

## Diagnosis of Parkinson's Disease using Fuzzy C-Means Clustering and Pattern Recognition

Indira Rustempasic, Mehmet Can

International University of Sarajevo, Faculty of Engineering and Natural Sciences  
Hrasnicka Cesta 15, 71000 Sarajevo, Bosnia and Herzegovina  
irustempasic@ius.du.ba, mcan@ius.edu.ba

Parkinson's disease (PD) is a global public health problem of enormous dimension. In this study, we aimed to discriminate between healthy people and people with Parkinson's disease (PD). Various studies revealed, that voice is one of the earliest indicator of PD, and for that reason, Parkinson dataset that contains biomedical voice of human is used. The main goal of this paper is to automatically detect whether the speech/voice of a person is affected by PD. We examined the performance of fuzzy c-means (FCM) clustering and pattern recognition methods on Parkinson's disease dataset. The first method has the main aim to distinguish performance between two classes, when trying to differentiate between normal speaking persons and speakers with PD. This method could greatly be improved by classifying data first and then testing new data using these two patterns. Thus, second method used here is pattern recognition. The experimental results have demonstrated that the combination of the fuzzy c-means method and pattern recognition obtained promising results for the classification of PD.

*Index Terms*—Euclidean distance, classification, fuzzy c-means (FCM) clustering, Parkinson's disease, pattern recognition

### I. INTRODUCTION

Parkinson's disease is a chronic progressive neurological disease that affects a small area of nerve cells called neurons in the area of the brain called the substantia nigra. These cells normally produce dopamine, a chemical (neurotransmitter) that transmits signals between areas in the brain that, when working normally, coordinate smooth and balanced muscle movement. Parkinson's disease causes these nerve cells to die, and as a result, body

movements are affected. The exact cause of this cell-death is still unknown. PD usually affects people over the age of 60, and it is more common in men than in women [16], [17].

The first signs are likely to be barely noticeable, a feeling of weakness or stiffness in one limb, perhaps, or a fine trembling of one hand when it is at rest (activity causes the tremor to disappear). Eventually, the shaking worsens and spreads, muscles tend to stiffen, and balance and coordination deteriorate. Thus, the most obvious symptoms are shaking, rigidity, slowness of movement, difficulty with walking and gait, and

communication. As symptoms get worse, people with the disease may have trouble walking, talking or doing simple tasks. Other symptoms may include depression and other emotional changes; difficulty in swallowing, chewing; urinary problems or constipation; skin problems and sleep disruptions [17].

Parkinson's causes are unknown, but genetics, aging, and toxins are being researched. There are currently no blood or laboratory tests that have been proven to help in diagnosing PD. Particularly in early stages, the disease can be difficult to diagnose accurately. Doctors may sometimes request brain scans or laboratory tests in order to rule out other diseases. Blood tests, brain imaging techniques such as magnetic resonance image (MRI), positron emission tomography (PET scan), and single photon emission computed tomography (SPECT), may be used to help doctors exclude other medical conditions, such as stroke or brain tumors, that produce symptoms similar to those of Parkinson's disease [10]. Amongst others, one of the method for disease diagnosis is detecting and analyzing voice disorders by using acoustic tools that record the changes in pressure at lips or inside the vocal tract. Recently, upon signal processing a group of experts found some features in the voices of the people with Parkinson's disease that can be used as discriminatory measures to differentiate those who have the disease from those who do not. After a Parkinson's diagnosis, Parkinson's disease treatments are given to help relieve symptoms. Although there is presently no cure, there are treatment options such as medication and surgery to manage its symptoms. Since there is no cure now available, early diagnosis is critical for maximizing the effect of treatment and improving the quality of the patient's life.

Scientists are doing a lot of research to look for the answer for causes of the disease. They are studying many possible causes, including aging and poisons in the environment. Abnormal genes seem to lead to Parkinson's disease in some people. But so far, there is not enough proof to show that it is always inherited. Studies in medical biometrics on detecting PD in the early stage are under way and have drawn a lot of attention from the biometrics community in recent years.

