

SVM Feature Selection Based Rotation Forest Ensemble Classifiers to Improve Computer-Aided Diagnosis of Parkinson Disease

Akin Ozcift

Received: 24 January 2011 / Accepted: 27 February 2011 / Published online: 10 March 2011
© Springer Science+Business Media, LLC 2011

Abstract Parkinson disease (PD) is an age-related deterioration of certain nerve systems, which affects movement, balance, and muscle control of clients. PD is one of the common diseases which affect 1% of people older than 60 years. A new classification scheme based on support vector machine (SVM) selected features to train rotation forest (RF) ensemble classifiers is presented for improving diagnosis of PD. The dataset contains records of voice measurements from 31 people, 23 with PD and each record in the dataset is defined with 22 features. The diagnosis model first makes use of a linear SVM to select ten most relevant features from 22. As a second step of the classification model, six different classifiers are trained with the subset of features. Subsequently, at the third step, the accuracies of classifiers are improved by the utilization of RF ensemble classification strategy. The results of the experiments are evaluated using three metrics; classification accuracy (ACC), Kappa Error (KE) and Area under the Receiver Operating Characteristic (ROC) Curve (AUC). Performance measures of two base classifiers, i.e. KStar and IBk, demonstrated an apparent increase in PD diagnosis accuracy compared to similar studies in literature. After all, application of RF ensemble classification scheme improved PD diagnosis in 5 of 6 classifiers significantly. We, numerically, obtained about 97% accuracy in RF ensemble of IBk (a K-Nearest Neighbor variant) algorithm, which is a quite high performance for Parkinson disease diagnosis.

Keywords Rotation forest · Ensemble classification · Parkinson · Breast cancer · Diabetes · Feature selection · Support vector machine · Computer aided diagnosis

Introduction

Knowledge discovery is the concept of computer science that describes the process of automatically searching data for useful patterns. These patterns are in general used to train a computer aided diagnosis system to make decision for the classification of previously unseen instances of data [1]. Computer aided diagnosis (CAD) systems in medicine is an emerging field of high importance for providing prognosis of diseases. CAD based classification schemes with adequate accuracies might substitute invasive approaches, for benefit of clients [2]. CAD systems support medical analysis in variety of applications to identify disease with the use of machine learning algorithms [3]. Large numbers of disease prognosis applications in the literature utilize supervised learning strategies. Numerous CAD systems make use of single base learners. However, ensemble learning techniques have potential to improve classification accuracy of base learning algorithms. In the literature, some CAD applications making use of an ensemble learning technique are breast cancer [4], heart disease [5], hepatitis [6], diabetes [7], thyroid [8] and Alzheimer [9]. In this study, an important neurological disorder, i.e. Parkinson's disease, is analyzed with a RF ensemble classification scheme.

Parkinson disease characteristically begins about 40 years of age with slow progressive attitude. The cause behind PD is not known exactly and some of the typical symptoms of disease are rigidity, bradykinesia, asymmetric onset, micrographia, decreased olfaction, and postural instability and dysphonia

A. Ozcift (✉)
Gaziantep Vocational School of Higher Education,
Computer Programming Division, University of Gaziantep,
Gaziantep, Turkey
e-mail: akinozcift@hotmail.com

[10, 11]. PD has no medical treatment up-to-date and some medication is only available to alleviate the symptoms of disease [12]. For this reason, PD clients are dependent of clinically being monitored regularly. In order to reduce cost of monitoring clients, internet based remote monitoring techniques might be useful for PD sufferers. This kind of remote screening could be possible with the observation of dysphonia, i.e. one of the most significant symptoms of PD. This medical term is defined as the presence of some form of vocal impairment in PD clients and the researches show that approximately 90% of persons with PD have such vocal evidence. A dysphonic person exhibit impairment in the normal production of vocal sounds [13]. More explicitly, dysphonia is a phonation disorder and a dysphonic voice is in general hoarse, weak, breathy, harsh and rough [14].

The *dysphonic* indicators of PD make speech measurements an important part of diagnosis. The conventional techniques of such measures are fundamental frequency of vocal oscillation, absolute sound pressure level amount to characterize loudness of speech, jitter measure to examine variation in speech, shimmer test to detect variation in amplitude of speech, pitch period entropy (a new sensitive measure specific to examine variations of PD client speech) and noise-to-harmonics ratios to quantify the amplitude of noise in relation to tonal components of speech. While making screening of PD possible, these measures are used to examine the extent of dysphonia as well [11, 14].

