to m do begin
(3) randomly select a sample R_i;
(4) find k nearest hits H_j;
(5) for each class $C \neq \text{class}(R_i)$ do
(6) from class C find k nearest misses $M_j(C)$;
(7) for $A := 1$ to a do
(8) $W[A] := W[A] - \sum_{j=1}^{k} \text{diff}(A, R_i, H_j)/mk + \sum_{C \neq \text{class}(R_i)} [P(C)/1 - P(\text{class}(R_i)) \sum_{j=1}^{k} \text{diff}(A, R_i, M_j(C))]/mk$;
(9) end;

Algorithm 1: Pseudocode of ReliefF.

3. Proposed System

3.1. Overview. The proposed hybrid classification system consists of two main components: (i) feature selection using the RFRS subsystem and (ii) data classification using the classification system. A flow chart of the proposed system is shown in Figure 1. We describe the preprocessing and classification systems in the following subsections.

3.2. RFRS Feature Selection Subsystem. We propose a two-phase feature selection method based on the ReliefF algorithm and the RS (RFRS) algorithm. The idea is to use the feature estimation from the ReliefF phase as the input and heuristics for the subsequent RS reduction phase. In the first phase, we adopt the ReliefF algorithm to obtain feature weights and select important features; in the second phase, the feature estimation obtained from the first phase is used to guide the initialization of the parameters required for the genetic algorithm. We use a GA-based search engine to find satisfactory reducts.

The RFRS feature selection subsystem consists of three main modules: (i) data discretization, (ii) feature extraction using the ReliefF algorithm, and (iii) feature reduction using the heuristic RS reduction algorithm we propose.

3.2.1. Data Discretization. RS reduction requires categorical data. Consequently, data discretization is the first step. We used an approximate equal interval binning method to bin the data variables into a small number of categories.

3.2.2. Feature Extraction by the ReliefF Algorithm. Module 2 is used for feature extraction by the ReliefF algorithm. To deal with incomplete data, we change the diff function. Missing feature values are treated probabilistically [20]. We calculate the probability that two given instances have different values for a given feature conditioned over the class value [20]. When one instance has an unknown value, then

$$\text{diff}(A, I_1, I_2) = 1 - P(\text{value}(A, I_2) \mid \text{class}(I_1)). \quad (6)$$

When both instances have unknown values, then

$$\text{diff}(A, I_1, I_2) = 1 - \sum_{V}^{\#\text{values}(A)} (P(V \mid \text{class}(I_1)) \times P(V \mid \text{class}(I_2))). \quad (7)$$

Conditional probabilities are approximated by relative frequencies in the training set. The process of feature extraction is shown as follows.

The Process of Feature Extraction Using ReliefF Algorithm

Input. A decision table $S = (U, P, Q)$, $P = \{a_1, a_2, \ldots, a_m\}$, $Q = \{d_1, d_2, \ldots, d_n\}$ $(m \geq 1, n \geq 1)$.

Output. The selected feature subset $K = \{a_1, a_2, \ldots, a_k\}(1 \leq k \leq m)$.

Step 1. Obtain the weight matrix of each feature using ReliefF algorithm$W = \{w_1, w_2, \ldots, w_i, \ldots, w_m\}$ $(1 \leq i \leq m)$.

Step 2. Set a threshold, δ.

Step 3. If $w_i > \delta$, then feature a_i is selected.

3.2.3. Feature Reduction by the Heuristic RS Reduction Algorithm. The evaluation result obtained by the ReliefF algorithm is the feature rank. A higher ranking means that the feature has stronger distinguishing qualities and a higher weight [30]. Consequently, in the process of reduct searching, the features in the front rank should have a higher probability of being selected.

We proposed the RS reduction algorithm by using the feature estimation as heuristics and a GA-based search engine to search for the satisfactory reducts. The pseudocode of the algorithm is provided in Algorithm 2. The algorithm was implemented in MATLAB R2014a.