Max Little, from the University of Oxford, has been developing software that learns to detect differences in voice patterns, in order to spot distinctive clues associated with Parkinson's. He is using machine learning, collecting a large amount of data when it is known that someone has the disease or not and he train. He introduce a new measure of dysphonia, pitch period entropy (PPE), which is robust to many uncontrollable confounding effects including noisy acoustic environments and normal, healthy variations in voice frequency. He collected sustained phonations from 31 people, 23 with PD, and using a kernel support vector machine (SVM) got overall correct classification performance of 91.4%. He concluded that nonstandard methods in combination with traditional harmonics-to-noise ratios are best able to separate healthy from PD subjects

[2]. Also, the proposal in [1] gave accuracy of around 90% in the diagnosis of PD using Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). Adaptive Neuro-Fuzzy Classifier with linguistic hedges gave recognition results with 95.38% training and 94.72% testing classifying performance [5]. Paper [12] propose algorithm based on the review synthesis that can tackle real-time constraints in pathological voice recognition for the assessment of Parkinson's disease severity. Pause detection, peak to average power rate clipping and zero thresholding rate calculations produce rich voice features in real-time. These features may be further processed using wavelet transforms and used with a neural network for detection and quantification of speech anomalies related to Parkinson's disease. [13] deals with the application of the artificial immune system to discriminate between healthy and people with Parkinson's disease (PWP). Taking inspiration from natural immune systems, they try to grab useful properties such as automatic recognition, memorization and adaptation. Medical data mining also has great potency for exploring the out of sight patterns in the respective medical data sets. Paper [14] provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use today for the classification of Parkinson Disease. The Random Tree Algorithm classifies the Parkinson Disease dataset accurately and provides the 100%. The Linear Discriminant Analysis, C4.5, CS-MC4 and K-NN yields the accuracy results above 90%. K-NN error rate is only 0.0256. Among all, the C-PLS algorithm classifies the dataset with least percentage of 69.74. The C-RT and CS-CRT produce the same error rate of 0.0462. Polat have distinguished people with PD from the healthy people using combination of feature weighting method called FMCFW and k-nearest neighbour (k-NN) classifier [8]. Bhattacharya, and Kharma have combined genetic programming and the expectation maximization algorithm (GP EM) to create learning feature functions on the basis of ordinary feature data (features of voice) [6]. In [10] a full investigation into the features extracted from voice signals of people with and without Parkinson's disease was performed, and three different classifiers were used: Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and some discrimination-function-based (DBF) classifiers. Based on the correct rate, among the mentioned, KNN has shown as the best classifier to differentiate between the people with PD and those without it.

As it is shown, there are many methods that are in use today in medical research and public health for recognizing Parkinson's disease. Classification systems have the potential of being good supportive for the expert. Classification systems can help in increasing accuracy and reliability of diagnosis and minimize the possible errors, as well as making the diagnosis more time efficient.

This paper deals with the application of fuzzy c-means

classification and pattern recognition to a medical dataset concerning PD with the aim of automatically classify patients in PD or non-PD. In order to make these processes happen, the clustering centers of features have been found. Then, the distance of each data to these centers have been calculated, so that the distinction between classes is increased in the classification of PD datasets. After classification, test data are assigned to the one of the two known patterns. In order to test the performance the classification accuracy, sensitivity, specificity, positive and negative predictive values measures were used.

The paper is organized as follows. Section 2 deals with the Parkinson Dataset that is used in this research work. Section 3 handles the proposed classification model to classify the patient as PD or non-PD. Section 4 explains the pattern recognition method. In Section 5 performance parameters are introduced. Section 6 is dealt with the experimental results and Section 7 with discussion of the classifier algorithms. Section 8 concludes research paper.

## II. PARKINSON DATASET

Voice measurement has shown a great progress in the advancement of Parkinson Disease detection. About 90% of people with Parkinson’s disease present some kind of vocal deterioration. And hence, in this paper dataset which mainly focus on the speech signals is chosen. This dataset is taken from UCI machine learning database [15]. The features of dataset are given in Table 1.

