CrossMark

ORIGINAL ARTICLE

A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method

Hüseyin Gürüler[1]

Abstract Parkinson's disease (PD) is a degenerative, central nervous system disorder. The diagnosis of PD is difficult, as there is no standard diagnostic test and a particular system that gives accurate results. Therefore, automated diagnostic systems are required to assist the neurologist. In this study, we have developed a new hybrid diagnostic system for addressing the PD diagnosis problem. The main novelty of this paper lies in the proposed approach that involves a combination of the k-means clustering-based feature weighting (KMCFW) method and a complex-valued artificial neural network (CVANN). A Parkinson dataset comprising the features obtained from speech and sound samples were used for the diagnosis of PD. PD attributes are weighted through the use of the KMCFW method. New features obtained are converted into a complex number format. These feature values are presented as an input to the CVANN. The efficiency and effectiveness of the proposed system have been rigorously evaluated against the PD dataset in terms of five different evaluation methods. Experimental results have demonstrated that the proposed hybrid system, entitled KMCFW–CVANN, significantly outperforms the other methods detailed in the literature and achieves the highest classification results reported so far, with a classification accuracy of 99.52 %. Therefore, the proposed system appears to be promising in terms of a more accurate diagnosis of PD. Also, the application confirms the conclusion that the reliability of the classification ability of a complex-valued algorithm with regard to a real-valued dataset is high.

Keywords Parkinson's disease · Automated diagnostic systems · K-means clustering-based feature weighting · Complex-valued artificial neural network

1 Introduction

PD is a neurodegenerative disease of the central nervous system that causes partial or complete loss of the motor reflexes, speech, behavioral and mental processes, and other vital functions [1, 2]. In this disease, loss of the neurons that produce dopamine molecules in the brain is observed. It was described and named in 1817 by Dr. James Parkinson [3]. In a comprehensive study that has been carried out recently, the incidence of the disease was given as 20/100,000 [4]. It is known that there are more than one million patients with PD in North America alone [5]. In addition, it is estimated that currently 20 % of patients are not diagnosed correctly [6]. PD affects a significant part of the population and impacts on approximately 1 % of those over 50 years of age [7]. This ratio is expected to increase as people live longer, thus aging is an important risk factor in PD [8].

Some of the PD symptoms can be reduced with pharmacological and/or surgical intervention, and the life span of the patients can consequently be extended. Currently, no specific method has been developed for PD diagnosis. Specialists use many different measurement techniques such as the Unified Parkinson's Disease Rating Scale (UPDRS), the Hoehn–Yahr Scale, the Schwab and England Scale of Activities of Daily Living, the Parkinson's Disease Questionnaire 39, and the Parkinson's Disease Quality of Life Questionnaire for measuring the severity of PD. The

✉ Hüseyin Gürüler
hguruler@mu.edu.tr

1 Department of Information Systems Engineering,
   Faculty of Technology, Mugla Sitki Kocman University,
   48000 Muğla, Turkey

UPDRS is the most commonly used technique [9]. These scales are based on the history of the disease and usually help to detect the existence and severity of symptoms. However, these processes are known to be both time- and effort-consuming [10, 11].

In recent years, computer-based solutions research has considerably increased the support provided to medical decision making. When these studies are reviewed, it appears that the relationship between speech disorders and PD is proved [12–14]. Also, many studies have indicated the reduction in the use of speech as the disease progresses [15, 16]. Therefore, speech samples of the patient are ideal in terms of a decision support system that can be used to perform a diagnosis. This is because it is a noninvasive technique, and the speech data can be collected easily. Speech samples have been used in several investigations with regard to the diagnosis of PD [17–22].

Recent studies have proposed some machine learning methods using audio recordings associated with PD. Little et al. [23] aimed to analyze the stage of the disease by measuring the dysphonia that occurs due to PD. In their study, they made sound recordings of the constant "a" vowel of 31 subjects, including 23 patients with PD. Then, the dysphonia criteria were removed from these sounds and attempts were made to determine the level of the disease by remote monitoring. Shahbaba et al. [24] presented a nonlinear model based on a Dirichlet mixture for diagnostic purposes. Das [25] carried out a comparative analysis using four different methods. Guo et al. [26] proposed a method based on a genetic algorithm (GA) and expectation maximization (EM). Luukka [27] proposed a new method using fuzzy entropy measures and similarity classifiers. Li et al. [28] used a fuzzy-based nonlinear transformation approach with a support vector machine (SVM) with regard to a PD dataset. Ozcift et al. [29] submitted a new classification scheme based on SVM selected attributes to train rotation forest (RF) ensemble classifiers in order to improve the diagnosis of PD. Spadoto et al. [30] have proposed an evolutionary-based method involving an optimum-path forest (OPF) classifier for the diagnosis of PD. Polat [31] applied a fuzzy c-means clustering feature weighting (FCMFW) method with a k-nearest neighbor (KNN) classifier. Zuo et al. [32] used a new diagnostic model based on particle swarm optimization (PSO) for the diagnosis of PD. Sakar and Kursun [33] applied a common knowledge-based feature selection with permutation tests to determine the validity and statistical significance of the relationship between features of the illness with UPDRS scores and created a classification model by giving selected features to the SVM classifier. Chen et al. [34] proposed a detection system using a fuzzy k-nearest neighbor approach with principal component analysis (PCA). Ma et al. [35] obtained high accuracy rates with a kernel-based extreme learning machine followed by a subtractive clustering features weighting. Comparative information about previously performed studies of the diagnosis of PD is given before conclusion section.