3.3. Classification Subsystem. In the classification subsystem, the dataset is split into training sets and corresponding test

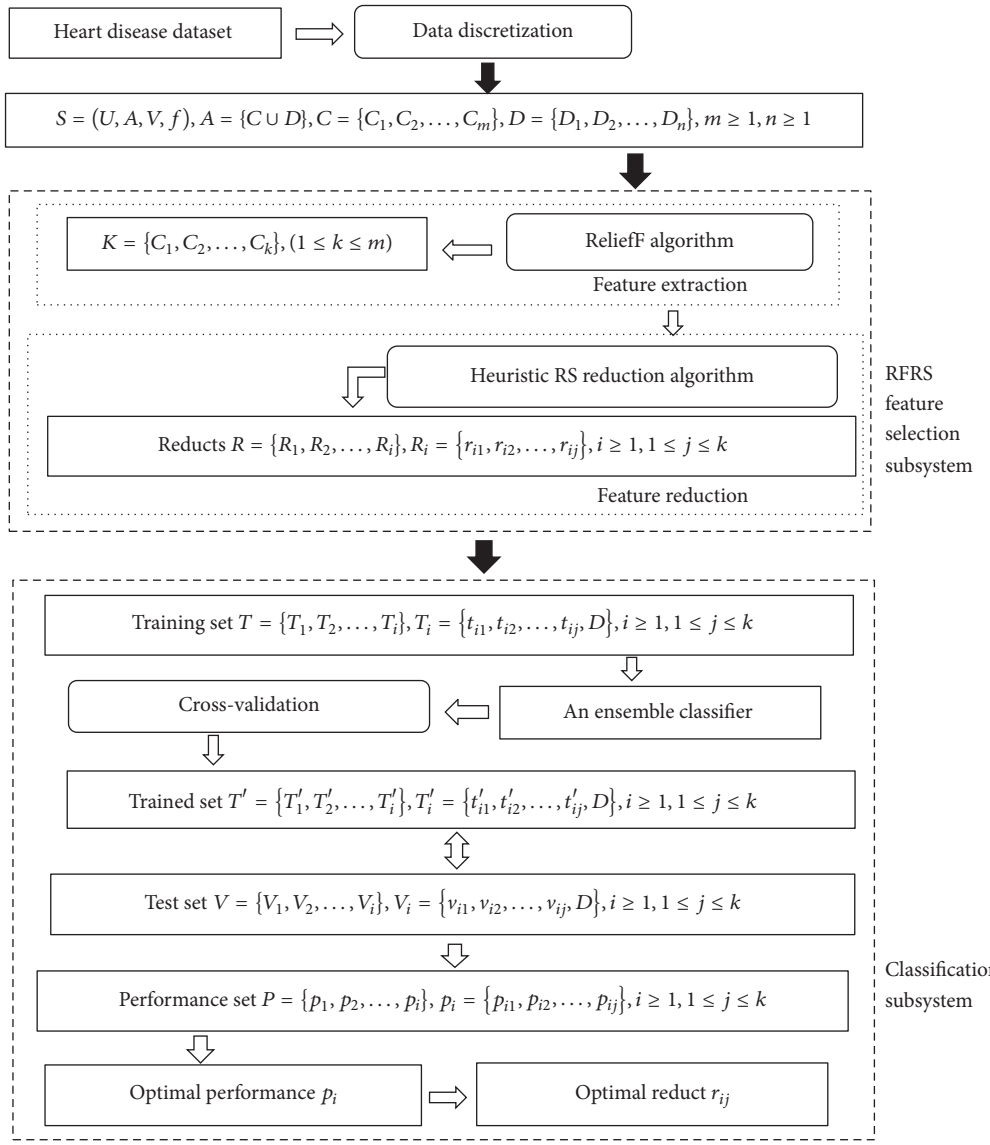

Figure 1: Structure of RFRS-based classification system.

sets. The decision tree is a nonparametric learning algorithm that does not need to search for optimal parameters in the training stage and thus is used as a weak learner for ensemble learning [49]. In this paper, the ensemble classifier uses the C4.5 decision tree as the base classifier. We use the boosting technique to construct ensemble classifiers. Jackknife cross-validation is used to increase the amount of data for testing the results. The optimal reduct is the reduct that obtains the best classification accuracy.

4. Experimental Results

4.1. Dataset. The Statlog (Heart) dataset used in our work was obtained from the UCI machine learning database [3]. This dataset contains 270 observations and 2 classes: the presence and absence of heart disease. The samples include 13 condition features, presented in Table 1. We denote the 13 features as C_1 to C_{13}.

4.2. Performance Evaluation Methods

4.2.1. Confusion Matrix, Sensitivity, Specificity, and Accuracy. A confusion matrix [50] contains information about actual and predicted classifications performed by a classification system. The performance of such systems is commonly evaluated using the data in the matrix. Table 2 shows the confusion matrix for a two-class classifier.