TABLE I  
TABLE DESCRIBING THE ATTRIBUTES  
THAT ARE REMOVED AFTER APPLYING THE CORRELATION FILTER

No.	Attribute Name	Rem.	Description
1	MDVP:Fo(Hz)	Yes	Average vocal fundamental frequency
2	MDVP:Fhi(Hz)	Yes	Maximum vocal fundamental frequency
3	MDVP:Flo(Hz)	Yes	Minimum vocal fundamental frequency
4	MDVP:Jitter(%)	Yes	Variation in fundamental frequency
5	MDVP:Jitter(Abs)	No	Variation in fundamental frequency
6	MDVP:RAP	Yes	Variation in fundamental frequency
7	MDVP:PPQ	Yes	Variation in fundamental frequency
8	Jitter:DDP	No	Variation in fundamental frequency
9	MDVP:Shimmer	Yes	Measures of variation in amplitude
10	MDVP:Shimmer(dB)	Yes	Measures of variation in amplitude
11	Shimmer:APQ3	Yes	Measures of variation in amplitude
12	Shimmer:APQ5	Yes	Measures of variation in amplitude
13	MDVP:APQ	No	Measures of variation in amplitude
14	Shimmer:DDA	No	Measures of variation in amplitude
15	NHR	No	Ratio of noise to tonal components
16	HNR	No	Ratio of noise to tonal components
17	status	No	The status of the patient (1)-Parkinson’s Disease, (0)-Healthy
18	RPDE	No	Dynamic complex measurement
19	DFA	No	Signal fractal scaling exponent
20	spread1	Yes	Non-linear measure of fundamental frequency
21	spread2	Yes	Non-linear measure of fundamental frequency
22	D2	No	Dynamic complex measurement
23	PPE	No	Non-linear measure of fundamental frequency

The Parkinson Disease dataset used for classification purpose was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado. This organization recorded the speech signals. The dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). The time since diagnoses ranged from 0 to 28 years, and the ages of the subjects ranged from 46 to 85 years. 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. The main aim of the data is to discriminate healthy people from those with PD. Thus, the dataset is divided into two classes according to its "status" column which is set to 0 for healthy subjects and 1 for those with PD. It is a two-decision classification problem.

Little applied a correlation filter and of these 23 attributes 12 are removed. Each correlation coefficient, which is less than 0.95 is considered not to contribute to classification accuracy, thus the attribute is removed. A total of 11 attributes are kept after the correlation filter has been applied. Table 2 indicates which features are kept. First 10 are used as inputs to the classifiers.

TABLE II FEATURES USED IN THIS PAPER

Index	Attribute Name	Description
1	MDVP:Jitter(Abs)	Variation in fundamental frequency
2	Jitter:DDP	Variation in fundamental frequency
3	MDVP:APQ	Measures of variation in amplitude
4	Shimmer:DDA	Measures of variation in amplitude
5	NHR	Ratio of noise to tonal components
6	HNR	Ratio of noise to tonal components
7	RPDE	Dynamic complex measurement
8	DFA	Signal fractal scaling exponent
9	D2	Dynamic complex measurement
10	PPE	Non-linear measure of fundamental frequency
11	status	The status of the patient (1)-Parkinson’s Disease, (0)-Healthy

## III. CLUSTERING

For fields dealing with diagnosis, we often seek to find structure in the data obtained from observation. Finding the structure in data is the essence of classification. Our experimental observations lead us to develop relationships between the inputs and outputs of an experiment. As we are able to conduct more experiments, we see the relationships forming some recognizable, or classifiable, structure. By finding structure, we are classifying the data according to similar patterns, attributes, features, and other characteristics. The general area is known as *classification*, also termed *clustering*. Clustering is an unsupervised learning task that aims at decomposing a set of object into subgroups or cluster based on

similarity. The goal is to divide the dataset in such a way that objects belonging to the same cluster are as similar as possible, whereas objects belonging to different clusters are as dissimilar as possible. In nonfuzzy or hard clustering, data is divided into crisp clusters, where each data point belongs to exactly one cluster. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to the different clusters.

Clustering can also be thought of as a form of data compression, where a large number of samples are converted into a small number of representative prototypes or clusters. Depending on the data and the application, different types of similarity measures may be used to identify classes, where the similarity measure controls how the clusters are formed. Some examples of values that can be used as similarity measures include distance, connectivity, and intensity.

