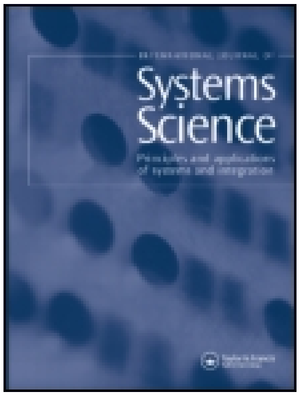

International Journal of Systems Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsys20

Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering

Kemal Polat [a]
[a] Electrical and Electronics Engineering Department, Faculty of Engineering, Bartın University, Bartın 74100, Turkey
Published online: 20 May 2011.

International Journal of Systems Science
Vol. 43, No. 4, April 2012, 597–609

Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering

Kemal Polat*

Electrical and Electronics Engineering Department, Faculty of Engineering, Bartın University, Bartın 74100, Turkey

(*Received 24 September 2010; final version received 8 April 2011*)

This study presents the application of fuzzy c-means (FCM) clustering-based feature weighting (FCMFW) for the detection of Parkinson's disease (PD). In the classification of PD dataset taken from University of California – Irvine machine learning database, practical values of the existing traditional and non-standard measures for distinguishing healthy people from people with PD by detecting dysphonia were applied to the input of FCMFW. The main aims of FCM clustering algorithm are both to transform from a linearly non-separable dataset to a linearly separable one and to increase the distinguishing performance between classes. The weighted PD dataset is presented to *k*-nearest neighbour (*k*-NN) classifier system. In the classification of PD, the various *k*-values in *k*-NN classifier were used and compared with each other. Also, the effects of *k*-values in *k*-NN classifier on the classification of Parkinson disease datasets have been investigated and the best *k*-value found. The experimental results have demonstrated that the combination of the proposed weighting method called FCMFW and *k*-NN classifier has obtained very promising results on the classification of PD.

Keywords: Parkinson's disease; fuzzy C-means (FCM) clustering-based feature weighting; *k*-NN classifier

1. Introducton

In this article, the fuzzy c-means (FCM) clustering-based feature weighting (FCMFW) method has been proposed and applied to the classification of Parkinson disease (PD) dataset. Also, in order to evaluate further, Iris and Wine datasets taken from University of California – Irvine (UCI) machine database [UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases] have been used as the benchmark datasets. FCM clustering algorithm (Bezdek 1981) is the improved version of *k*-means clustering (KCM) (MacQueen 1967). The aim of the proposed feature weighting method is twofold. The first one is to decrease the variance between classes. The second one is to improve the classification accuracy by the transformation from a linearly non-separable dataset to a linearly separable one.

PD is an important disorder, which affects nerve cells in a division of the brain that controls muscle movements. The symptoms of PD are as follows: (1) trembling of hands, arms, legs, jaw and face, (2) stiffness of the arms, legs and trunk, (3) slowness of movement and (4) poor balance and coordination. As symptoms get worse, people with the disease may have trouble walking, talking or doing simple tasks. They may also have problems such as depression, sleep problems or trouble chewing, swallowing or speaking [http://www.nlm.nih.gov/medlineplus/parkinsonsdisease.html]. Parkinson's usually begins around age 60; however, it can start earlier. It is more common in men than in women. There is no cure for PD [http://www.nlm.nih.gov/medlineplus/parkinsonsdisease.html; Elbaz et al. 2002; Little, McSharry, Roberts, Costello, and Moroz 2007; Little, McSharry, Hunter, and Ramig 2009].

There have been some works related to the classification of PD in literature. Little et al. (2009) researched dysphonia measurements for the telemonitoring of PD and achieved the overall correct classification performance of 91.4% using a kernel support vector machine (SVM) on the detection of PD. Little et al. introduced two new tools to speech analysis, recurrence and fractal scaling, which overcome the range limitations of existing tools by addressing, directly, these two symptoms of disorder, together reproducing a ‘hoarseness’ diagram and obtained overall correct classification performance of 91.8% using quadratic discriminant analysis on the classification of PD (Little et al. 2007). Bhattacharya and Bhatia (2010) have distinguished people with PD from the healthy people using SVM classifier. Also, they applied LIBSVM software to find the best possible accuracy on different kernel values for the given dataset (Bhattacharya and Bhatia 2010). Bolat and Sert (2010) used SVM classification to distinguish PD patients to discriminate healthy people from the PD patients with several measurements extracted from the sound samples of 31 people, 23 with PD. Guo, Bhattacharya, and Kharma (2010) have combined

*Email: kemal_polat2003@yahoo.com

ISSN 0020–7721 print/ISSN 1464–5319 online

http://dx.doi.org/10.1080/00207721.2011.581395
http://www.informaworld.com

genetic programming and the expectation maximisation algorithm (GP-EM) to create learning feature functions on the basis of ordinary feature data (features of voice) (Guo et al. 2010). Sakar and Kursun (2010) used the mutual information measure with the permutation test for evaluating the relevance and the statistical significance of the relations between the features and the PD-score, showing whether or not the sample belongs to a person with PD.

Also, there are few works related to fuzzy systems and applications in literature. Among these, Takagi and Sugeno (1985) have proposed a mathematical tool to build a fuzzy model of a system with fuzzy implications and reasoning. Also, they showed the method of identification of a system using its input–output data. Lin (2010) has proposed a classifier of ensemble using fuzzy functions and aggregators (like OWA) and applied it to driving.

As can be seen from the above papers, no weighting methods have been used prior to the classification of PD. In order to fill this void, FCMFW has been proposed and successfully applied to the detection of PD.

In this article, FCM-based feature weighting method has been proposed. The main aims of FCM clustering algorithm are both to transform a linearly non-separable dataset to a linearly separable one and to increase the distinguishing performance between classes. In order to make these processes happen, the clustering centres of features have been found. Then, the ratios of centres of features to the means of features have been calculated and these ratios have been multiplied with each datum in features dimension by dimension. In this way, the variance within intra-classes has been decreased so that the distinction between classes is increased in the classification of PD datasets. The classification method with two stages comprising FCMFW and k-nearest neighbour (*k*-NN) classifier was used in the classification of PD. In order to test the performance and efficiency of the proposed weighting method, the classification accuracy, sensitivity, specificity and *f*-measures were used. In the classification of PD, the various *k*-values in *k*-NN classifier were used and compared with each other. Also, the effects of *k*-values in *k*-NN classifier on the classification of PD datasets have been investigated and the best k-value found.

The remaining of this article is organised as follows. The material is presented in the next section. In Section 3, the method is explained. The experimental data and results are given in Section 4 to present the effectiveness of proposed method. Finally, the conclusions are given in Section 5 with future directions.

