Expert Systems with Applications 40 (2013) 263–271

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An efficient diagnosis system for detection of Parkinson's disease using fuzzy *k*-nearest neighbor approach

Hui-Ling Chen [a,*], Chang-Cheng Huang [a], Xin-Gang Yu [b], Xin Xu [c], Xin Sun [d], Gang Wang [d], Su-Jing Wang [d]

[a] College of Physics and Electronic Information, Wenzhou University, Wenzhou, Zhejiang 325035, China
[b] Guangzhou Military Region, Guang Zhou, Guang Dong 510000, China
[c] China Classification Society, Digital Easy Technology Development Co. LTD., Beijing 100007, China
[d] College of Computer Science and Technology, Jilin University, Changchun 130012, China

ARTICLE INFO

Keywords:
Fuzzy *k*-nearest neighbor method
Support vector machine
Feature extraction
Medical diagnosis
Parkinson's disease

ABSTRACT

In this paper, we present an effective and efficient diagnosis system using fuzzy *k*-nearest neighbor (FKNN) for Parkinson's disease (PD) diagnosis. The proposed FKNN-based system is compared with the support vector machines (SVM) based approaches. In order to further improve the diagnosis accuracy for detection of PD, the principle component analysis was employed to construct the most discriminative new feature sets on which the optimal FKNN model was constructed. The effectiveness of the proposed system has been rigorously estimated on a PD data set in terms of classification accuracy, sensitivity, specificity and the area under the receiver operating characteristic (ROC) curve (AUC). Experimental results have demonstrated that the FKNN-based system greatly outperforms SVM-based approaches and other methods in the literature. The best classification accuracy (96.07%) obtained by the FKNN-based system using a 10-fold cross validation method can ensure a reliable diagnostic model for detection of PD. Promisingly, the proposed system might serve as a new candidate of powerful tools for diagnosing PD with excellent performance.

1. Introduction

Parkinson's disease (PD) is one kind of degenerative diseases of the nervous system, which is characterized by a group of conditions called motor system disorders because of the loss of dopamine-producing brain cells. Primary symptoms of PD include tremor, or trembling in hands, arms, legs, jaw, and face; rigidity, or stiffness of the limbs and trunk; bradykinesia, or slowness of movement; and postural instability, or impaired balance and coordination. PD usually affects people over the age of 50, which has influenced a large part of worldwide population up to now (http://www.ninds.nih.gov/disorders/parkinsons_disease/parkinsons_disease.htm, last accessed: April 2012). Till now, the cause of PD is still unknown, however, it is possible to alleviate symptoms significantly at the onset of the illness in the early stage (Singh, Pillay, & Choonara, 2007). It is claimed that approximately 90% of the patients with PD show vocal impairment (Ho, Iansek, Marigliani, Bradshaw, & Gates, 1998), the patients with PD typically exhibit a group of vocal impairment symptoms, which is known as dysphonia. The dysphonic indicators of PD make speech measurements an important part of diagnosis. Recently, dysphonic measures have been proposed as a reliable tool to detect and monitor PD (Little, McSharry, Hunter, Spielman, & Ramig, 2009; Rahn, Chou, Jiang, & Zhang, 2007).

Previous study on the PD problem has been undertaken by various researchers. Little et al. (2009) conducted a remarkable study about PD identification, they employed an Support Vector Machine (SVM) classifier with Gaussian radial basis kernel functions to predict PD, and also performed feature selection to select the optimal subset of features from the whole feature space, and the best accuracy rate of 91.4% was obtained by the best model. Shahbaba and Neal (2009) introduced a new nonlinear model based on Dirichlet process mixtures for classification of PD, the results have been compared with that of multinomial logit models, decision trees, and SVM, the best classification accuracy of 87.7% was obtained by the proposed approach. Das (2010) presented a comparative study of using Neural Networks (ANN), DMneural, Regression and Decision Tree for effective diagnosis of PD, the experimental results have shown that the ANN classifier yielded the best results,

* Corresponding author.
E-mail address: chenhuiling.jlu@gmail.com (H.-L. Chen).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.07.014

the overall classification score of 92.9% was achieved. Sakar and Kursun (2010) used the mutual information based feature selection methods integrated with the SVM classifier for PD diagnosis, and the classification accuracy of 92.75% was achieved. Psorakis, Damoulas, and Girolami (2010) introduced novel convergence measures, sample selection strategies and model improvements for multiclass multi-kernel relevance vector machines (mRVMs), and finally, the improved mRVMs achieved the classification accuracy rate of 89.47% when applied to prediction of PD. Guo, Bhattacharya, and Kharma (2010) combined genetic programming and the expectation maximization algorithm (GP-EM) to detect PD, and the best classification accuracy of 93.1% was obtained. Recently, Luukka (2011) employed the feature selection method based on fuzzy entropy measures together with the similarity classifier to predict PD, and mean classification accuracy of 85.03% with only two features was obtained. Li, Liu, and Hu (2011) proposed a fuzzy-based non-linear transformation method in combination with the SVM classifier for prediction of PD, and the best classification accuracy of 93.47% was achieved. Ozcift and Gulten (2011) combined the correlation based feature selection (CFS) algorithm with the rotation forest (RF) ensemble classifiers of 30 machine learning algorithms to identify PD, and the best classification accuracy of 87.13% was achieved by the proposed CFS-RF model. AStröm and Koker (2011) proposed a parallel feed-forward neural network structure for prediction of PD, the highest classification accuracy of 91.20% was obtained. Spadoto et al. (2011) applied evolutionary-based techniques in combination with the Optimum-Path Forest (OPF) classifier to detect PD, and the best classification accuracy of 84.01% was achieved.

From these works, we can see that most of the common classifiers from machine learning community have been utilized for diagnosis of PD. It is evident that the choice of an excellent classifier is of significant importance to the PD diagnosis problem. In this study, an attempt is made to investigate the fuzzy *k*-nearest neighbor (FKNN) classifier in constructing an automatic diagnostic system for diagnosis of PD. Compared with ANN and SVM, FKNN as an improvement over the standard KNN classifier is much simpler and more easily interpretable while maintaining the acceptable classification accuracy. The main idea behind FKNN (Keller, Gray, & Givens, 1985) is that it uses concepts from fuzzy logic to assign degree of membership to different classes while considering the distance of its *k*-nearest neighbors. Points closer to the query point contributes larger value to be assigned to the membership function of their corresponding class in comparison to far away neighbors. Class with the highest membership function value is taken as the winner. One unique characteristic of FKNN method is that it can assign a confidence degree for each predicted class. Thanks to its good properties, it has found its application in a wide range of classification tasks such as protein subcellular locations prediction (Huang & Li, 2004), protein solvent accessibility prediction (Sim, Kim, & Lee, 2005), hyperspectral satellite image classification (Yu, De Backer, & Scheunders, 2002), manufacturing applications (Warren Liao & Li, 1997), bankruptcy prediction (Chen et al., 2011a, 2011b), medical diagnosis (Liu et al., 2011; Seker, Odetayo, Petrovic, & Naguib, 2003) and so on. To the best of our knowledge, FKNN has not been examined for PD diagnosis although it has been used frequently for the classification of biological and medical data. Aiming at improving the efficiency and effectiveness of the classification accuracy for PD diagnosis, in this study, a diagnosis system based on FKNN classifier is introduced. The rationale underlying the proposed system is firstly to use principle component analysis (PCA) to eliminate the redundant information in the original PD data, then to train an optimal FKNN model whose parameters are identified by the cross validation (CV) analysis on the reduced feature space. Finally, the optimal model is utilized to perform the PD diagnostic tasks. The effectiveness of the proposed system is examined in terms of the classification accuracy, sensitivity, specificity and AUC on the PD data set taken from UCI machine learning repository. Promisingly, as can be seen that the developed diagnosis system for this data set in which a more reliable result is found (96.07% mean accuracy) by 10-fold CV method.