#### A. Overview of Fuzzy C-Means (FCM)

Fuzzy c-means clustering (FCM) is one of the most popular fuzzy clustering algorithm. Fuzzy c-means (FCM) is a classification or clustering algorithm that has been applied successfully to a number of problems involving feature analysis, clustering and classifier design, such as in agricultural engineering, remote sensing, astronomy, chemistry, geology, image analysis, medical diagnosis and shape analysis. It is an unsupervised clustering algorithm. Clustering refers to identifying the number of subclasses of  $c$  clusters in a data universe  $X$  comprising  $n$  data samples, and partitioning  $X$  into  $c$  clusters ( $2 \leq c < n$ ). Two important issues to consider in this regard are how to measure the similarity between pairs of observations and how to evaluate the partitions once they are formed. One of the simplest similarity measures is distance between pairs of feature vectors in the feature space. If one can determine a suitable distance measure and compute the distance between all pairs of observations, then one may expect that the distance between points in the same cluster will be considerably less than the distance between points in different clusters.

To describe a method to determine the fuzzy c-partition matrix  $U$  for grouping a collection of  $n$  data sets into  $c$  classes, we define an objective function  $J_m$  for a fuzzy c-partition. The FCM algorithm is based on an iterative optimization of an objective function  $J_m$ :

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \quad (1)$$

The algorithm given in equation (1) is a *least squares function*, where the parameter  $n$  is the number of data sets and  $c$  is the number of classes (partitions) into which one is trying to classify the data sets. The squared distance,  $d_{ik}^2$  is then weighted by a measure,  $(u_{ik})^m$ , of the membership of  $x_k$  in the  $i$ th cluster. A new

parameter is introduced in equation (1) called a *weighting parameter*,  $m > 1$ . This parameter controls the amount of fuzziness in the classification process.

The value of  $J_m$  is then a measure of the sum of all the weighted squared errors; this value is then minimized with respect to two constraint functions. First,  $J_m$  is minimized with respect to the squared errors within each cluster, that is, for each specific value of  $c$ . Simultaneously, the distance between cluster centers is maximized.

FCM method uses concepts in  $n$ -dimensional Euclidean space to determine the geometric closeness of data points by assigning them to various clusters or classes and then determining the distance between the clusters:

$$d_{ik} = d(x_k - v_i) = \sqrt{\sum_{j=1}^m (x_{kj} - v_{ij})^2} \quad (2)$$

The distance measure,  $d_{ik}$  is a Euclidean distance between  $v_i$  the  $i$ th cluster center and the  $k$ th data set (data point in  $m$  space). Each of the centers for each class can be calculated as:

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik}^m x_{ki}}{\sum_{k=1}^n u_{ik}^m} \quad (3)$$

where  $j$  is a variable on the feature space, that is,  $j = 1, 2, \dots, m$ . The optimum fuzzy c-partition will be the smallest of the partitions described in equation (1) [3].

In general, FCM algorithms depend on certain assumptions in order to define the subgroups present in a dataset. These assumptions include the optimal number of classes,  $c$ , the initial centroid values, the initial partition,  $U^0$ , the optimal fuzzy exponent,  $m$  value, and the iteration termination threshold value. The optimal number of classes,  $c$ , may be known a priori or determined by cluster validity process.

In this paper, we determine optimal number of clusters by partitioning the dataset into two clusters. Many past works have proposed methods of determining optimal fuzzy exponent  $m$ . However, many researchers simply use  $m=2$  as the ideal fuzzy exponent [3]. Also, different termination threshold can be chosen. In this paper, two iteration termination threshold value were used for two methods, difference between two adjacent partition matrices less than  $10^{-10}$ , and difference between two adjacent iterations cluster centers less than  $10^{-6}$ .

As with many optimization processes, the solution to equation (1) cannot be guaranteed to be a global optimum, that is, the best of the best. What we seek is the best solution available within a prespecified level of accuracy. An effective algorithm for fuzzy classification, called iterative optimization, is explained in the following section.

### B. FCM Algorithm

The algorithm works as follows. Initially a random selection of fuzzy partition matrix  $U$  is chosen, and centers are calculated. The centers  $v_i$  and the membership strengths are calculated through the use of equations. Once the cluster centers are found, points are classified according to their Euclidean distance to each of the centers identified. A point is assigned to that cluster that results in a minimum distance from its center. Usually, there is some stopping criterion with its threshold, and when the result is below a certain threshold, the algorithm terminates. Fuzzy c-means clustering algorithm can be described in the following steps:

1. Consider a set of  $n$  data points to be clustered,  $x_i$ .
2. Assume the number of clusters (classes)  $c$ , is known,  $2 \leq c < n$ .
3. Choose an appropriate level of cluster fuzziness,  $m \in \mathbb{R} > 1$ .
4. Initialize the  $(n \times c)$  sized membership matrix  $U$  to random values such that  $u_{ik} \in [0,1]$  and  $\sum u_{ik} = 1$ .
5. Calculate the cluster centers  $v_j$  using

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik}^m x_{ki}}{\sum_{k=1}^n u_{ik}} \quad (4)$$

for  $j=1, \dots, c$ .