2. Material: PD dataset

This dataset is taken from UCI machine learning database [UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases]. The PD dataset was created by Max little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals (Elbaz et al. 2002; Little et al. 2007, 2009).

The data used in this study are composed of 195 sustained vowel phonations from 31 male and female subjects, of which 23 were diagnosed with PD. The time since diagnoses ranged from 0 to 28 years, and the ages of the subjects ranged from 46 to 85 years (mean 65.8, standard deviation 9.8). Averages of six phonations were recorded from each subject, ranging from 1 to 36 s in length. There are 195 instances comprising 48 normal and 147 PD cases in the dataset. Table 1 presents the statistical values of features of PD dataset (Elbaz et al. 2002; Little et al. 2007, 2009).

3. Method

3.1. *The overall structure*

In this article, a novel feature weighting method called FCMFW is proposed and combined with *k*-NN classifier to classify the PD. Figure 1 shows the block diagram of the proposed method.

3.2. *FCMFW: pre-processing stage*

FCM clustering is the improved version of KCM algorithm. KMC algorithm is one of the simplest unsupervised learning algorithms used to solve the clustering problems (MacQueen 1967). The procedure of KMC algorithm pursues a simple way to classify a given dataset through a certain number of clusters (assume k clusters).

The KCM algorithm is composed of four steps as follows (MacQueen 1967; Kanungo et al. 2002):

- Place K points into the space depicted by the objects that are being clustered. These points show initial group centroids.
- Assign each object to the group, which has the closest centroid.
- When all the objects have been assigned, recalculate the positions of the K centroids.
- Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from

Table 1. The statistical values of features of PD dataset [UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases].

Description	Feature label	Minimum value	Maximum value	Average value	SD
Average vocal fundamental frequecy	MDVP: Fo (Hz)	88.33	260.105	154.22	41.39
Maximum vocal fundamental frequecy	MDVP: Fhi (Hz)	102.14	592.03	197.10	91.491
Minimum vocal fundamental frequency	MDVP: Flo (Hz)	65.476	239.17	116.32	43.521
Several measures of variation in fundamental frequency	MDVP: Jitter (%)	0.00168	0.03316	0.00622	0.0048
	MDVP: Jitter (Abs)	7×10^{-6}	0.00026	4.39×10^{-5}	3.48×10^{-5}
	MDVP: RAP	0.00068	0.02144	0.0033	0.00296
	MDVP: PPQ	0.00092	0.01958	0.0034	0.00275
	Jitter: DDP	0.00204	0.06433	0.0099	0.00890
Several measures of variation in amplitude	MDVP: Shimmer	0.00954	0.11908	0.0297	0.01885
	MDVP: Shimmer (dB)	0.085	1.302	0.2822	0.19487
	Shimmer: APQ 3	0.0045	0.0564	0.0156	0.01015
	Shimmer: APQ 5	0.0057	0.0794	0.0178	0.01202
	MDVP: APQ	0.00719	0.1377	0.0240	0.01694
	Shimmer: DDA	0.01364	0.1694	0.0469	0.03045
Two measures of ratio of noise to tonal components in the voice	NHR	0.00065	0.3148	0.0248	0.04041
	HNR	8.441	33.047	21.885	4.42576
Two non-linear dynamical complexity measures	RPDE	0.2565	0.6851	0.49853	0.10394
	D2	0.57428	0.825	0.7180	0.05533
Signal fractal scaling exponent	DFA	−7.96498	−2.434	−5.684	1.0902
Three non-linear measures of fundamental frequency variation	Spread 1	0.00627	0.4504	0.2265	0.0834
	Spread 2	1.423	3.6711	2.3818	0.3827
	PPE	0.04453	0.5273	0.2065	0.0901

Note: $N = 195$ observations comprising 48 normal and 147 PD.

which the metric to be minimised can be calculated.

Clustering algorithms have been used widely not only to collect similar or dissimilar data, but also as useful meansfor data compression and data reduction. The most used clustering algorithms are KCM (MacQueen 1967; Kanungo et al. 2002), FCM clustering (Bezdek 1981), the mountain clustering (Yager and Filev 1994) and subtractive clustering (Chiu 1994). Among these clustering methods, the FCM clustering is chosen in weighting process since FCM is the extended version of KCM and is mostly used in applications and literature (MacQueen 1967; Kanungo et al. 2002).

The goal of the feature weighting method is to map the features according to their distributions in a dataset and also transformation from a non-linearly separable dataset to a linearly separable one. Feature weighting method aims to decrease the variance within features forming dataset. Thanks to this weighting method, similar data in the same feature are gathered and the discrimination ability of classifier is increased. In order to explain further, an example class distribution has been given in Figure 2. In this example, there are two classes including blue and red classes. As seen from Figure 2, the class distribution is confusing so the discrimination of these classes using classifier alone will be a very challenging task. In order to solve this problem, the FCM-based feature weighting method was used prior to classification stage.

The FCM is briefly explained as follows: FCM is one of the most popular fuzzy clustering algorithms (Höppner, Klawonn, and Kruse 1999). FCM was realised by Bezdek (1981). FCM attempts to find a partition for a set of data points $x_j \in R^d, j = 1, \ldots, N$ while minimising the cost function (Equation (1)) (Xu and Wunsch 2005)

$$J(U, M) = \sum_{i=1}^{c} \sum_{j=1}^{N} (u_{i,j})^m D_{ij} \tag{1}$$

where $U = [u_{i,j}]_{cxN}$ is the fuzzy partition matrix and $u_{i,j} \in [0, 1]$ the membership coefficient of the jth object in the ith cluster (Xu and Wunsch 2005); $M = [m_1, \ldots, m_c]$ the cluster prototype matrix; $m \in [1, \infty]$ the fuzzification parameter; and $D_{ij} = D(x_j, m_i)$ the distance measure between x_j and m_i.