The remainder of this paper is organized as follows. Section 2 offers brief background knowledge on FKNN. The detail of implementations of the FKNN-based diagnosis system is described in Section 3. In the next section, the detailed experimental design is presented, and Section 5 describes all the empirical results and discussion. Finally, Conclusions and future work are summarized in Section 6.

2. Fuzzy *k*-nearest neighbor method

The *k*-nearest neighbor (KNN) is one of the oldest and simplest non-parametric pattern classification methods (Cover & Hart, 1967), in which a class is assigned according to the most common class amongst its *k*-nearest neighbors. As an improved version of the KNN method, FKNN (Keller et al., 1985) incorporates the fuzzy set theory into KNN. In FKNN, rather than individual classes as in KNN, the fuzzy memberships of samples are assigned to different categories according to the following formulation:

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij}\left(1/\|x - x_j\|^{2/(m-1)}\right)}{\sum_{j=1}^{k}\left(1/\|x - x_j\|^{2/(m-1)}\right)}, \tag{1}$$

where $i = 1,2,\ldots,C$, and $j = 1,2,\ldots,k$, with C number of classes and k number of nearest neighbors. The fuzzy strength parameter m is used to determine how heavily the distance is weighted when calculating each neighbor's contribution to the membership value, and its value is usually chosen as $m \in (1,\infty)$. $\|x - x_j\|$ is the distance between x and its jth nearest neighbor x_j, usually Euclidean distance is chose as the distance metric. u_{ij} is the membership degree of the pattern x_j from the training set to the class i, among the k nearest neighbors of x. There are two ways to define u_{ij}, one way is the crisp membership, i.e., each training pattern has complete membership in their known class and non-memberships in all other classes. The other way is the constrained fuzzy membership, i.e., the k nearest neighbors of each training pattern (say x_k) are found, and the membership of x_k in each class is assigned as:

$$u_{ij}(x_k) = \begin{cases} 0.51 + (n_j/K)^{*}0.49, & \text{if } j = i, \\ (n_j/K)^{*}0.49, & \text{if } j \neq i. \end{cases} \tag{2}$$

The value n_j is the number of neighbors found which belong to the jth class. Note that, the memberships calculated by Eq. (2) should satisfy the following equations:

$$\begin{aligned} &\sum_{l=1}^{C} \mu_{ij} = 1, \quad j = 1,2,\ldots,n, \\ &0 < \sum_{j=1}^{n} u_{ij} < n, \\ &u_{ij} \in [0,1]. \end{aligned} \tag{3}$$

In our preliminary experiments, we have found that the second way leads to better classification accuracy. After calculating all the memberships for a query sample, it is assigned to the class with which it has the highest membership value, i.e.,

$$C(x) = \arg\max_{i=1}^{C}(u_i(x)) \tag{4}$$

The details of the FKNN algorithm are presented in Algorithm 1.

Algorithm 1. The FKNN algorithm.

Input: The training set X with the labeled patterns $\{x_i | i = 1,2,\ldots,n\}$ and the test pattern y.
Output: Class label of y and confidence for each class label.
Step 1: **For** $i = 1,2,\ldots,$ to n
Step 2: Compute the distance from x_i to y using the Euclidean distance;
Step 3: **If** $i \leqslant k$
Step 4: Include x_i in the set of k nearest neighbors;
Step 5: **Else If** (x_i is closer to y than any previous nearest neighbors)
Step 6: Delete the farthest of the k nearest neighbors;
Step 7: Include x_i in the set of k nearest neighbors;
Step 8: **End If**
Step 9: **End for**
Step 10: **For** $c = 1$ to C
Step 11: Compute $u_i(x)$ using Eq. (1);
Step 12: **End For**
Step 13: Crisp class label of y is assigned to the class with which it has the highest membership value using Eq. (4).

3. The proposed diagnosis system

In this section, we describe the proposed FKNN-based diagnosis system. The proposed approach is comprised of two stages as shown in Fig. 1. In the first stage, feature reduction is conducted by using PCA to eliminate the redundant features and thus enhance further the classification performance. In the second stage, FKNN model is firstly trained on the training sets via 10-fold CV to get the optimal parameter pair (k,m), and then the obtained optimal FKNN model is used to perform the classification tasks.

3.1. Feature reduction phase using PCA

Many studies have shown that feature extraction plays an important role in classifier modeling, especially for medical applications (Chen et al., 2012; Chen, Liu, Yang, Liu, & Wang, 2011; Chen, Yang, Liu, & Liu, 2011; Ghazavi & Liao, 2008; Li, Ouyang, Chen, & Liu, 2012). Feature extraction mainly performs two tasks: transforming input parameter vector into a feature vector and reducing its dimensionality. PCA (Duda, Hart, & Stork, 2001; Smith, 2002) is one of the most popular feature extraction tools, which