6. Calculate the distance measure for all clusters

$$d_{ik} = d(x_k - v_i) = \sqrt{\sum_{j=1}^m (x_{kj} - v_{ij})^2} \quad (5)$$

$i=1, \dots, c$  and data point  $k=1, \dots, n$ .

7. Update the fuzzy membership matrix  $U$  according to the  $d_{ik}$  data. If  $d_{ik} > 0$  then

$$u_{ik}^{(r+1)} = 1 / \sum_{j=1}^c \frac{d_{ik}^{2(r)}}{d_{jk}^{2(r)}} \quad (6)$$

8. Repeat from (5) until the stopping criterion is less than a given tolerance.

## IV. PATTERN RECOGNITION

Pattern recognition can be defined as a process of identifying structure in data by comparisons to known structure; the known structure is developed through methods of classification. In the statistical approach to numerical pattern recognition, each input observation is represented as a multidimensional data vector (feature vector) where each component is called a feature. The purpose of the pattern recognition system is to assign each input to one of  $c$  possible pattern classes (or data clusters).

Presumably, different input observations should be assigned to the same class if they have similar features and to different classes if they have dissimilar features. The data used to design a pattern recognition system are usually divided into two categories: design (or training) data and test data. Basically, classification establishes (or seeks to determine) the structure in data, whereas pattern recognition attempts to take new data and assign them to one of the classes defined in the classification process. Simply stated, classification *defines* the patterns and pattern recognition *assigns* data to a class.

A typical problem in pattern recognition is to collect data from a physical process and classify them into known patterns. Suppose we have  $c$  typical patterns represented as fuzzy sets  $A_i$  on  $X$  ( $i=1, 2, \dots, c$ ) and a new piece of data, perhaps consisting of a group of observations, is represented by a fuzzy set  $B$  on  $X$ . The task now is to find which  $A_i$  the sample  $B$  most closely matches. To address this issue, we use fuzzy vectors. Let us define  $\mathbf{a}$  and  $\mathbf{b}$  as a fuzzy vectors of length  $n$ , and define

$$\mathbf{a} \bullet \mathbf{b}^T = \square(a_i \square b_i) \quad (7)$$

as the fuzzy inner product of  $\mathbf{a}$  and  $\mathbf{b}$ , and

$$\mathbf{a} \square \mathbf{b}^T = \square(a_i \square b_i) \quad (8)$$

as the fuzzy outer product of  $\mathbf{a}$  and  $\mathbf{b}$ . These two norms, the inner product and the outer product, can be used simultaneously in pattern recognition studies because they measure *closeness* or *similarity*.

We can extend fuzzy vectors to the case of fuzzy sets. Let  $P(X)$  be a group of fuzzy sets. Now we define two fuzzy sets from this family of sets, that is,  $A, B \in P(X)$ ; then, either of the expressions

$$(A, B)_1 = (A \bullet B) \square (\overline{A \oplus B}) \quad (9)$$

$$(A, B)_2 = \frac{1}{2} [(A \bullet B) + (\overline{A \oplus B})] \quad (10)$$

describe two metrics to assess the degree of similarity of the two sets  $A$  and  $B$ :

$$(A, B) = (A, B)_1 \quad \text{or} \quad (A, B) = (A, B)_2 \quad (11)$$

In particular, when either of the values of  $(A, B)$  approaches 1, the two fuzzy sets  $A$  and  $B$  are “more closely similar”; when either of the values  $(A, B)$  approaches a value of 0, the two fuzzy sets are “more far apart” (dissimilar). The metric in equation (9) uses a minimum property to describe similarity, and the expression in equation (10) uses arithmetic metric to describe similarity [3].

Another way to measure the similarity between two fuzzy sets  $A$  and  $B$  is using Euclidean distance.

$$D(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (12)$$

This distance is smaller when two sets are closer to each other.