In the confusion matrix, TP is the number of true positives, representing the cases with heart disease that are correctly classified into the heart disease class. FN is the number of false negatives, representing cases with heart disease that

Heuristic RS reduction algorithm
Input: a decision table $S = (U, C, D)$, $C = \{c_1, c_2, \ldots, c_m\}$, $D = \{d_1, d_2, \ldots, d_n\}$
Output: Red
Step 1. Return Core
(1) Core $\leftarrow$ {}
(2) For $i = 1$ to m
(3) Select c_i from C;
(4) Calculate Ind(C), Ind($C - c_i$) and Ind(D);
(5) Calculate $\text{pos}_C(D)$, $\text{pos}_{C-c_i}(D)$ and $r_C(D)$;
(6) Calculate Sig($\{a_i\}$), Sig($\{a_i\}$) $= r_C(D) - r_{C-c_i}(D)$;
(7) If sig($\{c_i\}$) $\neq 0$
(8) core = core $\cup$ $\{c_i\}$;
(9) End if
(10) End for
Step 2. Return Red
(1) Red = Core
(2) $C' = C -$ Red
(3) While Sig(Red, D) $\neq$ Sig(C, D) do
Compute the weight of each feature c in C' using the ReliefF algorithm;
Select a feature c according to its weight, let Red=Red $\cup$ $\{c\}$;
Initialize all the necessary parameters for the GA-based search engine according to the results of the last step and search for satisfactory reducts;
End while

Algorithm 2: Pseudocode of heuristic RS reduction algorithm.

are classified into the healthy class. TN is the number of true negatives, representing healthy cases that are correctly classified into the healthy class. Finally, FP is the number of false positives, representing the healthy cases that are incorrectly classified into the heart disease class [50].

The performance of the proposed system was evaluated based on sensitivity, specificity, and accuracy tests, which use the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) terms [33]. These criteria are calculated as follows [41]:

$$\begin{aligned}
\text{Sensitivity (Sn)} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \\
\text{Specificity (Sp)} &= \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100\%, \\
\text{Accuracy (Acc)} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%.
\end{aligned} \tag{8}$$

4.2.2. Cross-Validation. Three cross-validation methods, namely, subsampling tests, independent dataset tests, and jackknife tests, are often employed to evaluate the predictive capability of a predictor [51]. Among the three methods, the jackknife test is deemed the least arbitrary and the most objective and rigorous [52, 53] because it always yields a unique outcome, as demonstrated by a penetrating analysis in a recent comprehensive review [54, 55]. Therefore, the jackknife test has been widely and increasingly adopted in many areas [56, 57].

Accordingly, the jackknife test was employed to examine the performance of the model proposed in this paper. For jackknife cross-validation, each sequence in the training dataset is, in turn, singled out as an independent test sample and all the parameter rules are calculated based on the remaining samples, without including the one being treated as the test sample.

4.2.3. Receiver Operating Characteristics (ROC). The receiver operating characteristic (ROC) curve is used for analyzing the prediction performance of a predictor [58]. It is usually plotted using the true positive rate versus the false positive rate, as the discrimination threshold of classification algorithm is varied. The area under the ROC curve (AUC) is widely used and relatively accepted in classification studies because it provides a good summary of a classifier's performance [59].

4.3. Results and Discussion

4.3.1. Results and Analysis on the Statlog (Heart) Dataset. First, we used the equal interval binning method to discretize the original data. In the feature extraction module, the number of k-nearest neighbors in the ReliefF algorithm was set to 10, and the threshold, δ, was set to 0.02. Table 3 summarizes the results of the ReliefF algorithm. Based on these results, C_5 and C_6 were removed. In Module 3, we obtained 15 reducts using the heuristic RS reduction algorithm implemented in MATLAB 2014a.

Trials were conducted using 70%–30% training-test partitions, using all the reduced feature sets. Jackknife cross-validation was performed on the dataset. The number of desired base classifiers k was set to 50, 100, and 150. The calculations were run 10 times, and the highest classification

Table 1: Feature information of Statlog (Heart) dataset.