In our study, we use dysphonia measures of Little et al. [11] to evaluate performance of our algorithms in diagnosis of PD. The dataset consists of voice measurements from 31 people, 23 with PD. In the dataset, each column is a particular voice measure attribute, and every row corresponds to one of 195 voice records from these clients. The number of columns in the dataset consists of 22 speech measure attributes and a status column to specify the class (PD or healthy) of the instance.

A general issue in classification algorithms is that use of large number of features in the classification may cause irrelevant features to exert undue influence on the classification decisions because of the finite size of the training sample. Therefore, this study first focuses on the problem of identification of relevant features that is expected to yield an accurate classifier [15].

Our feature selection (FS) algorithm uses a SVM based strategy to find powerful features of PD dataset while reducing the number of attributes to train classifiers. The detail of mentioned FS strategy is given Section 2.

Two significant works that focus on the classification of PD are performed by Little et al. [11] and Das [10]. Little et al. uses a Support Vector Machine in their study and Das discusses a Neural Network based prediction strategy. The reported accuracies of two classification strategies are 91.4% and 92.9% respectively.

In this context, the goal of this study is to improve PD diagnosis accuracy with the use of SVM feature selection based RF ensemble classification approach.

Linear SVM based feature selection

In statistical learning theory, *dimension reduction* is the process of reducing the number of features of data under consideration. There are in literature mainly two approaches for feature reduction of high dimensional datasets, i.e. Feature Selection (FS) and Feature Extraction (FE).

Feature extraction strategies mainly consist of data transformation techniques to represent a high feature space with a lower dimensional vector while attempting to preserve information content of original data. The data transformation may be linear as in the case of Principal Component Analysis (PCA) or it is a non-linear conversion with the utilization of feed-forward Neural Networks [16]. One of widely used linear dimension reduction techniques is Principal Component Analysis that performs linear mapping of data to a lower dimension maximizing the variance of transformed data [17].

The second set of Feature Reduction methods, i.e. *feature selection* approaches, comprise techniques to select most discriminative features from high feature set to obtain a smaller subset of variables. FS techniques attempts to remove redundant features from original dataset by the use of mainly three variable selection approaches: (i) feature rankers that only consider intrinsic properties of the data, (ii) wrapper methods that embed the model hypothesis search within the feature subset search and (iii) embedded techniques where the search for an optimal subset of features is built into the classifier construction [18].

In Linear SVM classifier, classes of two patterns are linearly separated by the use of a maximum margin hyperplane that is defined as the hyperplane maximizing the sum of the distances between the hyperplane and the margin (its closest points of each of the two classes). If the two classes are not linearly separable, then a variant of SVM (a soft-margin SVM) is used. While training a linear SVM classifier, a decision function similar to Eq. 1 is formed [19].

$$f(x_1, \dots, x_N) = \sum_{i=1}^N w_i x_i + b \quad (1)$$

In case of feature selection problem, a linear SVM model (defined by Eq. 1) is used to decide the relevance of each feature. In the decision function (1), the absolute value of larger weight (w_j) means that the j th feature (x_j) is a more powerful feature. In consequence, features are ranked according to $|w_j|$ [20]. To evaluate worth of a feature,

SVM assigns a weight to each feature and this absolute value is used to rank each variable. In this scheme, FS for a multiclass problem is handled by ranking attributes for each class separately using a one-vs.-all method and then “dealing” from the top of each pile to give a final ranking [21].

The mentioned SVM-FS strategy is implemented in WEKA data mining environment and feature dimension is reduced from 23 to 11. While ranking features with SVM, we used ten fold cross validation approach. Subsequent to SVM-FS feature ranking procedure, the problem was to determine a threshold value to select top subset of features. This threshold is determined by using the following procedure:

- i). Six machine learning algorithms are selected for evaluation
- ii). Each algorithm is trained starting with the most powerful (first feature in the list having highest weight) attribute and accuracies of the classifiers are calculated
- iii). The other features are added to dataset one by one to inspect the change in the accuracies of algorithms. This feature addition is continued up to the accuracy of the classifiers tend to increase.
- iv). If accuracy of the classifiers discontinues increasing the procedure is stopped and the resultant feature subset is assumed to include the most relevant features

The selected ten feature names from PD dataset, except status column, are given in Table 1, with respect to their ranks from highest to lower.