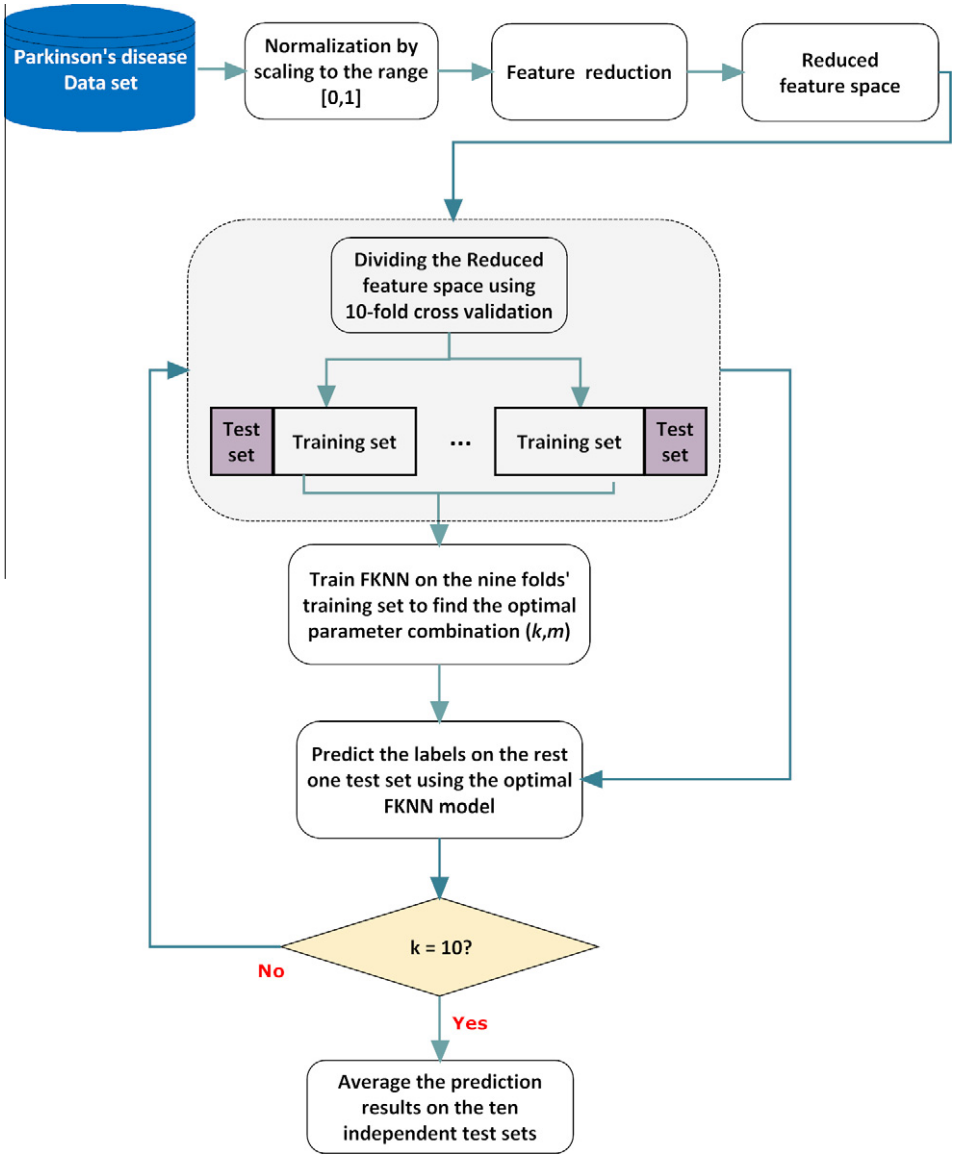

Fig. 1. Overall procedure of the proposed FKNN-based diagnosis system.

seeks to find the largest variations in the original feature space. In all, a well-defined feature extraction algorithm makes the classification process more effective and efficient. In this study, an attempt was made to investigate the effectiveness of PCA for feature reduction on the PD diagnosis problem. The detailed pseudo-code of PCA is given below:

Pseudo-code for the feature reduction procedure

Step 1. Represent the PD data as a matrix O, whose rows represent instances, and columns are features.

Step 2. Subtract the mean from each column of O. The mean is the average across each dimension. This produces a new matrix M whose mean is zero, so that it satisfies the work condition of PCA.

Step 3. Make a single value decomposition of M, namely, $M = U\Sigma V^T$, where U is an s^*s orthonormal matrix containing the left singular vectors of M, Σ is an s^*t rectangular diagonal matrix with nonnegative real numbers on the diagonal, and the t^*t matrix V is the matrix of the right eigenvectors of M.

Step 4. Choose a reduced dimension number L, project M down into the reduced space defined by only the first L singular vectors U_L, then the new matrix $N = (U_L)^T M$ is obtained.

3.2. Classification phase using FKNN

In the second stage, FKNN model performs the classification tasks using the new feature set done by PCA. It includes two main sub procedures. At first, we should set up all the parameters of FKNN model. Since the fuzzy strength parameter has an importance influence on the performance of FKNN. Thus we design an experimental strategy to choose the optimal fuzzy strength parameter for the FKNN classifier. The main idea is that we give a range of [1, 2] with the step of 0.01 for the fuzzy strength parameter m, and then validate the classification performance via the 10-fold CV analysis on several numbers of neighbors k. For each choice of m, we test the average accuracy obtained by FKNN via CV analysis, finally the one with the highest average accuracy is selected as the optimal fuzzy strength parameter. After choosing the optimal fuzzy strength parameter, the FKNN classifier was employed to compute the classification accuracy using the reduced feature set, and then the results were averaged. The detailed pseudo-code for the classification phase is as follows:

Pseudo-code for the classification procedure

/*performance estimation by using k-fold CV where $k = 10$*/
Begin
  For i = 1:Mmax
    For j = 1:k
      Training set = $k - 1$ subsets;
      Test set = remaining subset;
      Train the FKNN model on the training set to find the optimal fuzzy strength parameter m when the neighborhood size k is set to 1, 3, 5 and 7 respectively;
      Test it on the test set and assigns the accuracy to $V(j)$, where V is a vector whose element is the corresponding accuracy obtained by each folder;
    End for
    Compute the mean value of vector V, and store the mean CV accuracy to the vector $M(i)$;
  End for
  Get the optimal m value whose corresponding mean CV accuracy is the highest in $M(i)$;
End
Begin
  For l = 1:k
    Training set = $k - 1$ subsets;
    Test set = remaining subset;
    Train the FKNN model on the training set using the obtained optimal parameter combination;
    Test it on the test set and save the mean CV accuracy;
  End for
  Return the average classification accuracy rates of FKNN over l test set.
End

Table 1
The detail of the 22 attributes of the PD data set.