The FCM can be briefly summarised as follows, in which the Euclidean or L_2 norm distance

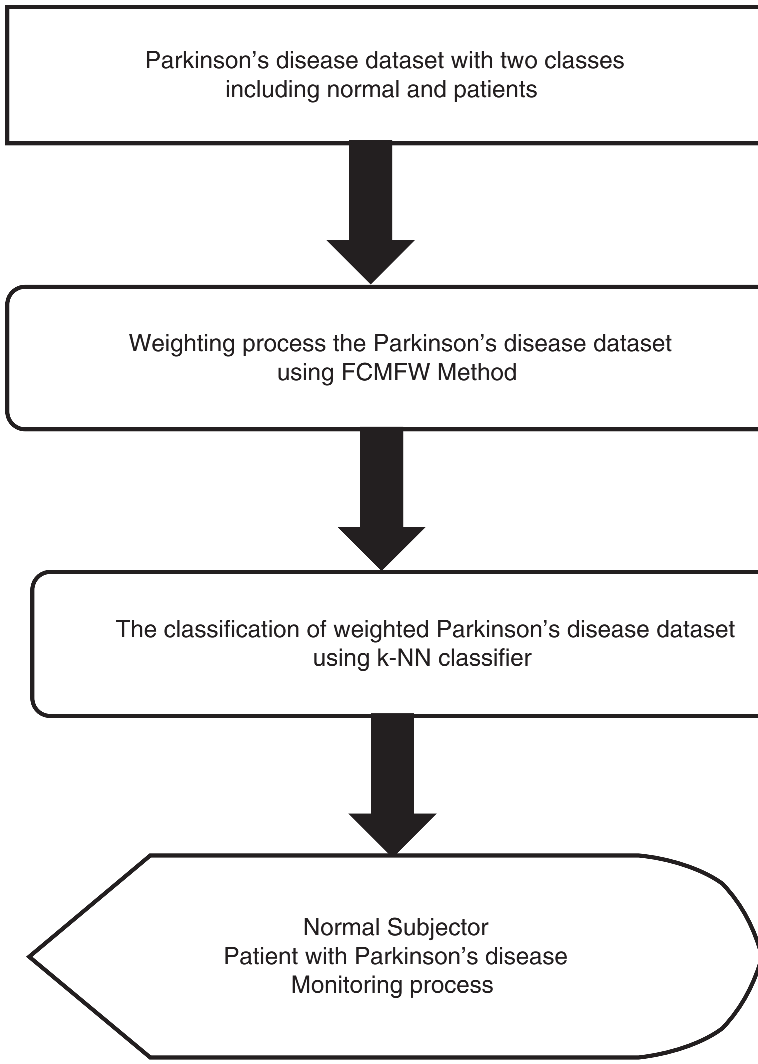

Figure 1. The block diagram of the proposed method.

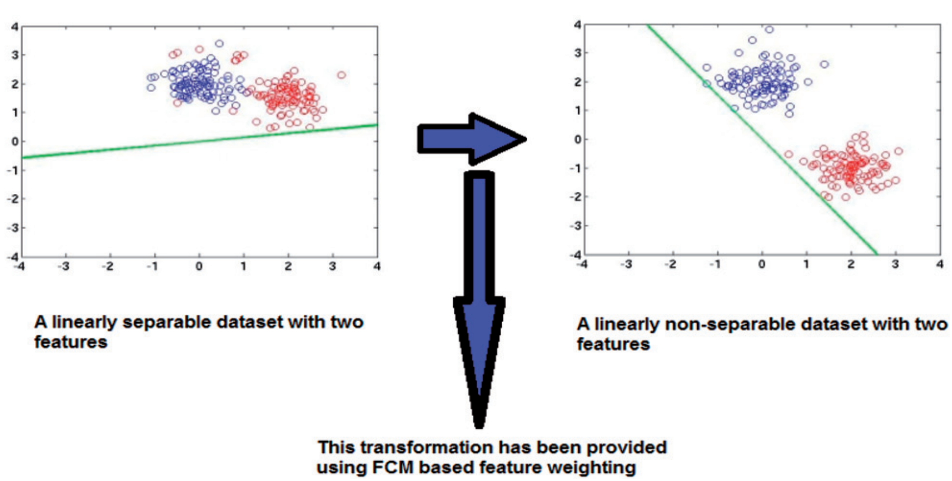

Figure 2. An example class distribution explaining the FCM-based feature weighting.

function is used (Xu and Wunsch 2005; Guldemır and Sengur 2006):

(1) Select the cluster centres $c_i, i = 1, 2, \ldots, c$ randomly from the n points $\{X_1, X_2, X_3, \ldots, X_n\}$.
(2) Compute the membership matrix U using Equation (2)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \tag{2}$$

where $d_{ij} = \|c_i - x_j\|$ is the Euclidean distance between ith cluster centre and jth data point and m the fuzziness index.

(1) Compute the cost function according to the following equation. Stop the process if it is below a certain threshold

$$J(U, c_1, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} d_{ij}^{2} \tag{3}$$

(1) Compute new c fuzzy cluster centres $c_i, i = 1, 2, \ldots, c$ using the following equation

$$c_i = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} X_j}{\sum_{j=1}^{n} \mu_{ij}^{m}} \tag{4}$$

Otherwise go to 2.

This weighting method works as follows: first, the cluster centres of each feature are calculated using FCM method. After computing the centres of features, the ratios of means of features to their cluster centres are calculated and these ratios multiplied with the data of each feature. The pseudo code of the proposed feature weighting method is shown in Figure 3. Also, the flowchart of proposed weighting process is presented in Figure 4. Table 2 presents the clustering centres of features of PD dataset using FCM and KCM algorithms.

3.3. *k-NN classifier: classification stage*

After applying FMCFW to PD dataset, k-NN classifier algorithm has been used. The k-NN algorithm is one of the simplest algorithms among all machine learning algorithms. In k-NN classification, the training dataset is used to classify each data of a 'target' dataset. The structure of the data is that there is a classification variable of interest and a number of additional predictor variables. k-NN algorithm works as follows [http://www.resample.com/xlminer/help/k-NN/knn_intro.htm, http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm (Dasarathy 1991;

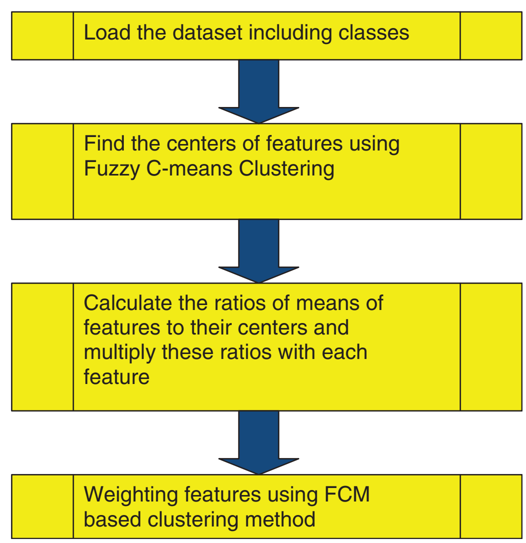

Figure 4. The flowchart of FCMFW method.

Initialize *total values (inputs, outputs, and distances) to zero*

Loading *the A dataset including n features and m output*

For *each feature in dataset*

Step 1. Calculating the cluster centers using fuzzy C-means clustering algorithm

Step 2. Computing the ratios of means of features to their centers and multiply these ratios with each feature

Step 3. Multiplying these ratios with each data in features (Weighting process)

End For

Figure 3. The pseudo code of the proposed feature weighting method.