## V. PERFORMANCE MEASURES

As the performance measures, the classification accuracy, sensitivity, specificity, positive and negative predictive values have been used and explained as follows. A confusion matrix contains information about actual and predicted classifications done by a classification system. 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.

TABLE III  
CONFUSION MATRIX

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

$$accuracy(\%) = \frac{TP + TN}{TP + FP + FN + TN} \times 100 \quad (13)$$

$$sensitivity(\%) = \frac{TP}{TP + FN} \times 100 \quad (14)$$

$$specificity(\%) = \frac{TN}{FP + TN} \times 100 \quad (15)$$

$$Positive\ predictive\ value(\%) = \frac{TP}{TP + FP} \times 100 \quad (16)$$

$$Negative\ predictive\ value(\%) = \frac{TN}{FN + TN} \times 100 \quad (17)$$

Sensitivity, specificity, and accuracy are the terms which are most commonly associated with a binary classification test and they statistically measure the performance of the test. In a binary classification, we divide a given data set into two categories on the basis of whether they have common properties or not by identifying their significance and in a binary classification test, as the name itself conveys, we deal with two datasets. Of these two categories, in general, sensitivity indicates, how well the test predicts one category and specificity measures how well the test predicts the other category. Whereas accuracy is expected to

measure how well the test predicts both categories. In our case, test sensitivity is the ability of a test to correctly identify those with the disease (true positive rate), whereas test specificity is the ability of the test to correctly identify those without the disease (true negative rate). Accuracy indicates total success of both positive and negative cases. Positive predictive value, or precision rate is the proportion of positive test results that are true positives (such as correct diagnoses). Similar, negative predictive value is defined as the proportion of subjects with a negative test result who are correctly diagnosed.

## VI. RESULTS

The main goal of this paper was to understand how different classifiers would behave when encountering the chosen data and to compare their performance. In this study, fuzzy c-mean clustering and pattern recognition have been used for classification of Parkinson's disease.

First method used here is fuzzy c-means classification method, with  $c=2$  and with randomly chosen initial partition matrix  $U$ . As the stopping criterion in this case we used  $10^{-10}$  difference between two partition matrices. The classification true/false counts for this case are listed in column 2 of Table 4.

TABLE IV  
CONFUSION MATRIX FOR FCM CLASSIFICATION

Confusion matrix	
True positive (%)	51.02
True negative (%)	81.25
False positive (%)	18.75
False negative (%)	48.98
<b>Success (%)</b>	<b>58.46</b>

Because of the unbalanced data, FCM clustering applied on Parkinson's dataset yields relatively poor classification result of 58.46 % success. The low percentage of false positives still makes it usable in diagnosis of the disease. In order to improve results, neural networks can be used.

In the second part of the research, we used pattern recognition method. Whole data is divided into training and testing sets. First half of the positive data set is used for training and second half for testing. Same thing we did with negative data. Therefore, total train and test data have same number of positive and negative samples. The FCM algorithm is used for a prior training by using the representative samples of both healthy and sick people and generating the two different clusters. We ended the simulation when difference between two adjacent centers of the clusters was less than  $10^{-6}$ .

After two classes have been identified, the Euclidean distance measures between mean of centers of these two classes and new testing data are applied to pattern recognitions. In this paper we

also used the similarity measure in (9), and (10) for pattern recognition, but we get the best results by using Euclidean distance.

The detection performances of combined training and test sets are illustrated in Table 5. Table gives the results obtained from FCM and pattern recognition methods.

TABLE V  
CONFUSION MATRIX FOR DISEASE DIOGNOSIS

Confusion matrix	FCM and Pattern
True positive (%)	80.88
True negative (%)	37.93
False positive (%)	19.12
False negative (%)	62.07
Success (%)	68.04

The best recognition result (80.88 %) is achieved when using equal number train-test data for FCM clustering and pattern recognition.

If we calculate classification accuracy, sensitivity, specificity, positive and negative predictive value we get the following results:

TABLE VI  
PERFORMANCE MEASURES

Performance measures	FCM
Accuracy [%]	68.04
Sensitivity [%]	75.34
Specificity [%]	45.83
Positive pred. value [%]	80.88
Negative pred. value [%]	37.93

By definition sensitivity relates to the test's ability to identify positive results. For example, a sensitivity of 100% means that the test recognizes all actual positives – i.e. all sick people are recognized as being ill. So does the specificity for negative results. A high rate of specificity shows the ability of our test to rule out the disease in the subject. Conclusively, the highest rate is desired for all these criteria given in the Table 6. Table 6 provides a performance measure for FCM and pattern recognition methods with same number of train and test data, and without repeating any data. As it can be seen basic FCM classifier used on PD dataset provides high performance in positive predictive value meaning it can be used to diagnose positive cases.