Feature	Code	Description	Domain	Data type	Mean	Standard deviation
Age	C_1	—	29–77	Real	54	9
Sex	C_2	Male, female	0, 1	Binary	—	—
Chest pain type	C_3	Angina, asymptomatic, abnormal	1, 2, 3, 4	Nominal	—	—
Resting blood pressure	C_4	—	94–200	Real	131.344	17.862
Serum cholesterol in mg/dl	C_5	—	126–564	Real	249.659	51.686
Fasting blood sugar > 120 mg/dl	C_6	—	0, 1	Binary	—	—
Resting electrocardiographic results	C_7	Norm, abnormal, hyper	0, 1, 2	Nominal	—	—
Maximum heart rate achieved	C_8	—	71–202	Real	149.678	23.1666
Exercise-induced angina	C_9	—	0, 1	Binary	—	—
Old peak = ST depression induced by exercise relative to rest	C_{10}	—	0–6.2	Real	1.05	1.145
Slope of the peak exercise ST segment	C_{11}	Up, flat, down	1, 2, 3	Ordered	—	—
Number of major vessels (0–3) colored by fluoroscopy	C_{12}	—	0, 1, 2, 3	Real	—	—
Thal	C_{13}	Normal, fixed defect, reversible defect	3, 6, 7	Nominal	—	—

Table 2: The confusion matrix.

	Predicted patients with heart disease	Predicted healthy persons
Actual patients with heart disease	True positive (TP)	False negative (FN)
Actual healthy persons	False positive (FP)	True negative (TN)

performances for each training-test partition are provided in Table 4.

In Table 4, R_2 obtains the best test set classification accuracy (92.59%) using the ensemble classifiers when k = 100. The training process is shown in Figure 2. The training and test ROC curves are shown in Figure 3.

4.3.2. Comparison with Other Classifiers. In this section, our ensemble classification method is compared with the individual C4.5 decision tree and Naïve Bayes and Bayesian Neural Networks (BNN) methods. The C4.5 decision tree and Naïve Bayes are common classifiers. Bayesian Neural Networks (BNN) is a classifier that uses Bayesian regularization to train feed-forward neural networks [60] and has better performance than pure neural networks. The classification accuracy results of the four classifiers are listed in Table 5. The ensemble classification method has better performance than the individual C4.5 classifier and the other two classifiers.

4.3.3. Comparison of the Results with Other Studies. We compared our results with the results of other studies. Table 6 shows the classification accuracies of our study and previous methods.

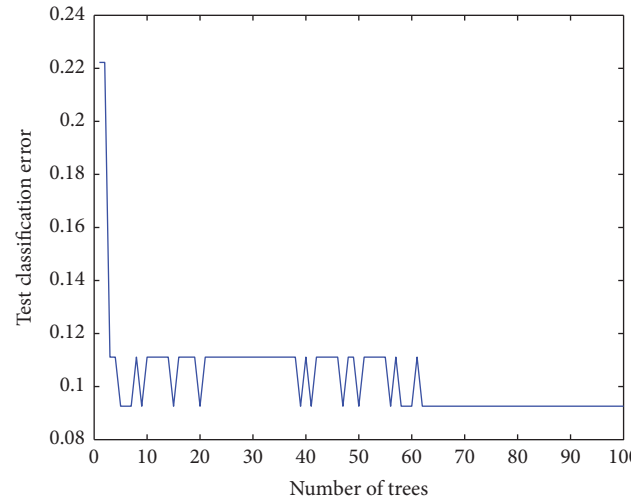

Figure 2: Training process of R_7.

The results show that our proposed method obtains superior and promising results in classifying heart disease patients. We believe that the proposed RFRS-based classification system can be exceedingly beneficial in assisting physicians in making accurate decisions.

Table 3: Results of the ReliefF algorithm.

Feature	C_2	C_{13}	C_7	C_{12}	C_9	C_3	C_{11}	C_{10}	C_8	C_4	C_1	C_6	C_5
Weight	0.172	0.147	0.126	0.122	0.106	0.098	0.057	0.046	0.042	0.032	0.028	0.014	0.011

Table 4: Performance values for different reduced subset.