Table 2. The clustering centres of the features of PD dataset using FCM and KCM algorithms.

Number of feature	Centres of class 0 using FCM (normal)	Centres of class 1 using FCM (PD)	Centres of class 0 using KMC (normal)	Centres of class 1 using KMC (PD)
1	129.35	198.26	196.66	132.014
2	153.65	262.23	280.94	153.221
3	101.90	146.628	144.91	101.36
4	0.006337	0.00617	0.00629	0.00617
5	5.08×10^{-5}	3.34×10^{-5}	3.48×10^{-5}	4.87×10^{-5}
6	0.0033	0.00337	0.00341	0.003247
7	0.0034	0.00349	0.00353	0.0034
8	0.0099	0.01012	0.010254	0.00974
9	0.0305	0.02847	0.02762	0.03079
10	0.285	0.27864	0.27262	0.287
11	0.0161	0.01495	0.01452	0.0162
12	0.0181	0.0175	0.01689	0.0183
13	0.0243	0.02354	0.02262	0.02484
14	0.0484	0.04487	0.04357	0.0487
15	0.0235	0.02811	0.02919	0.02257
16	21.894	21.885	22.0818	21.7834
17	0.5185	0.4653	0.46247	0.5174
18	0.7317	0.6931	0.69132	0.732
19	−5.475	−6.049	−6.0443	−5.49
20	0.231	0.2164	0.21556	0.232
21	2.309	2.485	2.478	2.331
22	0.221	0.1812	0.1814	0.219

Shakhnarovich, Darrell, and Indyk 2005; Latifoglu, Polat, Kara, and Günes 2008)]:

(1) For each row in the target dataset (the set to be classified), locate the k closest members (the k-NN) of the training dataset. A Euclidean distance measure is used to calculate how close each member of the training set is to the target row that is being examined.
(2) Examine the k-NN – which classification do most of them belong to? Assign this category to the row being examined.
(3) Repeat this procedure for the remaining rows in the target set.
(4) The user selects a maximum value for k and builds models parallel on all values of k upto the maximum specified value; scoring is done on the best of these models.

More strong models may be obtained by locating k, where $k > 1$, neighbours and allowing the majority vote decide the outcome of the class labelling. A higher value of k results in a smoother function. The disadvantage of increasing the value of k is that as k approaches n, where n is the size of the instance base, the performance of the classifier will approach that of the most straightforward statistical baseline, the assumption that all unknown instances belong to the class most frequently represented in the training data [http://www.fon.hum.uva.nl/praat/manual/kNN_classifiers_1__What_is_a_kNN_classifier.html].

The various k-values in k-NN classifier algorithm have been used and compared with each other. By this way, the best k-value has been found for classification of PD.

4. The experımental results and dıscussıon

4.1. *Benchmark dataset and performance measures*

In addition to PD dataset, Iris and Wine datasets taken from UCI machine learning database [UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases] have been used as benchmark dataset to further evaluate the proposed method. In order to test the performance of the proposed method, the classification accuracy, sensitivity, specificity and f-measure values have been used.

4.1.1. *Benchmark datasets*

4.1.1.1. *Iris dataset*. Iris dataset is the best known database to be found in the pattern recognition applications. The dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plant. In Iris dataset [UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases]:

number of features, 4
number of instances, 150
number of classes, 3.

4.1.1.2. *Wine dataset*. The data in this dataset are the results of a chemical analysis of wines grown in the same region in Italy, however coming from three-different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. In Wine dataset [UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases]:

number of features, 13
number of instances, 178
number of classes, 3.

4.1.2. *Performance measures*

As the performance measures, the classification accuracy, sensitivity, specificity and *f*-measure values have been used and explained as follows. In training and testing of *k*-NN classifier, 50–50% train–test data split has been used and randomly chosen from dataset. The confusion matrix is shown in Table 3 (actual *vs*. predicted) and the other parameters which are computed using confusion matrix are shown with the following formulas.

For classification accuracy, sensitivity and specificity analysis, and *f*-measure, the following expressions have been used.

$$\text{Classification accuracy (\%)} = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \tag{5}$$

$$\text{Sensitivity (\%)} = \frac{TP}{TP+FN} \times 100 \tag{6}$$

$$\text{Specificity} = \frac{TN}{FP+TN} \times 100 \tag{7}$$

$$\text{Precision} = TP/(TP+FP) \tag{8}$$

Table 3. Confusion matrix.

Confusion matrix	No	Yes
No	TN	FP
Yes	FN	TP

Note: TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively.