Rotation forest ensemble classifier approach

The classification performance of CAD systems is undoubtedly important and therefore, the accuracy of relatively weak classifiers is to be improved by means of

Table 1 Feature set selected by support vector machine

No	Feature name
1	Spread1
2	MDVP_Fo_Hz
3	D2
4	Spread2
5	MDVP_Fhi_Hz
6	MDVP_APQ
7	DFA
8	HNR
9	PPE
10	RPDE

relevant approaches such as ensemble of conventional algorithms. Generally, use of multiple (ensemble) classifiers is better than the use of single algorithms for higher classification accuracy. In this manner, an ensemble of classifiers is, in general, expected to be more accurate than a single classifier [22]. Primarily, an ensemble classifier consists of base classifiers that learn a target function by combining their prediction mutually. A few ensemble learning approaches seen in literature is composite classifier systems, mixture of experts, consensus aggregation, dynamic classifier selection, classifier fusion and committees of neural networks [23].

In particular, Rotation Forest ensemble approach is a newly proposed multi classifier scheme and the framework of the algorithm might be described as follows:

In RF, the training dataset for each base classifier is created by randomly splitting F (features of dataset) into K subsets (K is a parameter of the algorithm) and PCA is applied to those subsets. In order to preserve information content of data, all principal components are retained. Hence, new attributes of base classifier is obtained with K axis rotations. RF algorithm accomplishes diversity (by extracting features for each base classifier) and individual accuracy (by keeping all principal components and using the whole data set to train each base classifier) within the ensemble [24]. The structure of the algorithms is more clearly explained with pseudocode in Fig. 1 and this flow is inspired from the developers of algorithm [25].

Being a relatively new ensemble approach, Rotation Forest is used in a few biomedical classification applications and in some experimental studies compared to other well known ensemble techniques [25, 26].

In Section 4, we will explain the classifiers used in the evaluation of RF ensemble scheme and corresponding performance estimation metrics.

Base classifiers and classification metrics used in this study

In this study, we used three different types of classifiers, two neural network architectures, two lazy learners and two decision trees for creating six RF ensembles of these base classifiers.

The evaluated classifiers are from Weka data mining software (a java based rich open source environment) and the information about classifiers are given in brief terms as follows:

- i). Multi Layer Perceptron (MLP) : This algorithm is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate output. It is a

Fig. 1 Rotation forest algorithm pseudocode

```

PHASE 1: TRAINING
Parameters of pseudocode
     $X$ : the samples of dataset
     $Y$ : the labels of training set
     $L$ : the classifier number forming the ensemble
     $K$ : the feature subset number
     $\{w_1, \dots, w_c\}$ : classes of dataset
Do for all classifiers (from classifier 1 to classifier L)
    → Setup the rotation matrix  $R_i^a$ 
        - Split randomly  $F$  into  $K$  subsets as  $F_{i,j}$  (where  $j=1 \dots K$ )
        - For all  $K$  subsets do the following
            i: Setup the dataset  $X_{i,j}$  using  $F_{i,j}$  subsets
            ii: Remove a random class subset from  $X_{i,j}$ 
            iii: Select 75% bootstrap sample from  $X_{i,j}$  to get  $X'_{i,j}$ 
            iv: Transform  $X'_{i,j}$  with PCA to get  $C_{i,j}$  coefficients
        - Assemble rotation matrix  $R_i$  using all  $C_{i,j}$  coefficients
        - Rearrange columns of  $R_i$  to match the feature order in  $F$  and build  $R_i^a$ 
    → Build classifier  $D_i$  by training with dataset  $(X, R_i^a, Y)$ 

PHASE 2: CLASSIFICATION
    → For each given instance  $x$  calculate the confidence of  $x$  that belongs to class  $w_j$  using average combination method
    → Assign  $x$  to class having the largest confidence
  