Code	Reduct	Number	Test classification accuracy (%) Ensemble classifier		
			K	Sn	Sp	ACC
R_1	$C_3, C_4, C_7, C_8, C_{10}, C_{12}, C_{13}$	7	50	83.33	87.5	85.19
			100	83.33	95.83	88.89
			150	86.67	83.33	85.19
R_2	$C_1, C_3, C_7, C_8, C_{11}, C_{12}, C_{13}$	7	50	86.67	91.67	88.89
			100	93.33	87.50	92.59
			150	93.33	87.04	90.74
R_3	$C_1, C_2, C_4, C_7, C_8, C_9, C_{12}$	7	50	86.67	83.33	85.19
			100	93.33	79.17	87.04
			150	80	91.67	85.19
R_4	$C_1, C_4, C_7, C_8, C_{10}, C_{11}, C_{12}, C_{13}$	8	50	86.67	83.33	85.19
			100	93.33	83.33	88.89
			150	86.67	87.5	87.04

Table 5: Classification results using the four classifiers.

Classifiers	Test classification accuracy of R_2 (%)		
	Sn	Sp	Acc
Ensemble classifier (k = 50)	86.67	91.67	88.89
Ensemble classifier (k = 100)	93.33	87.50	92.59
Ensemble classifier (k = 150)	93.33	87.04	90.74
C4.5 tree	93.1	80	87.03
Naïve Bayes	93.75	68.18	83.33
Bayesian Neural Networks (BNN)	93.75	72.72	85.19

5. Conclusions and Future Work

In this paper, a novel ReliefF and Rough Set- (RFRS-) based classification system is proposed for heart disease diagnosis. The main novelty of this paper lies in the proposed approach: the combination of the ReliefF and RS methods to classify heart disease problems in an efficient and fast manner. The RFRS classification system consists of two subsystems: the RFRS feature selection subsystem and the classification subsystem. The Statlog (Heart) dataset from the UCI machine learning database [3] was selected to test the system. The experimental results show that the reduct R_2 (C_1, C_3, C_7, C_8, C_{11}, C_{12}, C_{13}) achieves the highest classification accuracy (92.59%) using an ensemble classifier with the C4.5 decision tree as the weak learner. The results also show that the RFRS method has superior performance compared to three common classifiers in terms of ACC, sensitivity, and specificity. In addition, the performance of the proposed system is superior to that of existing methods in the literature. Based on empirical analysis, the results indicate that the proposed classification system can be used as a promising alternative tool in medical decision making for heart disease diagnosis.

However, the proposed method also has some weaknesses. The number of the nearest neighbors (k) and the weight threshold (θ) are not stable in the ReliefF algorithm [20]. One solution to this problem is to compute estimates for all possible numbers and take the highest estimate of each feature as the final result [20]. We need to perform more experiments to find the optimal parameter values for the ReliefF algorithm in the future.

Table 6: Comparison of our results with those of other studies.

Author	Method	Classification accuracy (%)
Our study	RFRS classification system	92.59
Lee [4]	Graphical characteristics of BSWFM combined with Euclidean distance	87.4
Tomar and Agarwal [5]	Feature selection-based LSTSVM	85.59
Buscema et al. [6]	TWIST algorithm	84.14
Subbulakshmi et al. [7]	ELM	87.5
Karegowda et al. [8]	GA + Naïve Bayes	85.87
Srinivas et al. [9]	Naïve Bayes	83.70
Polat and Güneş [10]	RBF kernel F-score + LS-SVM	83.70
Özşen and Güneş [11]	GA-AWAIS	87.43
Helmy and Rasheed [12]	Algebraic Sigmoid	85.24
Wang et al. [13]	Linear kernel SVM classifiers	83.37
Özşen and Güneş [14]	Hybrid similarity measure	83.95
Kahramanli and Allahverdi [15]	Hybrid neural network method	86.8
Yan et al. [16]	ICA + SVM	83.75
Şahan et al. [17]	AWAIS	82.59
Duch et al. [18]	*K*NN classifier	85.6

BSWFM: bounded sum of weighted fuzzy membership functions; LSTSVM: Least Square Twin Support Vector Machine; TWIST: Training with Input Selection and Testing; ELM: Extreme Learning Machine; GA: genetic algorithm; SVM: support vector machine; ICA: imperialist competitive algorithm; AWAIS: attribute weighted artificial immune system; *K*NN: *k*-nearest neighbor.