Label	Attribute	Description
F1	MDVP:Fo(Hz)	Average vocal fundamental frequency
F2	MDVP:Fhi(Hz)	Maximum vocal fundamental frequency
F3	MDVP:Flo(Hz)	Minimum vocal fundamental frequency
F4	MDVP:Jitter(%)	Several measures of variation in fundamental frequency
F5	MDVP:Jitter(Abs)	
F6	MDVP:RAP	
F7	MDVP:PPQ	
F8	Jitter:DDP	
F9	MDVP:Shimmer	Several measures of variation in amplitude
F10	MDVP:Shimmer(dB)	
F11	Shimmer:APQ3	
F12	Shimmer:APQ5	
F13	MDVP:APQ	
F14	Shimmer:DDA	
F15	NHR	Two measures of ratio of noise to tonal components in the voice
F16	HNR	
F17	RPDE	Two nonlinear dynamical complexity measures
F18	D2	
F19	DFA	Signal fractal scaling exponent
F20	Spread1	Three nonlinear measures of fundamental frequency variation
F21	Spread2	
F22	PPE	

3.3. Method for comparison

In order to validate the superiority of the proposed FKNN-based diagnosis system, the prominent SVM classifier was chosen for algorithm comparison. SVM was first introduced by Vapnik (1995), which seeks to minimize the upper bound of the generalization error based on the structural risk minimization principal. The linear SVM finds an optimal separating margin by solving the following optimization task:

$$\begin{aligned} &\text{Min} \quad g(\mathrm{w},\xi) = \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{s.t.,} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geqslant 1 - \xi_i, \quad \xi_i \geqslant 0 \end{aligned} \tag{5}$$

where C is a penalty value, ξ_i is the positive slack variables. This primal problem can be reduced to the Lagarangian dual problem by introducing Lagrangian multipliers α_i. According to the Karush Kuhn–Tucker (KKT) condition, we can get the optimal solution α_i. If $\alpha_i > 0$, the corresponding data points are called SVs. Afterwards, we can get the optimal hyperplane parameters w and b. Then the linear discriminant function can be given by

$$g(\mathbf{x}) = \mathrm{sgn}\left(\sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i^T\mathbf{x} + b\right) \tag{6}$$

In order to make the linear learning machine work well in non-linear cases, the original input space can be mapped into some higher-dimensional feature space via a mapping function. With this mapping, $x^T x$ in the input space can be represented as the form of $\phi(\mathbf{x}_i)^T\phi(\mathbf{x})$ in the feature space. The functional form of the mapping $\phi(\mathbf{x}_i)$ does not need to be known since it is implicitly defined by one selected kernel: $K(\mathbf{x}_i,\mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$. The most widely used kernel in SVM is the Gaussian kernel (or Radial-Basis function, RBF), which is defined as:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i-\mathbf{x}_j\|^2) \tag{7}$$

where γ is the predefined parameter controlling the width of the Gaussian kernel. By introducing the kernel function, the decision function can be expressed as follows:

$$g(\mathbf{x}) = \mathrm{sgn}\left(\sum_{i=1}^{n}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}) + b\right). \tag{8}$$

In Section 5, we will conduct the comparative study between FKNN-based and SVM-based PD diagnosis model, and then the validity of the feature reduction will be investigated accordingly.

4. Experiments design

4.1. Data description

In this study, we have performed our conduction on the PD data set taken from UCI machine learning repository. (http://archive.ics.uci.edu/ml/datasets/Parkinsons, last accessed: May 2012). The purpose of this data set is to discriminate healthy people from those with PD, given the results of various medical tests carried out on a patient. This data set is composed of a range of biomedical voice measurements from 31 people, 23 with PD. The time since diagnoses ranged from 0 to 28 years, and the ages of the subjects ranged from 46 to 85 years, with a mean age of 65.8. Each subject provides an average of six phonations of the vowel (yielding 195 samples in total), each 36 seconds in length (for details consult (Little et al., 2009)). It should be noted that there is no missing values in the data set, and the whole features are real valued. The whole 22 features are presented in Table 1, along with its description.

4.2. Experimental setup

FKNN classifier was implemented from scratch on an Intel Quad-Core Xeon 2.0 GHz CPU using MATLAB language under Windows Server 2003 environment. For SVM, LIBSVM implementation is utilized, which is originally developed by Chang and Lin (2001).

Normalization is employed to avoid feature values in greater numerical ranges dominating those in smaller numerical ranges, as well as to avoid the numerical difficulties during the calculation. In this study, the data are scaled into the interval of [0, 1] according to the Eq. (9), where x is the original value, x' is the scaled value, max_a is the maximum value of feature a, and min_a is the minimum value of feature a.

$$x' = \frac{x - min_a}{max_a - min_a} \tag{9}$$

In order to gain an unbiased estimate of the generalization accuracy, the 10-fold CV was used to evaluate the classification accuracy (Kohavi, 1995). The main advantage of this method is that all of the test sets are independent and the reliability of the results could be improved. The main idea of 10-fold CV procedure is that each time one of the 10 subsets is used as the test set and the remaining 9 subsets are used as a training set. Then the average error across all 10 trials is computed. In order to ensure the same class distribution in the subset, the data is split via stratified sampling in which the sample proportion in each data subset is the same as that in the population. Note that only one repetition of the 10-fold CV will not generate enough classification accuracies for comparison. Because of the arbitrariness of partition of the data set, the predicted accuracy of a model at each iteration is not necessarily the same. To evaluate accurately the performance of the data sets, the 10-fold CV will be repeated 10 independent times and then the results were averaged.

4.3. Measure for performance evaluation

Classification accuracy (ACC), sensitivity, specificity and AUC were used to test the performance of the proposed PCA-FKNN model. ACC, sensitivity and specificity are defined as follows according to the confusion matrix which is shown in Table 2:

$$ACC = \frac{TP+TN}{TP+FP+FN+TN}\times 100\% \tag{10}$$

$$Sensitivity = \frac{TP}{TP+FN}\times 100\% \tag{11}$$

$$Specificity = \frac{TN}{FP+TN}\times 100\% \tag{12}$$

In the confusion matrix, TP is the number of true positives, which means that some cases with 'PD' class is correctly classified as PD; FN, the number of false negatives, which means that some cases with the 'PD' class are classified as healthy persons; TN, the number of true negatives, which means that some cases with the 'Healthy' class are correctly classified as healthy persons; and FP, the number of false positives, which means that some cases with the 'Healthy' class are classified as PD. The receiver operating characteristic (ROC) curve is a graphical display that gives the measure of the predictive accuracy of a logistic model. The curve displays the true positive rate and false positive rate. AUC is the area under the ROC curve, which is one of the best methods for comparing classifiers in two-class problems (Fawcett, 2006), in this study the method proposed in Fawcett (2004) was implemented to compute the AUC.

5. Experimental results and discussions

In order to verify the effectiveness of the proposed model, firstly FKNN was compared with the advanced SVM classifier on the original feature space. For the FKNN classifier, Fig. 2 shows the relationship between the classification accuracy and the fuzzy strength parameter m which varies in the range of [1, 2] with the step size of 0.01 using different numbers of k. It can be observed that the classification accuracy fluctuates between 90% and 98% with different values of m. It reveals that the fuzzy strength parameter has a big impact to the performance of FKNN classifier. The best classification accuracy was achieved with the parameter pair of (1, 1.17), (3, 1.04), (5, 1.06) and (7, 1.02) as shown in Fig. 2(a)–(d) when k is equal to 3, 5 and 7 respectively. These optimal different parameter pairs, namely (1, 1.17), (3, 1.04), (5, 1.06) and (7, 1.02) are used in the subsequent experiments, and for convenience they are named FKNN1, FKNN2, FKNN3 and FKNN4 respectively. Table 3 summarized the detailed results of classification performance in terms of ACC, sensitivity, specificity and AUC obtained by FKNN1, FKNN2, FKNN3 and FKNN4 in the form of average

Table 2
The confusion matrix.