```

variant of the traditional linear perceptron and it uses three or more layers of neurons with nonlinear activation functions. This algorithm is more powerful than the perceptron, since it can distinguish the linearly inseparable data [27]. Another special kind of artificial neural network, Radial basis function (RBF) networks, uses radial basis functions as activation functions. RBF networks typically have three layers such as an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. A radial basis function is a special kind of class function and its characteristics is that the response decreases or increases monotonically with distance from a central point [27].

iii) **Lazy Learner Classifiers:** This group of classifiers is also known instance-based learners. The classifiers, instead of performing explicit generalization, compare new problem instances with instances seen in training that are stored in memory. In this aspect, an instance-based learner constructs hypotheses directly from the training instances themselves. K-Nearest Neighbor (KNN) is one of the most popular learner of this type and in our study we used two classifiers from Weka software; IBk a Weka implementation of KNN and KSTAR (or K*) a kind of KNN classifier that use entropic distance measure in its classification [28].

Table 2 The classification performances of algorithms and their corresponding ensembles

Classifier	ACC(%)	eACC(%)	Diff (%)	Kappa	eKappa	Diff	AUC	eAUC	Diff
MLP	88.21	90.8	2.59	0.69	0.75	0.06	0.89	0.91	0.02
RBF	88.71	88.71	0	0.68	0.68	0	0.89	0.89	0
LADTree	89.23	92.82	3.59	0.71	0.81	0.1	0.89	0.93	0.04
J48	86.15	92.3	6.15	0.63	0.78	0.15	0.86	0.92	0.06
KSTAR	94.91	96.41	1.51	0.86	0.91	0.05	0.95	0.97	0.02
IBk	95.89	96.93	1.04	0.89	0.92	0.03	0.96	0.97	0.01

iii) **Decision Tree Learners:** Decision tree learners, as the name implies, uses decision tree diagrams to model decisions and their possible consequences, including chance event outcomes and resource costs. A decision tree classifier is a kind of multistage decision making and the basic idea of this approach is to break up a complex decision into a union of several simpler decisions. The approach hopes to obtain a final solution to resemble the intended desired solution [29]. We used a multi-class alternating decision tree (LADTree that uses a logit boost strategy) and J48 decision tree classifier (java implementation of C4.5 algorithm) implementation.

In this study, we utilized three performance metrics, *accuracy* (ACC), *kappa statistic* (KS) value and *Area under the Receiver Operating Characteristic* (AUC) value, in order to evaluate the performances of our classifiers. Some brief information about classification calculations in general and about the mentioned performance evaluation metrics will be given as follows:

Most of the biomedical disease diagnosis problems deal with two class predictions. The goal of such problems is to map data samples into one of the groups, i.e. benign or malignant, with possible maximum estimation accuracy. For such a two-class problem, the outcomes are labeled as positive (*p*) or negative (*n*). The possible outcomes with respect to this classification scheme is frequently defined in statistical learning as *true positive* (TP), *false positive* (FP), *true negative* (TN) and *false negative* (FN). These four outcomes are used to derive most of the well known performance metrics such as sensitivity, specificity, accuracy, positive prediction value, F-measure, AUC and ROC curve [30].

Accuracy (ACC) is a widely used metric to determine class discrimination ability of classifiers, and it is calculated using Eq. 2.