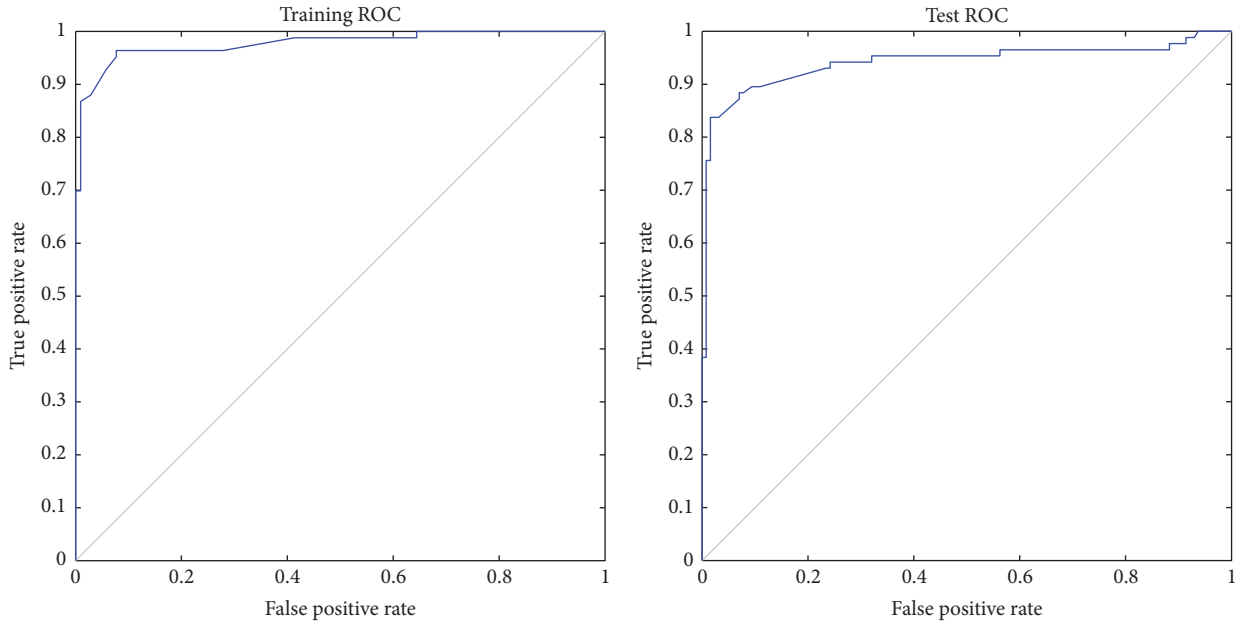

Figure 3: ROC curves for training and test sets.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant no. 71432007).

References

[1] K. D. Kochanek, J. Xu, S. L. Murphy, A. M. Miniño, and H.-C. Kung, "Deaths: final data for 2009," *National Vital Statistics Reports*, vol. 60, no. 3, pp. 1–116, 2011.

[2] H. Temurtas, N. Yumusak, and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8610–8615, 2009.

[3] UCI Repository of Machine Learning Databases, http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29.

[4] S.-H. Lee, "Feature selection based on the center of gravity of BSWFMs using NEWFM," *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 482–487, 2015.

[5] D. Tomar and S. Agarwal, "Feature selection based least square twin support vector machine for diagnosis of heart disease," *International Journal of Bio-Science and Bio-Technology*, vol. 6, no. 2, pp. 69–82, 2014.

[6] M. Buscema, M. Breda, and W. Lodwick, "Training with Input Selection and Testing (TWIST) algorithm: a significant advance in pattern recognition performance of machine learning," *Journal of Intelligent Learning Systems and Applications*, vol. 5, no. 1, pp. 29–38, 2013.

[7] C. V. Subbulakshmi, S. N. Deepa, and N. Malathi, "Extreme learning machine for two category data classification," in *Proceedings of the IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT '12)*, pp. 458–461, Ramanathapuram, India, August 2012.

[8] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Feature subset selection problem using wrapper approach in supervised learning," *International Journal of Computer Applications*, vol. 1, no. 7, pp. 13–17, 2010.

[9] K. Srinivas, B. K. Rani, and A. Govrdhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," *International Journal on Computer Science and Engineering*, vol. 2, pp. 250–255, 2010.

[10] K. Polat and S. Güneş, "A new feature selection method on classification of medical datasets: kernel F-score feature selection," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10367–10373, 2009.

[11] S. Özşen and S. Güneş, "Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems," *Expert Systems with Applications*, vol. 36, no. 1, pp. 386–392, 2009.

[12] T. Helmy and Z. Rasheed, "Multi-category bioinformatics dataset classification using extreme learning machine," in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '09)*, pp. 3234–3240, Trondheim, Norway, May 2009.