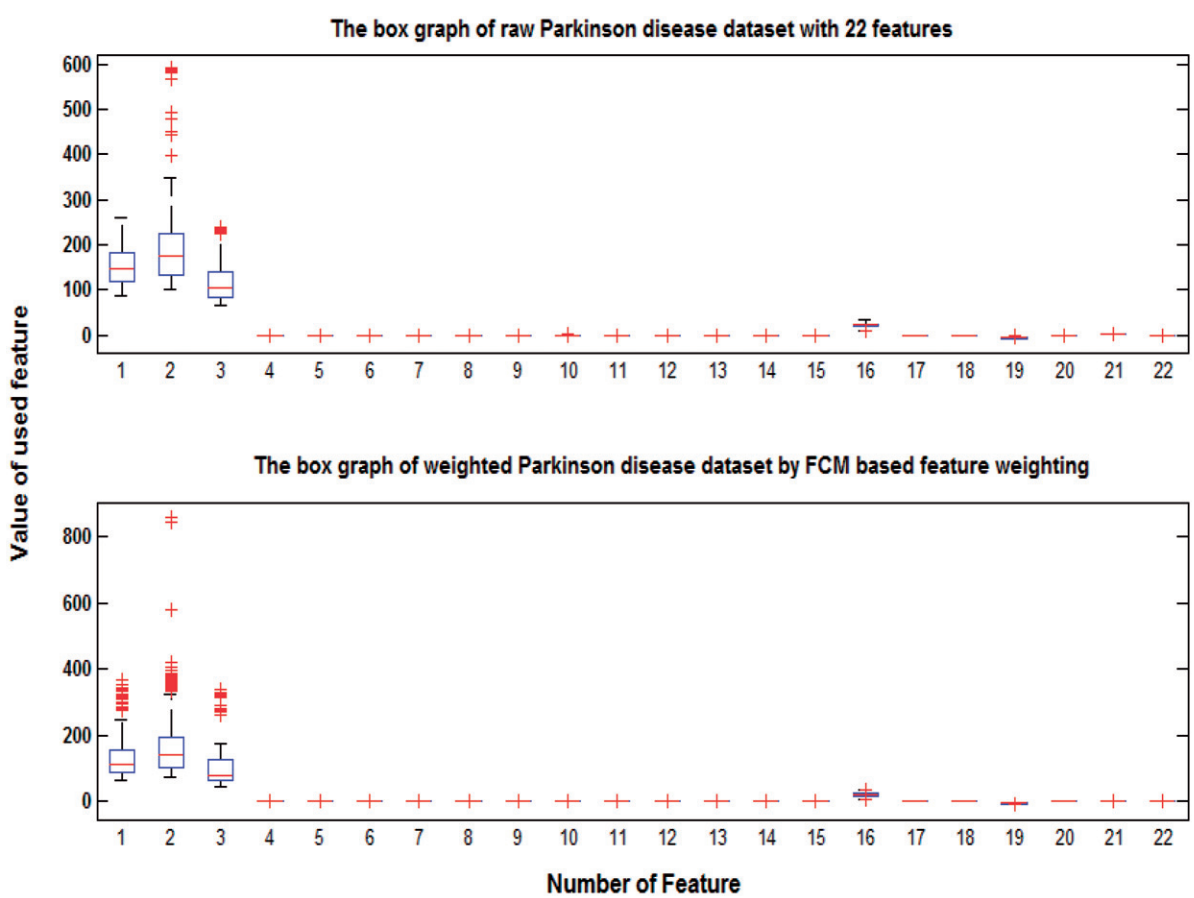

Figure 5. The box graph of raw and weighted PD datasets with 22 features.

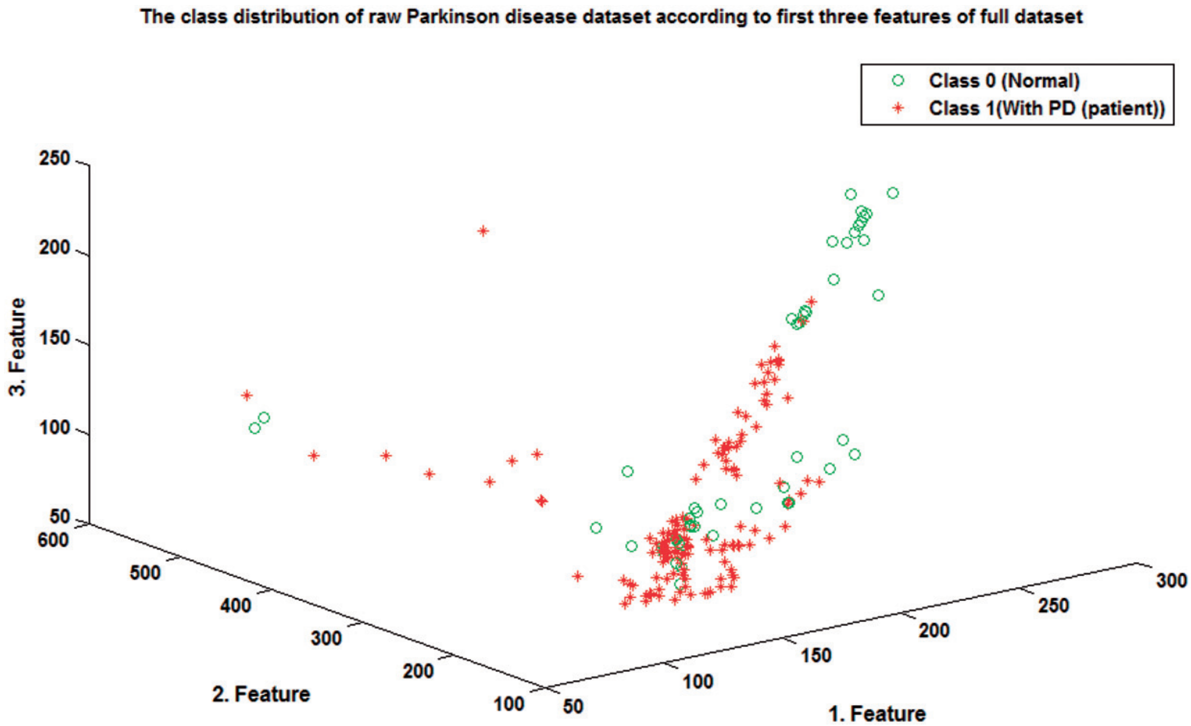

Figure 6. The distribution of PD dataset according to the first three features (first, second and third features).

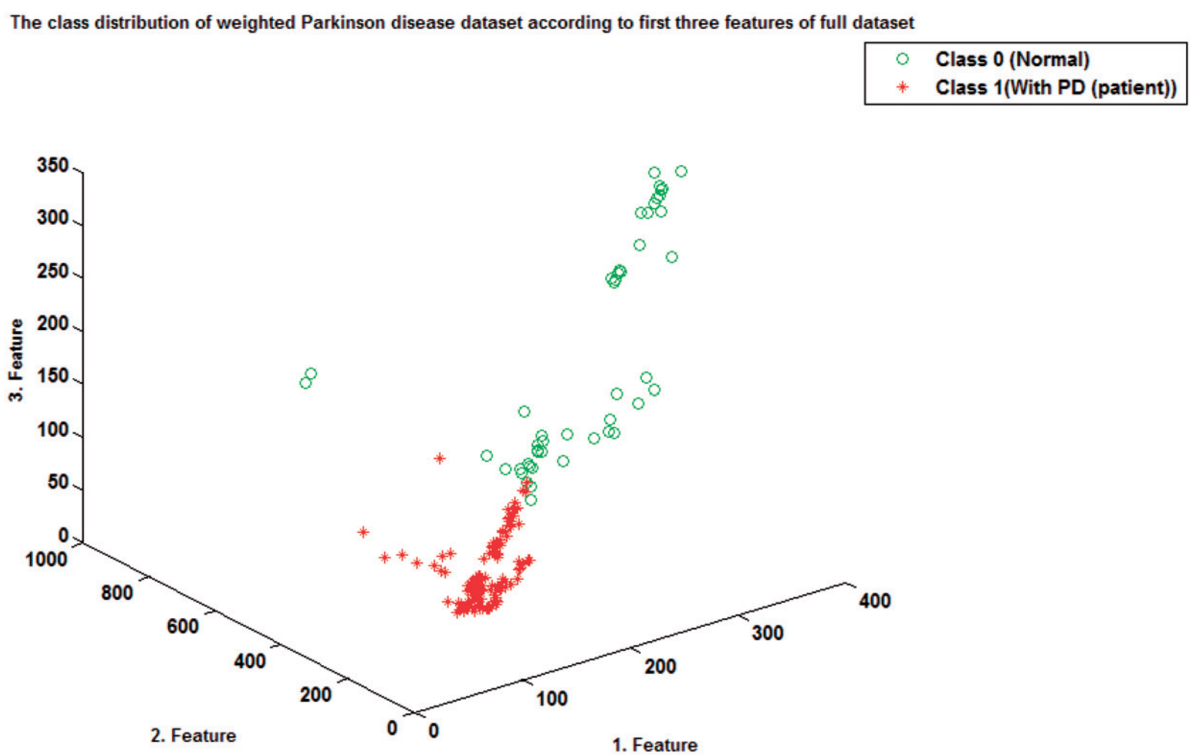

Figure 7. The distribution of weighted PD dataset with FCM-based feature weighting method according to the first three features (first, second and third features).