	Predicted positive	Predicted negative
Actual positive	True Positive (TP)	False Negative (FN)
Actual negative	False Positive (FP)	True Negative (TN)

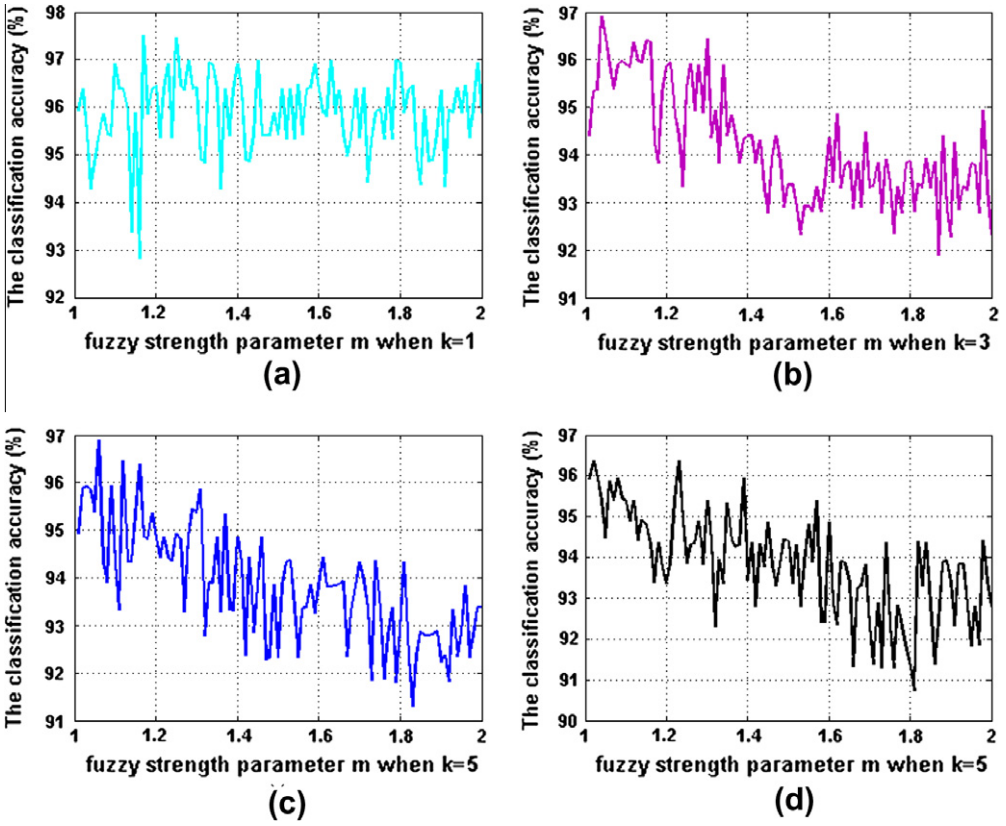

Fig. 2. The relationship between classification accuracy and fuzzy strength parameter m with different numbers of k.

Table 3
Results of classification performance of FKNN with different optimal parameter pairs.

FKNN	Performance metric	Mean	SD	Max	Min
FKNN1	ACC (%)	95.74	0.60	96.45	94.87
k = 1	Sensitivity (%)	96.25	0.45	96.79	95.28
m = 1.17	Specificity (%)	95.07	1.68	97.57	93.00
	AUC (%)	95.66	0.91	96.98	94.48
FKNN2	ACC (%)	95.76	0.71	96.45	94.42
k = 3	Sensitivity (%)	95.53	0.42	95.94	94.59
m = 1.04	Specificity (%)	96.22	2.02	98.33	92.07
	AUC (%)	95.87	1.04	97.12	93.82
FKNN3	ACC (%)	95.61	0.89	96.92	93.92
k = 5	Sensitivity (%)	95.93	0.96	97.45	94.38
m = 1.06	Specificity (%)	94.10	3.20	98.57	88.00
	AUC (%)	95.01	1.86	97.00	91.19
FKNN4	ACC (%)	95.79	0.64	96.92	94.82
k = 7	Sensitivity (%)	95.75	0.41	96.37	95.05
m = 1.02	Specificity (%)	95.45	1.88	98.00	91.67
	AUC (%)	95.60	1.01	97.19	93.81

accuracy (Mean), stand deviation (STD), maximal accuracy (Max) and minimal accuracy (Min). From Table 3, we can see that the results of the classification performance of FKNN classifiers with different optimal parameter pair are very close. Among them, FKNN1 achieved the highest sensitivity of 96.25%, FKNN2 achieved the highest specificity of 96.22% and the highest AUC of 95.87%, FKNN4 obtained the highest classification accuracy of 95.79%.

For the SVM classifier, we have implemented two SVM classifiers, one is the SVM classifier with linear kernel (hereafter SVMliner), and the other is the SVM classifier with RBF kernel (hereafter SVMRBF). For SVMliner, the penalty parameter C, which determines the trade-off between the fitting error minimization and

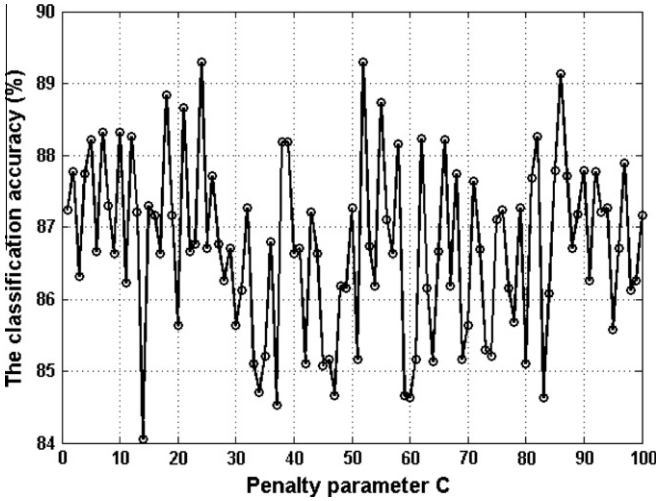

Fig. 3. The relationship between the penalty parameter C and the classification accuracy.

model complexity, should be specified. Here we choose the optimal C via CV method by giving the range of [1, 100] with the step size of 1, the one with the best CV accuracy was chosen as the optimal value of C. The relationship between the penalty parameter C and CV accuracy is shown in Fig. 3, from which we can observe that the best CV accuracy was obtained when C is equal to 24, thus 24 is used the optimal value of C for subsequent analysis. For SVMRBF, in addition to the penalty parameter C should be adjusted, the other kernel parameter γ, defines the non-linear mapping from