## VII. DISCUSSION

Comparing the presented results with those reported in other studies one can notice that the methods of FCM clustering and pattern recognition yields at least as good results as others. Even though this method has produced a little lower classification rate,

the benefits accruing from its simplicity and repeatability may far outweigh this in real applications. The algorithms showing higher performance are quite complicated. Methods thus seem to have excellent reproducibility besides being much simpler.

Thus, the fuzzy c-means clustering and pattern recognition can be widely used for segregating objects according to some measure of similarity and for revealing dissimilarity among pattern vectors of a dataset. Consequently, it can be used for detection of disease such as Parkinson's disease.

## VIII. CONCLUSION

Early detection of any kind of disease is an essential factor. This helps in treating the patient well ahead. In this research paper, we aimed to design a system that would assist doctors in medical diagnosis. This paper presents a diagnostic fuzzy cluster means and pattern recognition systems to help in diagnosis of Parkinson's disease using a set of speech signals. The paper is intended to verify the effectiveness of the application of these classifiers to the Parkinson Dataset. This dataset comprises of 11 attributes with various range of values. Combination of FCM and pattern recognition with randomly chosen test and train data improve the classification success. The best result is in positive predictive value which is 80.88%. This result is quite satisfactory, due to the fact that detecting the PD is a very complex problem, and used methods are very simple. After all, the results obtained here are very encouraging and open the doors of the future research towards the detection of Parkinson's disease.

## REFERENCES

- [1] A. David Gil, B. M. Johnson, Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines, *Global Journal of Computer Science and Technology*, pp. 63-71
- [2] Max A. Little, Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 4, April 2009
- [3] T. J. Ross, Fuzzy Logic with Engineering Applications, Third Edition, ISBN: 978-0-470-74376-8, John Wiley & Sons, 2010
- [4] F. Åström, R. Koker, A parallel neural network approach to prediction of Parkinson's Disease, *Expert systems with applications*, 38(10): 12470-12474, 2011.
- [5] M. F. Caglar, B. Cetisli, I. B. Toprak, Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier, *Journal of Engineering Science and Design*, Vol.1 No.2 pp.59-64, 2010
- [6] P-F. Guo, P. Bhattacharya, N. Khanna, Advances in Detecting Parkinson's Disease, pp. 306-314, 2010.
- [7] E. Micheli-Tzanakou, P. Wojnicki, J. Hamilton, Neuronal Signal Processing in Parkinson's Disease, *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference*, 0-7803-7278-6/02, Vol.1 pp. 77 - 81, IEEE , 2002
- [8] K. Polat, Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering, *International Journal of Systems Science*, 43:4, 597-609