[13] S.-J. Wang, A. Mathew, Y. Chen, L.-F. Xi, L. Ma, and J. Lee, "Empirical analysis of support vector machine ensemble classifiers," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6466–6476, 2009.

[14] S. Özşen and S. Güneş, "Effect of feature-type in selecting distance measure for an artificial immune system as a pattern recognizer," *Digital Signal Processing*, vol. 18, no. 4, pp. 635–645, 2008.

[15] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 82–89, 2008.

[16] G. Yan, G. Ma, J. Lv, and B. Song, "Combining independent component analysis with support vector machines," in *Proceedings of the in 1st International Symposium on Systems and Control in Aerospace and Astronautics (ISSCAA '06)*, pp. 493–496, Harbin, China, January 2006.

[17] S. Şahan, K. Polat, H. Kodaz, and S. Günes, "The medical applications of attribute weighted artificial immune system (AWAIS): diagnosis of heart and diabetes diseases," *Artificial Immune Systems*, vol. 3627, pp. 456–468, 2005.

[18] W. Duch, R. Adamczak, and K. Grabczewski, "A new methodology of extraction, optimization and application of crisp and fuzzy logical rules," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 277–306, 2001.

[19] M. P. McRae, B. Bozkurt, C. M. Ballantyne et al., "Cardiac ScoreCard: a diagnostic multivariate index assay system for predicting a spectrum of cardiovascular disease," *Expert Systems with Applications*, vol. 54, pp. 136–147, 2016.

[20] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, 2003.

[21] D. Wettschereck, D. W. Aha, and T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms," *Artificial Intelligence Review*, vol. 11, no. 1–5, pp. 273–314, 1997.

[22] I. Kononenko, E. Šimec, and M. R. Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997.

[23] K. Kira and L. Rendell, "A practical approach to feature selection," in *Proceedings of the International Conference on Machine Learning*, pp. 249–256, Morgan Kaufmann, Aberdeen, Scotland, 1992.

[24] I. Kononenko, "Estimating attributes: analysis and extension of ReliefF," in *Machine Learning: ECML-94: European Conference on Machine Learning Catania, Italy, April 6–8, 1994 Proceedings*, vol. 784 of *Lecture Notes in Computer Science*, pp. 171–182, Springer, Berlin, Germany, 1994.

[25] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Heuristic search over a ranking for feature selection," in *Proceedings of 8th International Work-Conference on Artificial Neural Networks (IWANN '05), Barcelona, Spain, June 2005*, vol. 3512 of *Lectures Notes in Computer Science*, pp. 742–749, Springer, Berlin, Germany, 2005.

[26] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "Filter approach feature selection methods to support multi-label learning based on ReliefF and information gain," in *Advances in Artificial Intelligence—SBIA 2012: 21th Brazilian Symposium on Artificial Intelligence, Curitiba, Brazil, October 20–25, 2012. Proceedings*, vol. 7589 of *Lectures Notes in Computer Science*, pp. 72–81, Springer, Berlin, Germany, 2012.

[27] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Fast feature ranking algorithm," in *Proceedings of the Knowledge-Based Intelligent Information and Engineering Systems (KES '03)*, pp. 325–331, Springer, Berlin, Germany, 2003.

[28] V. Jovanoski and N. Lavrac, "Feature subset selection in association rules learning systems," in *Proceedings of Analysis, Warehousing and Mining the Data*, pp. 74–77, 1999.

[29] J. J. Liu and J. T. Y. Kwok, "An extended genetic rule induction algorithm," in *Proceedings of the Congress on Evolutionary Computation*, pp. 458–463, LA Jolla, Calif, USA, July 2000.

[30] L.-X. Zhang, J.-X. Wang, Y.-N. Zhao, and Z.-H. Yang, "A novel hybrid feature selection algorithm: using ReliefF estimation for GA-Wrapper search," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 380–384, IEEE, Xi'an, China, November 2003.

[31] Z. Zhao, L. Wang, and H. Liu, "Efficient spectral feature selection with minimum redundancy," in *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, AAAI, Atlanta, Ga, USA, 2010.

[32] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Advances in Neural Information Processing Systems*, pp. 1813–1821, MIT Press, 2010.