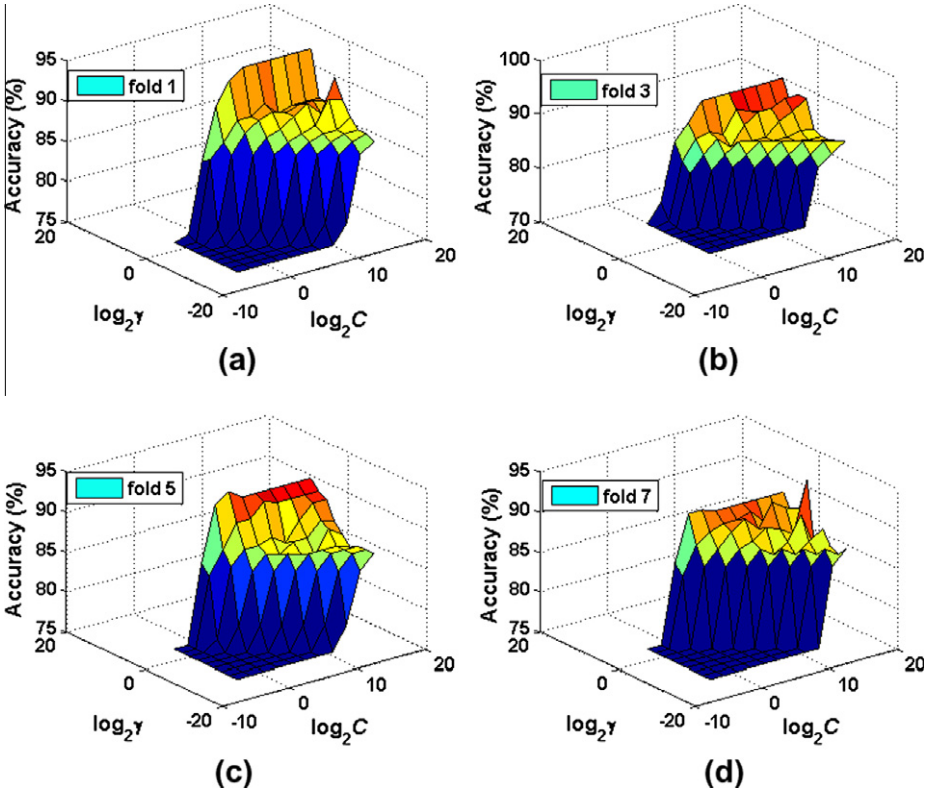

Fig. 4. Test accuracy surface with parameters on PD data set for several folds.

Table 4
Comparison results between the FKNN-based models and SVM-based models on the whole input feature space.

Classifiers	ACC (%)	Sensitivity (%)	Specificity (%)	AUC (%)
FKNN1	95.74 ± 0.60	**96.25 ± 0.45**	95.07 ± 1.68	95.66 ± 0.91
FKNN2	95.76 ± 0.71	95.53 ± 0.42	**96.22 ± 2.02**	**95.87 ± 1.04**
FKNN3	95.61 ± 0.89	95.93 ± 0.96	94.10 ± 3.20	95.01 ± 1.86
FKNN4	**95.79 ± 0.64**	95.75 ± 0.41	95.45 ± 1.88	95.60 ± 1.01
SVMlinear	86.27 ± 0.94	95.31 ± 1.19	57.79 ± 4.53	76.55 ± 2.24
SVMRBF	93.52 ± 0.72	95.88 ± 1.19	86.35 ± 3.98	91.12 ± 1.63

The best values of each performance metric have been shown in bold.

the input space to some high-dimensional feature space, should be determined when performing the classification tasks. In order to automatically tune the two parameter of SVMRBF, the grid search technique (Hsu, Chang, & Lin, 2003) was employed using 5-fold CV to find out the optimal parameter values. The range of the related parameters C and γ are varied between $C = \{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$ and $\gamma = \{2^{-15}, 2^{-13}, \ldots, 2^{1}\}$. $11 \times 9 = 99$ parameter combinations of (C, γ) were tried and the one with the best CV accuracy was chosen as the parameter values of the RBF kernel. Then the best parameter pair (C, γ) was used to create the model for training. Fig. 4 shows the test classification accuracy surface of fold 1, 3, 5 and 7 of 10 folds in one run of 10-fold CV procedure, where the x-axis and y-axis are $\log_2 C$ and $\log_2 \gamma$, respectively. Each mesh node in the (x, y) plane of the test accuracy stands for a parameter combination and the z-axis denotes the obtained test accuracy value with each parameter combination. The detailed results achieved by SVM models together with those of four FKNN models are detailed in Table 4. It can be seen from the table that FKNN-based models are much more superior to SVM-based ones by a large percentage in terms of ACC, sensitivity, specificity and AUC. We also can observe that the performance of SVMRBF was much higher than that of SVMlinear, the reason may lie in that the relation between class labels and attributes in PD data set is nonlinear, thus the linear kernel can't work well for the nonlinear case. In addition, it is interesting to see that the standard deviation for the acquired performance by the FKNN-based models is much smaller than that of SVM-based models in most cases, which indicates consistency and stability of the proposed model.

In order to investigate whether feature reduction can enhance the performance of the FKNN classifier, we further conducted the proposed model on the reduced feature space using PCA approach. For comparison purpose, we also implemented the SVM-based approaches on the same problem. Table 5 lists the results of FKNN-based models and SVM-based models with feature extraction done by PCA in terms of ACC and AUC where the principle components (PCs) range from 1 to 20 with the step size of 5. We can observed that FKNN1 achieves the best performance when PCs is equal to 10, FKNN2, FKNN3 and FKNN4 achieved the best performance when PCs is equal to 20, while SVM-based approaches achieved the best performance when PCs is equal to 15, and we find all these best results are higher than the ones obtained on the original feature space, namely, the whole 22 features. It indicates that there is irrelevant and redundant information in the original PD feature space and different classifiers get the best results on different reduced feature space. It is interesting to find when PCs is equal to 1, SVMlinear has outperforms all the other classifiers in terms of ACC and AUC. The reason may be that when the feature space is reduced to 1, the data distribution is most suitable for the SVM with the linear kernel to find the optimal hyperplane. From the table we can also find that when the features are projected to the new space, FKNN-based approaches can still outperform the SVM-based ones in most cases. Fig. 5 shows the comprehensive results obtained by the six classifiers in terms of ACC, AUC, sensitivity and specificity in one run of 10-fold CV on the reduced feature space where the PCs range from 1 to 22 with the step size of 1. It can be observed that FKNN-based

Table 5
Comparison results between the FKNN-based models and SVM-based models on the reduced input feature space.