- [9] T. Bocklet, E. Noth, G. Stemmer, H. Ruzickova, J. Ruz, Detection of Persons with Parkinson's Disease by Acoustic, Vocal, and Prosodic Analysis, *Automatic Speech Recognition and Understanding (ASRU), IEEE Workshop*, 978-1-4673-0367-5/11, 478-483, 2011
- [10] H. K. Rouzbahani, M. R. Daliri, Diagnosis of Parkinson's Disease in Human Using Voice Signals, *Basic and Clinical Neuroscience*, 2008126X, Vol. 2 Issue 3 pages/rec.No.12-20, 2011
- [11] P. Verma, R.D.S. Yadava, Fuzzy C-Means Clustering Based Uncertainty Measure for Sample Weighting Boosts Pattern Classification Efficiency, *Computational Intelligence and Signal Processing (CISP), 2012 2nd National Conference*, 978-1-4577-0720-9/12, pp. 31-35, 2012, IEEE
- [12] T. Khan, Parkinson's Disease Assessment using Speech Anomalies: A Review.
- [13] B. K. Kihel, M. Benyettou, Parkinson's Disease Recognition Using Artificial Immune System, *Journal of Software Engineering and Applications*, 2011, 4, 391-395
- [14] R. G. Ramani, G. Sivagami, Parkinson Disease Classification using Data Mining Algorithms, *International Journal of Computer Applications*, Vol. 32, No.9, October 2011
- [15] UCI Machine Learning Database, <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>
- [16] T. K. Koutouzis, D. Stone, Parkinson's Disease, [http://www.emedicinehealth.com/parkinson\\_disease/article\\_em.htm](http://www.emedicinehealth.com/parkinson_disease/article_em.htm), eMedicineHealth Archives
- [17] N. Sietske Heyn, Parkinson's Disease, [http://www.medicinenet.com/parkinsons\\_disease/article.htm](http://www.medicinenet.com/parkinsons_disease/article.htm)
- [18] N. M. Tahir, H. H. Manap, Parkinson Disease Gait Classification based on Machine Learning Approach. *Journal of Applied Sciences*, 12: 180-185, 2012.
- [19] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, *Expert Systems with Applications*, 2009
- [20] Y. Delican, L. Ozilmaz, T. Yildirim, Evolutionary Algorithms Based RBF Neural Networks For Parkinson s Disease Diagnosis, *ELECO 2011 7th International Conference on Electrical and Electronics Engineering*, 290-294, December 2011, Bursa, Turkey
- [21] M. Ene, Neural network-based approach to discriminate healthy people from those with Parkinson's disease, Vol. 35, 112-116, ISSN: 1223-6934, January 2008.
- [22] V. Cutsuridis, S. Perantonis, A neural network model of Parkinson's disease bradykinesia, *Neural Networks* 19 , 354-374, 2006
- [23] S.Wu, J. Guo, A Data Mining Analysis of the Parkinson's Disease, *iBusiness*, 3, 71-75, 2011.
- [24] D.S.V.G.K.Kaladhar, Confusion Matrix Analysis for Evaluation of Speech on Parkinson Disease using Weka and Matlab, *International Journal of Engineering Science and Technology*, Vol. 2 (7), 2734-2737, 2010.
- [25] A.Khemphila, V. Boonjing, Parkinsons Disease Classification using Neural Network and Feature selection, *World Academy of Science, Engineering and Technology* 64, 15-18, 2012.
- [26] A.H. Hadjahmadi, T. J. Askari, A Decision Support System for Parkinson's Disease Diagnosis using Classification and Regression Tree, *The Journal of Mathematics and Computer Science*, Vol. 4 No.2, 257 - 263, 2012.
- [27] National Institute of Neurological Disorders and Stroke, Parkinson's Disease Information Page, [http://www.ninds.nih.gov/disorders/parkinsons\\_disease/parkinsons\\_disease.htm](http://www.ninds.nih.gov/disorders/parkinsons_disease/parkinsons_disease.htm)
- [28] E. Kaya, O. Findik, I. Babaoglu, A. Arslan, Effect of Discretization Method on the Diagnosis, *International Journal of Innovative Computing, Information and Control*, ISSN 1349-4198, Vol. 7, No. 8, pp. 4669-4678, August 2011.
- [29] D. Wu, K. Warwick, Z. Ma, M. N. Gasson, J.G. Burgess, S. Pan, T. Z. Aziz, Prediction of Parkinson's Disease Tremor Onset using a Radial Basis Function Neural Network based on Particle Swarm Optimization, *International Journal of Neural Systems*, Vol. 20, No. 2, 109-116, 2010.
- [30] N. M. Tahir, H. H. Manap, Parkinson Disease Gait Classification based on Machine Learning Approach, *Journal of Applied Sciences*, pp 1-6, 2012.
- [31] S. M. Kamruzzaman, A. R. Hasan, A. B. Siddiquee, E. H. Mazumder, Medical Diagnosis using Neural Network, *3rd International Conference on Electrical & Computer Engineering, ICECE 2004*, 28-30 December 2004, Dhaka, Bangladesh.
- [32] U.S. National Library of Medicine, PubMed Health, <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001762/>.
- [33] M. El-Rashidy, T.Taha, N. Ayad, H. Sroor, An Effective Intelligent Model for Medical Diagnosis, *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 5, No. 1, September 2011.
- [34] J. Jankovic, Parkinson's Disease: Clinical Features and Diagnosis, *J Neurol Neurosurg Psychiatry* 2008, Vol. 79, 368-376