Table 4. The results obtained from the k-NN classifier and a combination of FCM-based feature weighting method and k-NN classifier using 50–50% training–testing partition on raw and weighted PD datasets.

Classifier method	k-Value in k-NN classifier	Raw PD dataset	Weighted PD dataset with FCM-based feature weighting method
k-NN	1	Classification Accuracy (%): 72.16 Sensitivity (%): 42.10 Specificity (%): 79.48 f-Measure: 0.371	Classification Accuracy (%): 95.87 Sensitivity (%): 100 Specificity (%): 94.80 f-Measure: 0.909
	2	Classification Accuracy (%): 72.16 Sensitivity (%): 43.47 Specificity (%): 81.08 f-Measure: **0.424**	Classification Accuracy (%): 96.90 Sensitivity (%): 95.65 Specificity (%): 97.29 f-Measure: 0.936
	3	Classification Accuracy (%): 72.16 Sensitivity (%): 42.10 Specificity (%): 79.48 f-Measure: 0.371	Classification Accuracy (%): 97.93 Sensitivity (%): 100 Specificity (%): 97.34 f-Measure: **0.956**
	4	Classification Accuracy (%): 69.07 Sensitivity (%): 37.50 Specificity (%): 79.45 f-Measure: 0.375	Classification Accuracy (%): 96.90 Sensitivity (%): 95.65 Specificity (%): 97.29 f-Measure: 0.936
	5	Classification Accuracy (%): 74.22 Sensitivity (%): 47.36 Specificity (%): 80.76 f-Measure: 0.418	Classification Accuracy (%): 96.90 Sensitivity (%): 95.65 Specificity (%): 97.29 f-Measure: 0.936
	6	Classification Accuracy (%): 72.16 Sensitivity (%): 42.85 Specificity (%): 80.26 f-Measure: 0.4	Classification Accuracy (%): 96.90 Sensitivity (%): 95.65 Specificity (%): 97.29 f-Measure: 0.936
	7	Classification Accuracy (%): 73.19 Sensitivity (%): 40 Specificity (%): 77.01 f-Measure: 0.234	Classification Accuracy (%): 95.87 Sensitivity (%): 95.45 Specificity (%): 96.00 f-Measure: 0.912
	8	Classification Accuracy (%): 72.16 Sensitivity (%): 36.36 Specificity (%): 76.74 f-Measure: 0.227	Classification Accuracy (%): 96.90 Sensitivity (%): 95.65 Specificity (%): 97.29 f-Measure: 0.936
	9	Classification Accuracy (%): 73.19 Sensitivity (%): 40 Specificity (%): 77.01 f-Measure: 0.234	Classification Accuracy (%): 95.87 Sensitivity (%): 95.45 Specificity (%): 96.00 f-Measure: 0.912
	10	Classification Accuracy (%): 73.19 Sensitivity (%): 40 Specificity (%): 77.01 f-Measure: 0.234	Classification Accuracy (%): 95.87 Sensitivity (%): 95.45 Specificity (%): 96.00 f-Measure: 0.912

Note: The boldface marks the highest accuracy in classification of raw and weighted PD datasets.

$$\text{Recall} = TP/(TP + FN) \tag{9}$$

$$f\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

4.2. *Results*

In this study, a new feature weighting technique based on FCM clustering in the classification of PD has been proposed. The main aims of this feature weighting method are both to transform from a linearly non-

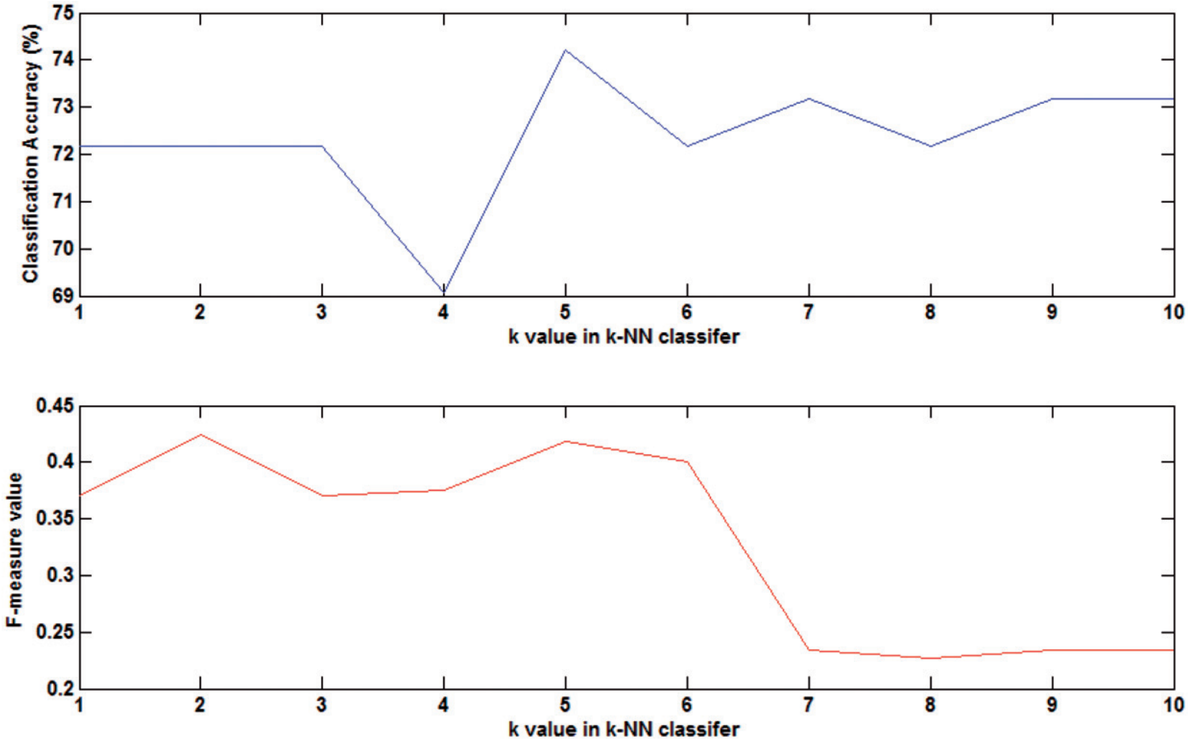

Figure 8. The effects of *k*-value in *k*-NN classifier on the classification accuracy and *f*-measure in the classification of raw PD dataset (without pre-processing).