No of PCs	Metric	FKNN1	FKNN2	FKNN3	FKNN4	SVM	
						Linear	RBF
1	AUC (%)	61.19 ± 3.38	62.44 ± 2.24	61.62 ± 2.69	61.89 ± 2.59	68.71 ± 1.52	66.53 ± 2.11
	ACC (%)	71.46 ± 1.77	72.47 ± 1.17	72.94 ± 1.57	69.67 ± 0.95	80.96 ± 0.56	80.92 ± 0.74
5	AUC (%)	93.42 ± 1.36	93.60 ± 1.46	93.49 ± 1.57	93.42 ± 1.01	74.73 ± 2.57	87.77 ± 2.21
	ACC (%)	93.55 ± 0.80	93.91 ± 0.86	94.06 ± 0.79	93.71 ± 0.67	85.37 ± 0.43	91.52 ± 1.44
10	AUC (%)	**95.90 ± 0.98**	94.99 ± 1.48	95.36 ± 1.63	95.22 ± 0.97	77.23 ± 1.98	89.71 ± 1.48
	ACC (%)	**96.07 ± 0.60**	95.44 ± 0.89	95.84 ± 0.84	95.76 ± 0.64	86.60 ± 1.20	92.75 ± 0.91
15	AUC (%)	95.15 ± 1.50	96.04 ± 0.96	95.58 ± 1.20	95.11 ± 1.04	**77.60 ± 1.96**	**92.31 ± 1.97**
	ACC (%)	95.45 ± 0.45	95.49 ± 0.83	95.59 ± 0.78	95.39 ± 0.85	**86.99 ± 1.60**	**94.82 ± 1.26**
20	AUC (%)	95.44 ± 1.02	**96.19 ± 1.45**	**95.73 ± 1.09**	**95.87 ± 1.33**	77.43 ± 1.45	90.56 ± 2.03
	ACC (%)	95.73 ± 0.65	**95.95 ± 0.73**	**95.86 ± 0.87**	**95.90 ± 0.73**	86.41 ± 0.83	93.32 ± 1.04

The best results have been shown in bold.

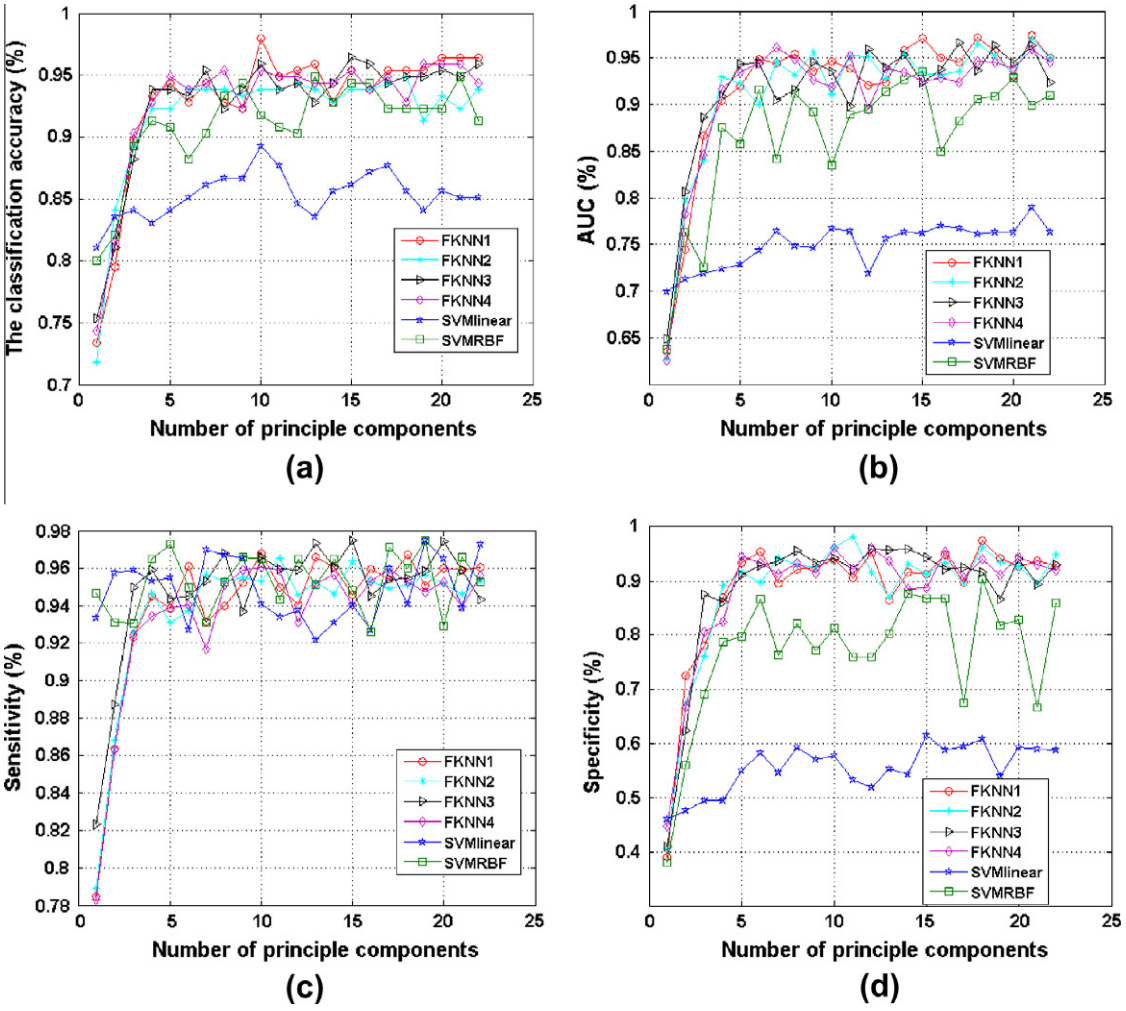

Fig. 5. The relationship between the classification performance and different reduced feature space.

approaches achieve the better results than SVM-based ones in terms of ACC, AUC and specificity on the reduced space in most cases. However, the sensitivity obtained by the SVM-based approaches is very close to the FKNN-based ones. It means that SVM can achieve the same ability to discriminate the patients with PD as that of FKNN.

For comparison purpose, Table 6 also lists the classification accuracies of the previous methods which investigated on the PD diagnosis problem. As shown in Table 6, our developed PCA-FKNN system can obtain better classification accuracy than all available methods proposed in previous studies.

Table 6
Classification accuracies obtained with our method and other methods.