[33] S.-Y. Jiang and L.-X. Wang, "Efficient feature selection based on correlation measure between continuous and discrete features," *Information Processing Letters*, vol. 116, no. 2, pp. 203–215, 2016.

[34] M. J. Kearns and U. V. Vazirani, *An Introduction to Computational Learning Theory*, MIT Press, Cambridge, Mass, USA, 1994.

[35] R. Bellman, *Adaptive Control Processes: A Guided Tour, RAND Corporation Research Studies*, Princeton University Press, Princeton, NJ, USA, 1961.

[36] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybernetics and Systems*, vol. 29, no. 7, pp. 661–688, 1998.

[37] A.-E. Hassanien, "Rough set approach for attribute reduction and rule generation: a case of patients with suspected breast cancer," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 11, pp. 954–962, 2004.

[38] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory, System Theory, Knowledge Engineering and Problem Solving*, R. Słowinski, Ed., vol. 11, pp. 331–362, Kluwer Academic, Dordrecht, Netherlands, 1992.

[39] Z. Pawlak, "Rough set approach to knowledge-based decision support," *European Journal of Operational Research*, vol. 99, no. 1, pp. 48–57, 1997.

[40] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 459–471, 2007.

[41] C. Ma, J. Ouyang, H.-L. Chen, and X.-H. Zhao, "An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach," *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 985789, 14 pages, 2014.

[42] A. H. El-Baz, "Hybrid intelligent system-based rough set and ensemble classifier for breast cancer diagnosis," *Neural Computing and Applications*, vol. 26, no. 2, pp. 437–446, 2015.

[43] J.-H. Eom, S.-C. Kim, and B.-T. Zhang, "AptaCDSS-E: a classifier ensemble-based clinical decision support system for cardiovascular disease level prediction," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2465–2479, 2008.

[44] Y. Ma, *Ensemble Machine Learning: Methods and Applications*, Springer, New York, NY, USA, 2012.

[45] S. A. Etemad and A. Arya, "Classification and translation of style and affect in human motion using RBF neural networks," *Neurocomputing*, vol. 129, pp. 585–595, 2014.

[46] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[47] Z. Pawlak, "Rough sets and intelligent data analysis," *Information Sciences*, vol. 147, no. 1–4, pp. 1–12, 2002.

[48] Y. Huang, P. J. McCullagh, and N. D. Black, "An optimization of ReliefF for classification in large datasets," *Data and Knowledge Engineering*, vol. 68, no. 11, pp. 1348–1356, 2009.

[49] J. Sun, M.-Y. Jia, and H. Li, "AdaBoost ensemble for financial distress prediction: an empirical comparison with data from Chinese listed companies," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9305–9312, 2011.

[50] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, no. 2-3, pp. 271–274, 1998.

[51] W. Chen and H. Lin, "Prediction of midbody, centrosome and kinetochore proteins based on gene ontology information," *Biochemical and Biophysical Research Communications*, vol. 401, no. 3, pp. 382–384, 2010.

[52] K.-C. Chou and C.-T. Zhang, "Prediction of protein structural classes," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 30, no. 4, pp. 275–349, 1995.

[53] W. Chen and H. Lin, "Identification of voltage-gated potassium channel subfamilies from sequence information using support vector machine," *Computers in Biology and Medicine*, vol. 42, no. 4, pp. 504–507, 2012.

[54] K.-C. Chou and H.-B. Shen, "Recent progress in protein subcellular location prediction," *Analytical Biochemistry*, vol. 370, no. 1, pp. 1–16, 2007.

[55] P. Feng, H. Lin, W. Chen, and Y. Zuo, "Predicting the types of J-proteins using clustered amino acids," *BioMed Research International*, vol. 2014, Article ID 935719, 8 pages, 2014.

[56] K.-C. Chou and H.-B. Shen, "ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information," *Biochemical and Biophysical Research Communications*, vol. 376, no. 2, pp. 321–325, 2008.

[57] K.-C. Chou and Y.-D. Cai, "Prediction of membrane protein types by incorporating amphipathic effects," *Journal of Chemical Information and Modeling*, vol. 45, no. 2, pp. 407–413, 2005.

[58] F. Tom, "ROC graphs: notes and practical considerations for researchers," *Machine Learning*, vol. 31, pp. 1–38, 2004.

[59] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.

[60] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proceedings of the International Conference on Neural Networks*, vol. 3, pp. 1930–1935, Houston, Tex, USA, 1997.