$$ACC = (TP + TN)/(P + N) \quad (2)$$

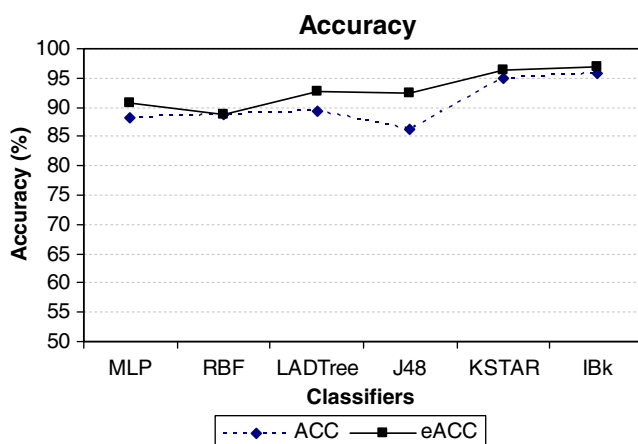


Fig. 2 The ACC measures of base classifiers and corresponding ensembles

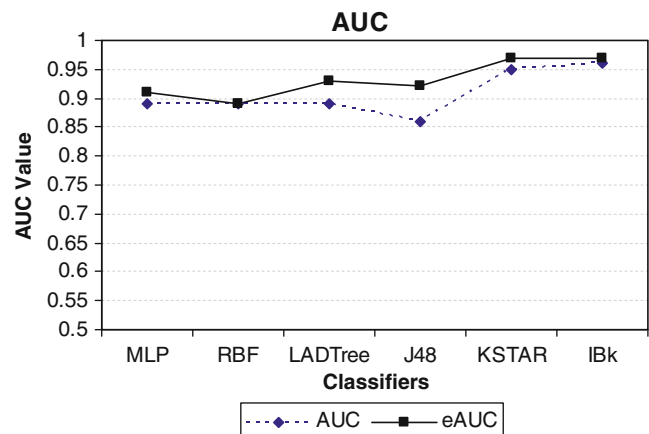


Fig. 3 The AUC measures of base classifiers and corresponding ensembles

This is one of primary metrics in evaluating classifier performances and it is defined as the percentage of test samples that are correctly classified by the algorithm. The inspection of ACC values is easy for an experimental study, and we selected this index to consider it with AUC, and Kappa error mutually.

Furthermore, AUC value is widely used in classification studies with relevant acceptance and it is a good summary of the performance of the classifier. AUC value is calculated from the area of under the ROC curve. ROC curves are usually plotted using *true positives rate* versus *false positives rate*, as the discrimination threshold of classification algorithm is varied. In this aspect, Since a ROC curve compares the classifiers' performance across the entire range of class distributions and error costs; an AUC value is accepted to be a good measure of comparative performances of classification algorithms and it is calculated with Eq. 3 [31].

$$AUC = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (3)$$

As the third metric of this study, Kappa error or Cohen's Kappa Statistics value is being used to compare the performances of classifiers. In classification algorithm performance comparisons, just using the percentage of misses as the single meter for accuracy can give misleading results. The cost of error must also be taken into account, while making such assessments. Kappa error, in this aspect, is a good measure to inspect classifications that may be due

Table 3 Class structure of breast cancer and diabetes datasets

Dataset	Number of features	Healthy instances	Disease instances
Breast Cancer	11	458	241
Diabetes	20	500	268

Table 4 The classification performances of algorithms and their corresponding ensembles for benchmark datasets

Classifier	Breast cancer			Diabetes		
	ACC(%)	eACC(%)	Diff (%)	ACC(%)	eACC(%)	Diff (%)
MLP	95.28	96.00	0.72	75.39	75.52	0.13
RBF	95.85	96.00	0.15	75.39	76.30	0.91
LADTree	95.57	97.30	1.73	74.09	76.30	2.21
J48	94.56	97.14	2.58	73.83	76.17	2.34
KSTAR	95.42	95.28	-0.14	69.14	72.00	2.86
IBk	95.13	95.42	0.29	70.18	70.44	0.3

to chance. In general, Kappa error takes values between $(-1,1)$. As the Kappa value calculated for classifiers approaches to '1', then the performance of the classifier is assumed to be more realistic rather than by chance. Therefore, in the performance analysis of classifiers, Kappa error is a recommended metric to consider for evaluation purposes and it is calculated with Eq. 4 [32].

$$KE = \frac{p_0 - p_c}{1 - p_c} \quad (4)$$

In Eq. 4, p_0 demonstrates total agreement probability and p_c agreement probability due to chance.

Experimental results for parkinson disease diagnosis

In this section, the results of the experiments are given in Table 2. In the table, 'e' means ensemble classifier performances and 'Diff' is used to mean 'Difference'.

Table 2 shows obviously the performance increase, i.e. difference calculations are positive, in three metrics with the utilization of ensemble Rotation Forest approach, except RBF that has no change in its performance indexes.

In Table 2, it is easy to see the affirmative correlation between metrics. If a classifier has a good accuracy, its other two metrics supports this relative increase. For instance, while Table 2 is observed for RBF algorithm, it is seen that the classifier has no improvement in accuracy by the use of its corresponding ensemble. And Kappa statistics value also supports this with no change at all. However, while IBk metrics are examined, the ACC and AUC values are observed to increase. This increase is also supported by Kappa statistic that the increase is confidential and not by chance.