Study	Method	Accuracy (%)
Little et al. (2009)	Pre-selection filter + Exhaustive search + SVM	91.4 (bootstrap with 50 replicates)
Shahbaba and Neal (2009)	Dirichlet process mixtures	87.7 (5-fold CV)
Das (2010)	ANN	92. (hold-out)
Sakar and Kursun (2010)	Mutual information based feature selection + SVM	92.75 (bootstrap with 50 replicates)
Psorakis et al. (2010)	Improved mRVMs	89.47 (10-fold CV)
Guo et al. (2010)	GP-EM	93.1 (10-fold CV)
Ozcift and Gulten (2011)	CFS-RF	87.1 (10-foldCV)
Li et al. (2011)	Fuzzy-based non-linear transformation + SVM	93.47 (hold-out)
Luukka (2011)	Fuzzy entropy measures + similarity classifier	85.03 (hold-out)
Spadoto et al. (2011)	Particle swarm optimization + OPF	73.53 (hold-out)
	Harmony search + OPF	84.01 (hold-out)
	Gravitational search algorithm + OPF	84.01 (hold-out)
AStröm and Koker (2011)	Parallel NN	91.20 (hold-out)
This Study	PCA-FKNN	96.07 (average 10-fold CV)

6. Conclusions and future works

This study introduces a new model for PD diagnosis. The main novelty of this model lies in employing the FKNN classifier together with the feature reduction technique to do the diagnosis tasks for PD. Experimental results demonstrated that the proposed system performed significantly well in distinguishing the patients with PD and healthy ones. Meanwhile, a comparative study was conducted between SVM and FKNN. The experimental results have shown that FKNN approach performs advantageously over the SVM methods in terms of the classification accuracy, sensitivity, specificity and AUC. Additionally, our developed system outperformed the existing methods proposed in the literature. Hence, it can be safely concluded that, the developed diagnosis system can serve as a promising alternative tool in medical decision-making for PD diagnosis. The future investigation will pay much attention to evaluate the proposed system in other medical diagnosis problems.

References

AStröm, F., & Koker, R. (2011). A parallel neural network approach to prediction of Parkinson's Disease. *Expert Systems with Applications, 38*, 12470–12474.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: a library for support vector machines. In *Software available* at http://www.csie.ntu.edu.tw/cjlin/libsvm.

Chen, H. L., Liu, D. Y., Yang, B., Liu, J., & Wang, G. (2011). A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis. *Expert Systems with Applications, 38*, 11796–11803.

Chen, H. L., Yang, B., Liu, J., & Liu, D. Y. (2011). A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert Systems with Applications, 38*, 9014–9022.

Chen, H.-L., Liu, D.-Y., Yang, B., Liu, J., Wang, G., & Wang, S.-J. (2011a). An adaptive fuzzy k -nearest neighbor method based on parallel particle swarm optimization for bankruptcy prediction. In J. Huang, L. Cao & J. Srivastava (Eds.), *Advances in Knowledge Discovery and Data Mining* (vol. 6634, pp. 249–264). Berlin/Heidelberg: Springer.

Chen, H. L., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S. J., et al. (2011b). A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowledge-Based Systems, 24*, 1348–1359.

Chen, H.-L., Yang, B., Wang, G., Wang, S.-J., Liu, J., & Liu, D.-Y. (2012). Support vector machine based diagnostic system for breast cancer using swarm intelligence. *Journal of Medical Systems*, http://dx.doi.org/10.1007/s10916-011-9723-0.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*, 21–27.

Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications, 37*, 1568–1572.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine Learning, 31*, 1–38.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Ghazavi, S. N., & Liao, T. W. (2008). Medical data mining by fuzzy modeling with selected features. *Artificial Intelligence in Medicine, 43*, 195–206.

Guo, P. F., Bhattacharya, P., & Kharma, N. (2010). Advances in detecting Parkinson's disease. *Medical Biometrics*, 306–314.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11*, 131–138.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. In *Technical report*, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003. Available at http://www.csie.ntu.edu.tw/cjlin/libsvm/.

Huang, Y., & Li, Y. (2004). Prediction of protein subcellular locations using fuzzy k-NN method. *Bioinformatics, 20*, 21–28.

Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy k-nearest neighbours algorithm. *IEEE Transacions on Systems, Man, Cybernetics, 15*, 580–585.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In C.S. Mellish (Ed.), *Proceedings IJCAI-95* (Vol. 14, pp. 1137–1145). Montreal, Que., Morgan Kaufmann, Los Altos, CA, 1995.

Li, D. C., Liu, C. W., & Hu, S. C. (2011). A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artificial Intelligence in Medicine, 52*, 45–52.

Li, L. N., Ouyang, J. H., Chen, H. L., & Liu, D. Y. (2012). A computer aided diagnosis system for thyroid disease using extreme learning machine. *Journal of Medical Systems*, http://dx.doi.org/10.1007/s10916-012-9825-3.

Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 56*(4), 1015–1022.

Liu, D. Y., Chen, H. L., Yang, B., Lv, X. E., Li, L. N., & Liu, J. (2011). Design of an enhanced fuzzy *k*-nearest neighbor classifier based computer aided diagnostic system for thyroid disease. *Journal of Medical Systems*, 1–12.

Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications, 38*, 4600–4607.

Ozcift, A., & Gulten, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput Methods Programs Biomed, 104*, 443–451.

Psorakis, I., Damoulas, T., & Girolami, M. A. (2010). Multiclass relevance vector machines: sparsity and accuracy. *Neural Networks, IEEE Transactions on, 21*, 1588–1598.

Rahn, D. A., Chou, M., Jiang, J. J., & Zhang, Y. (2007). Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice, 21*, 64–71.

Sakar, C. O., & Kursun, O. (2010). Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems, 34*, 1–9.

Seker, H., Odetayo, M. O., Petrovic, D., & Naguib, R. N. G. (2003). A fuzzy logic based-method for prognostic decision making in breast and prostate cancers. *IEEE Transactions on Information Technology in Biomedicine, 7*, 114–122.

Shahbaba, B., & Neal, R. (2009). Nonlinear models using Dirichlet process mixtures. *The Journal of Machine Learning Research, 10*, 1829–1850.

Sim, J., Kim, S. Y., & Lee, J. (2005). Prediction of protein solvent accessibility using fuzzy k-nearest neighbor method. *Bioinformatics, 21*, 2844–2849.

Singh, N., Pillay, V., & Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. *Progress in Neurobiology, 81*, 29–44.

Smith, L. I. (2002). A tutorial on principal components analysis. *Cornell University, USA, 51*, 52.

Spadoto, A. A., Guido, R. C., Carnevali, F. L., Pagnin, A. F., Falcao, A. X., & Papa, J. P. (2011). Improving Parkinson's disease identification through evolutionary-based feature selection. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 7857–7860).

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.

Warren Liao, T., & Li, D. (1997). Two manufacturing applications of the fuzzy K-NN algorithm. *Fuzzy Sets and Systems, 92*, 289–303.

Yu, S., De Backer, S., & Scheunders, P. (2002). Genetic feature selection combined with composite fuzzy nearest neighbor classifiers for hyperspectral satellite imagery. *Pattern Recognition Letters, 23*, 183–190.