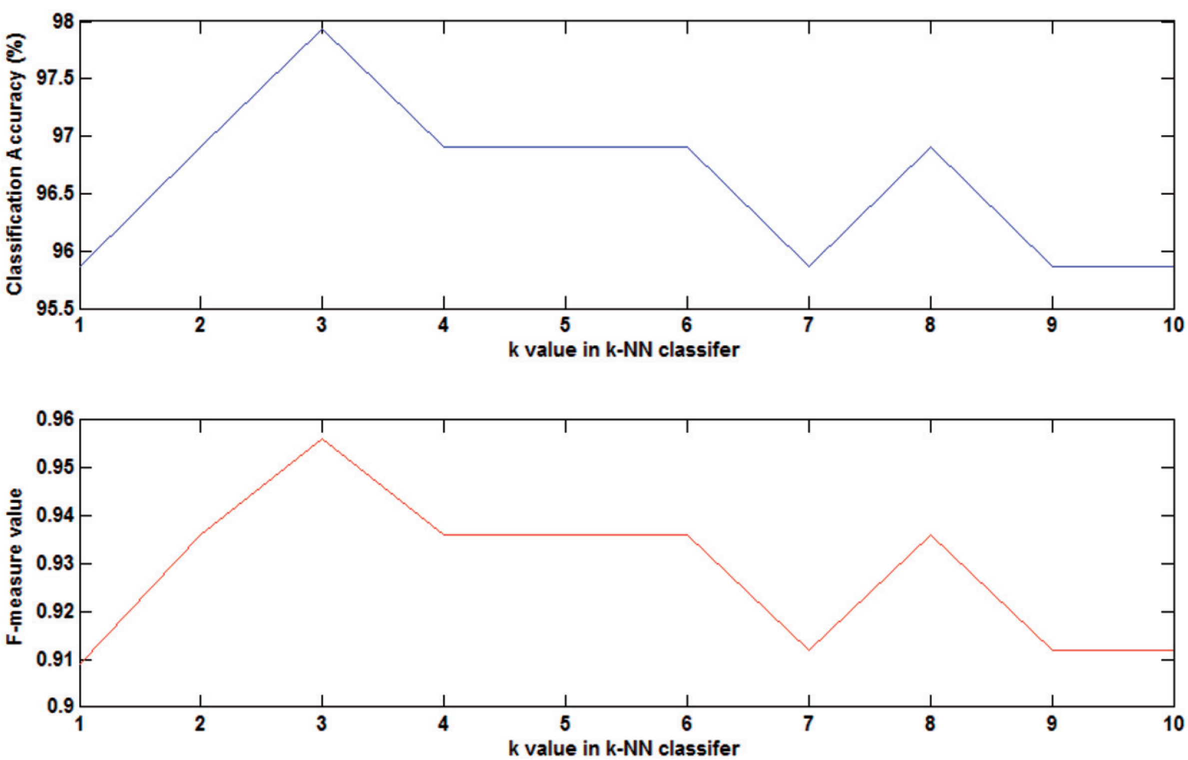

Figure 9. The effects of *k*-value in *k*-NN classifier on the classification accuracy and *f*-measure in the classification of weighted PD dataset by FCM feature weighting.

Table 5. The results obtained in the classification of raw and weighted iris datasets using k-NN classifier.

Used dataset	Raw iris dataset (without pre-processing)	Weighted iris dataset with FCM algorithm
Performance measure	Classification accuracy (%)	Classification accuracy (%)
$k=1$ in k-NN classifier	94.67	100
$k=2$ in k-NN classifier	92.00	100
$k=3$ in k-NN classifier	92.00	100
$k=4$ in k-NN classifier	94.67	100
$k=5$ in k-NN classifier	92.00	100

Table 6. The results obtained in the classification of raw and weighted wine datasets using k-NN classifier.

Used dataset	Raw wine dataset (without pre-processing)	Weighted wine dataset with FCM algorithm
Performance measure	Classification accuracy (%)	Classification accuracy (%)
$k=1$ in k-NN classifier	71.59	86.36
$k=2$ in k-NN classifier	67.05	87.50
$k=3$ in k-NN classifier	67.05	85.23
$k=4$ in k-NN classifier	68.18	86.36
$k=5$ in k-NN classifier	70.45	82.95

separable dataset to a linearly separable one and to increase the distinguishing performance between classes. In order to present the superiority of proposed method, PD dataset has been used. Besides, Iris and Wine datasets have been utilised as benchmark dataset.

The classification system with two stages has been used. In the first stage, to weight the PD dataset, FCM-based feature weighting method has been applied to PD dataset. In this way, PD dataset has been transformed from a linearly non-separable dataset to a linearly separable one. In the second stage, after FCMFW was applied to the dataset, the weighted PD dataset has been classified using k-NN classifier (for various k-values; 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10). In training and testing of k-NN classifier, 50–50% train–test data split has been used and randomly chosen from the dataset.

The working of FCMFW method is given as follows: (1) the centres of features in dataset were found with FCM clustering, (2) the ratios of means of features to their centres were calculated and these ratios were multiplied to each feature in dataset. The raw and weighted PD datasets were given to see the superiority of the proposed weighting method. Figure 5 presents the box graph of raw and weighted PD datasets with 22 features. Figure 6 shows the distribution of PD dataset according to the first three features (first, second and third features). Figure 7 demonstrates the distribution of weighted PD dataset with FCM-based feature weighting method according to the first three features (first, second and third features).

As can be seen from these figures, the distinguishing power of raw PD dataset has been increased by means of FCMFW method. The data distribution of non-linearly separable PD dataset has been transformed to a linear separable one thanks to FCMFW method.

Later, the weighted PD dataset has been applied to k-NN classifier system. In order to test the performance of proposed method, the classification accuracy, sensitivity, specificity and f-measure values have been used. Table 4 gives the results obtained from k-NN classifier alone and a combination of FCM-based feature weighting method and k-NN classifier for 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 values of k. In this table, the superiority of FCM-based feature weighting method could be seen from the results given below. Also, the effects of k-value in k-NN classifier on the classification accuracy in the classification of datasets including PD, Iris and Wine datasets have been investigated. For 3 of k-value in k-NN classifier, the best f-measure has been achieved in the classification of weighted PD dataset. For 2 of k-value in k-NN classifier, the best f-measure has been achieved in the classification of raw PD dataset. Figure 8 shows the effects of k-value in k-NN classifier on the classification accuracy in the classification of raw PD dataset (without pre-processing). Figure 9 presents the effects of k-value in k-NN classifier to classification accuracy in the classification of weighted PD dataset by FCM feature weighting.