In order the performances of classifiers to be seen visually, we developed two figures that take ACC and AUC values of classifiers. Figure 2, for ACC index and Fig. 3 for AUC index are given as follows:

Experimental results for two benchmark datasets

In order to demonstrate the efficiency of the rotation forest ensemble classification algorithm, we will make use of two benchmark datasets, i.e. Breast Cancer Wisconsin (Orig-

inal) and Diabetes datasets, from University of California-Irvine (UCI) machine learning repository. We will provide class structure of the datasets and we will only give classification accuracies of algorithms for the sake of convenience.

Breast cancer dataset was obtained from the University of Wisconsin Hospitals [33] and the class structure of the breast cancer is given in Table 3.

Diabetes dataset records were obtained using an automatic electronic recording device and paper records [34]. The detailed structure of the diabetes dataset is given in Table 3.

We used the same flow as in Parkinson diagnosis for the experiments of two datasets. The features of datasets are selected using SVM and then the resultant datasets are evaluated with our six base algorithms and corresponding ensembles. While making the experiments, 10-fold cross validation scheme is used to get resultant accuracies in Table 4.

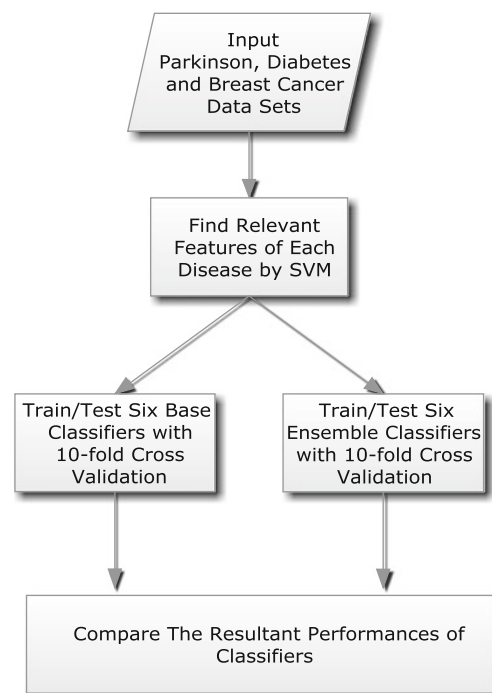
**Fig. 4** Flow of the algorithm to analyze three datasets

Table 4 shows an apparent performance increase, i.e. difference calculations are positive, for two datasets with the utilization of ensemble Rotation Forest approach, except KSTAR that has a small decrease in its accuracy.

The experimental flow of this study consists of two main steps and the layout of the algorithm for three datasets is given in Fig. 4.

Conclusion and remarks

In literature there are two specific works in the classification of PD, namely Little et al. [11] and Das [10]. Little et al. uses a kernel Support Vector Machine for their classification and Das discusses a Neural Network classification scheme. The reported performances of the two classifications respectively are 91.4% and 92.9% as percentage accuracies.

When Table 2 is examined, it is seen that ACC performances of our two base classifiers, i.e. KSTAR with ACC 94.91% and IBk with ACC 95.89%, is better than the reported accuracies. While the Rotation Forest ensemble of these classifiers is examined from Table 2, the mentioned accuracies of the ensemble classifiers increase to 96.41% and 96.93% for KSTAR and IBk respectively.

While Table 4 is examined for the accuracies of the six base classifiers and their corresponding ensembles, it is easy to see the performance increase for all six classifiers for diabetes datasets. The accuracies of five classifiers from six algorithms are increased for the breast cancer dataset.

One of the remarkable results of this study is the merit of Rotation Forest in creating ensemble of base classifiers with the possible improvement in classification performance. Thus, it is rational to use Rotation Forest ensemble classifiers for numerous classification schemes to design enhanced computer-aided diagnosis systems