In addition to PD dataset, Iris and Wine datasets have been used to evaluate the performance of the proposed method. Table 5 presents the results obtained in the classification of raw iris dataset and weighted iris dataset using k-NN classifier. Table 6 shows the obtained results in the classification of raw wine dataset and weighted wine dataset using k-NN classifier. As can be seen from these results, the proposed FCMFW has achieved promising results.

5. Conclusions

In this article, PD, which is an important disease seen in elderly people, has been classified using the combination of FCMFW and k-NN classifier. The aim of this article is twofold. The first one is to propose the novel feature weighting method called FMCFW. The second one is to increase the classification accuracy of PD by transforming from a linearly non-separable dataset to a linearly separable one. Also, as benchmark dataset, Iris and Wine datasets have been used to test the proposed feature weighting method. In this study, a real application of FCM-based feature weighting method has been carried out. In the end, based on the results obtained, the proposed method could be confidently used in the classification of datasets. In future, PD classification could be carried out online using the improved versions of the proposed method.

Notes on contributor

Dr Kemal Polat was born in Kırıkkale on 30 January 1981. He graduated from the Electrical and Electronics Engineering Department of Selcuk University with BSc in 2002 and from the Electrical and Electronics Engineering Department of Selcuk University with MSc degree in 2004. He took his PhD in Electrical and Electronic Engineering at Selcuk University in 2008. He is now working as assistant professor in Electrical and Electronic Engineering department, Faculty of Engineering, Bartın University since December, 2010. His current research interests are biomedical signal classification, statistical signal processing, digital signal processing, pattern recognition and classification. He has 50 articles published in the SCI journals and 12 international conference papers. He is an reviewer in various SCI-journals including *Information Sciences, Pattern Recognition Letters, Digital Signal Processing, Soft Computing, Artificial Intelligence in Medicine, Applied Soft Computing, Computers in Biology and Medicine, Expert Systems, Computational Statistics and Data Analysis, IEEE Transactions on Biomedical Engineering, IEEE Transactions on Evolutionary Computation, TALANTA, Soft Computing, Journal of Medical Systems and Scientific Research* and Essays, so on. He is a member of the editorial board of *Journal of Neural Computing and Applications* (SCI expanded).

References

Bezdek, J.C. (1981), *Pattern Recognition With Fuzzy Objective Function Algorithms*, New York: Plenum Press.

Bhattacharya, I., and Bhatia, M.P.S. (2010), 'SVM Classification to Distinguish Parkinson Disease Patients', in *A2CWiC '10 Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India*. Krishnashree Achuthan Amrita University, India. DOI: 10.1145/1858378.1858392.

Bolat, B., and Sert, S.N. (2010), 'Classification of Parkinson's Disease by Using Voice Measurements', *International Journal of Reasoning-based Intelligent Systems*, 2, 279–284.

Chiu, S.L. (1994), 'Fuzzy Model Identification Based on Cluster Estimation', *Journal of Intelligent and Fuzzy Systems*, 2, 267–278.

Dasarathy, B.V. (ed.) (1991) *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*, Los Alamitos, CA: IEEE Computer Society Press. ISBN 0-8186-8930-7.

Elbaz, A., Bower, J.H., Maraganore, D.M., McDonnell, S.K., Peterson, B.J., Ahlskog, J.E., Schaid, D.J., and Rocca, W.A. (2002), 'Risk Tables for Parkinsonism and Parkinson's Disease', *Journal of clinical epidemiology*, 55, 25–31.

Guldemır, H., and Sengur, A. (2006), 'Comparison of Clustering Algorithms for Analog Modulation Classification', *Expert Systems with Applications*, 30, 642–649.

Guo, P.F., Bhattacharya, P., and Kharma, N. (2010), 'Advances in Detecting Parkinson's Disease', *Lecture Notes in Computer Science*, 6165/2010, 306–314.

Höppner, F., Klawonn, F., and Kruse, R. (1999), *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis, and Image Recognition*, New York: Wiley.

http://www.fon.hum.uva.nl/praat/manual/kNN_classifiers_1__What_is_a_kNN_classifier.html (accessed 2011).

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., and Wu, A.Y. (2002), 'An Efficient k-Means Clustering Algorithm: Analysis and Implementation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 881–892.

Latifoglu, F., Polat, K., Kara, S., and Günes, S. (2008), 'Medical Diagnosis of Atherosclerosis from Carotid Artery Doppler Signals Using Principal Component Analysis (PCA), k-NN Based Weighting Pre-Processing and Artificial Immune Recognition System (AIRS)', *Journal of Biomedical Informatics*, 41, 15–23.

Lin, Y. (2010), 'Affective Driving', in *Emotional Engineering: Service Development* (Chap. 14, 1st ed.), ed. S. Fukuda, London: Springer (invited chapter), 395 pp.

Little, M.A., McSharry, P.E., Hunter, E.J., and Ramig, L.O. (2009), 'Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease', *IEEE Transactions on Biomedical Engineering*, 56, 1015–1022.

Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A., and Moroz, I.M. (2007), 'Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', *Biomedical Engineering Online*, 6, 23.

MacQueen, J.B. (1967), 'Some Methods for classification and Analysis of Multivariate Observations', in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, pp. 281–297. MR0214227. Zbl 0214.46201.

Sakar, C.O., and Kursun, O. (2010), 'Telediagnosis of Parkinson's Disease Using Measurement of Dysphonia', *Journal of Medical Systems*, 34, 591–599.

Shakhnarovich, G., Darrell, T., and Indyk, P. (2005), *Nearest-neighbor Methods in Learning and Vision*, Cambridge: The MIT Press. ISBN 0-262-19547-X.

Takagi, T., and Sugeno, M. (1985), 'Fuzzy Identification of Systems and its Applications to Modeling and Control', *IEEE Transactions on Systems, Man, and Cybernetics*, 15, 116–132.

Xu, R., and Wunsch II, D. (2005), 'Survey of Clustering Algorithms', *IEEE Transactions on Neural Networks*, 16, 645–678.

Yager, R.R., and Filev, D.P. (1994), 'Generation of Fuzzy Rules by Mountain Clustering', *IEEE Transactions on Systems, Man and Cybernetics*, 24, 209–219.