References

1. Wahed, M., and Wahba, K., Data mining based-assistant tools for physicians to diagnose diseases, Micro-nano mechatronics and human science, 2003 IEEE Int. Symp. 388–391, 2003.
2. Ozcift, A., and Gulen, A., Assessing effects of preprocessing mass spectrometry data on classification performance. *Eur. J. Mass Spectrom.* 267–273, 2008.
3. Kononenko, I., Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.* 23(1):89–109, 2001.
4. Djebbari, A., An ensemble machine learning approach to predict survival in breast cancer. *Int. J. Comput. Biol. Drug Des.* 1 (3):275–294, 2008.
5. Das, R., and Sengur, A., Evaluation of ensemble methods for diagnosing of valvular heart disease. *Expert Syst. Appl.* 2010.
6. Duangsoithong, R., and Windeatt, T., Relevant and redundant feature analysis with ensemble classification. Seventh International Conference on Advances in Pattern Recognition. 247–250, 2009.
7. Shadabi, F., Sharma, D., and Cox, R., Learning from ensembles: Using artificial neural network ensemble for medical outcomes prediction. *Innovations in Information Technology, IEEE.* 1–5, 2006.
8. Bruha, I., Meta-learner for unknown attribute values processing: Dealing with inconsistency of meta-databases. *J.I.I.S.* 71–87, 2004.
9. Polikar, R., An ensemble based data fusion approach for early diagnosis of Alzheimer's disease. *Inf. Fusion.* 83–95, 2008.
10. Das, R., A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst. Appl.* 1568–1572, 2010.
11. Little, M., and McSharry, P., Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nature Precedings.* 1–27, 2008.
12. Chaudhuri, K., and Healy, D. G., Non-motor symptoms of Parkinson's disease: Diagnosis and management. *Lancet Neurol.* 235–245, 2006.
13. Rosen, K., and Kent R. D., Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. *J. Speech Lang. Hear. Res.* 395–411, 2006.
14. Little, M., and McSharry, P., Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* 1–8, 2008.
15. Lee, M., and Boroczky, L., A two-step approach for feature selection and classifier ensemble construction in computer-aided diagnosis. *Computer-Based Med. Syst.* 548–553, 2008.
16. Martinez, A. M., and Zhu, M., Where are linear feature extraction methods applicable? *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp. 1934–1944, 2005.
17. Fodor, I. K., A survey of dimension reduction techniques, Department of Energy by the University of California. 1–7, 2002.
18. Saeys, Y., A review of feature selection techniques in bioinformatics. *Bioinformatics. Review.* 2507–2517, 2007.
19. Jong, K., Feature selection in proteomic pattern data with support vector machines, computational intelligence in bioinformatics and computational biology, Proceedings of the 2004 IEEE Symposium. 41–48, 2004.
20. Chang, Y., Feature ranking using linear SVM, JMLR: workshop and conference proceedings. 53–64, 2008.
21. Guyon, I., Gene selection for cancer classification using support vector machines. *Mach. Learn.* 389–422, 2002.
22. Kuncheva, L. I., Combining pattern classifiers: Methods and algorithms. Wiley. 251.267, 2004.
23. Polikar, R., Ensemble based system in decision making. *IEEE Circuits Syst. Mag.* 21–44, 2006.
24. Zhang, C. and Zhang, J. S., RotBoost: a technique for combining rotation forest and adaboost. *Pattern Recogn. Lett.* 1524–1536, 2008.
25. Rodriguez, J., and Kuncheva L., Rotation forest: a new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* 1619–1630, 2006.
26. Kuncheva, L., and Rodriguez, J., An experimental study on rotation forest ensembles. *Lect. Notes Comput. Sci.* 459–468, 2007.
27. Beale, R., and Jackson, T., *Neural computing: an introduction.* Institute of Physics Publishing, 1–22, 1990.
28. Aha, D. W., and Kibler, D., Instance-based learning algorithms. *Mach. Learn.* 37–66, 1991.
29. Alpaydin, E., *Introduction to machine learning.* MIT Press, 173–197, 2004.
30. Witten, I. H. and Ian, H., Data mining : practical machine learning tools and techniques. *Morgan Kaufmann Ser. Data Manage. Syst.* 153–168, 2005.
31. Huang, J., and Ling, C., Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* 299–310, 2005.
32. David, A., Comparison of classification accuracy using Cohen's weighted kappa. *Expert Syst. Appl.* 825–832, 2008.
33. Mangasarian, O., and Wolberg, W., Cancer diagnosis via linear programming. *SIAM News.* 1–18, 1990.
34. Kahn, M., Automated interpretation of diabetes patient data: Detecting temporal changes in insulin therapy. *Proc Symp. Comp. Appl. Med. Care*, IEEE Computer Society Press, 569–573, 1990.