In this study, a PD dataset comprising the features obtained from speech samples is used for the diagnosis of PD. As a method, a feature weighting and complex-valued classifier-based new hybrid model is proposed. Feature weighting is used to increase the classification performance. In this study, the KMCFW method is preferred as the weighting method. The aims of KMCFW are (i) to transform the nonlinearly separable dataset into a linearly separable dataset and (ii) to gather similar data points. New features were obtained after the weighting process was converted into a complex number format. In the final stage, these feature values were presented as complex-valued neural network (CVANN) input.

The outline of the study is as follows: General information about the dataset and the methods used in this study are presented in Sect. 2. In Sect. 3, the experimental results are presented. Finally, the outcome of the paper is given in Sect. 4.

2 Materials and methods

2.1 Data

The PD dataset used in this study, comprising speech samples, was created by Max Little with the cooperation of the National Voice and Speech Centre of the University of Colorado and the University of Oxford. It was obtained from the UCI (Machine Learning Repository) [36]. The dataset consists of 195 biomedical sound measurements taken from 31 people consisting of 8 healthy subjects and 23 with PD.

The features of the PD dataset used in this study are as follows: mean, maximum and minimum sound fundamental frequency, irregularity measures in terms of fundamental frequency, amplitude irregularity measurements, measurements of harmonics and the noise ratio, nonlinear dynamic complexity measurements, nonlinear fundamental frequency change measurements, and fractional exponent signal. Also, the PD dataset includes a status column defined as 0 for healthy and 1 for PD patients. Table 1 presents the statistical values of the features of the PD dataset with their definitions [37, 38].

2.2 K-means clustering-based feature weighting (KMCFW): Data preprocessing

The clustering method is a process of dividing the data into groups according to the similarity or uniqueness criteria

Table 1 Statistical values and definitions of the features in PD dataset

Feature label	Definition	Minimum value	Maximum value	Average value	SD
(Vocal fundamental frequencies)					
MDVP: Fo (Hz)	Average vocal fundamental frequency	88.33	260.105	154.22	41.39
MDVP: Fhi (Hz)	High vocal fundamental frequency	102.14	592.03	197.10	91.491
MDVP: Flo (Hz)	Low vocal fundamental frequency	65.476	239.17	116.32	43.521
(Variations in fundamental frequency)					
MDVP: jitter (%)	Jitter percent	0.00168	0.03316	0.00622	0.0048
MDVP: jitter (Abs)	Absolute jitter	7×10^{-6}	0.00026	4.39×10^{-5}	3.48×10^{-5}
MDVP: RAP	Relative average perturbation	0.00068	0.02144	0.0033	0.00296
MDVP: PPQ	Period perturbation quotient	0.00092	0.01958	0.0034	0.00275
Jitter: DDP	Difference of differences of periods	0.00204	0.06433	0.0099	0.00890
(Variations in amplitude)					
MDVP: shimmer	Shimmer percent	0.00954	0.11908	0.0297	0.01885
MDVP: shimmer (dB)	Shimmer in dB	0.085	1.302	0.2822	0.19487
Shimmer: APQ 3	Amplitude perturbation quotient	0.0045	0.0564	0.0156	0.01015
Shimmer: APQ 5	Quotient of amplitude perturbation in 3-point.	0.0057	0.0794	0.0178	0.01202
MDVP: APQ	Quotient of amplitude perturbation in 5-point.	0.00719	0.1377	0.0240	0.01694
Shimmer: DDA	Mean absolute difference between consecutive amplitude differences of consecutive periods.	0.01364	0.1694	0.0469	0.03045
(Ratio of noise to harmonics in the voice)					
NHR	Noise-to-harmonics ratio	0.00065	0.3148	0.0248	0.04041
HNR	Harmonics-to-noise ratio	8.441	33.047	21.885	4.4257
(Nonlinear dynamical complexity measures)					
RPDE	Recurrence period density entropy	0.2565	0.6851	0.49853	0.10394
D2	Correlation dimension	0.57428	0.825	0.7180	0.05533
(Fractional exponent signal)					
DFA	Detrended fluctuation analysis	−7.9649	−2.434	−5.684	1.0902
(Nonlinear measures of fundamental frequency variation)					
Spread 1	Quantifications the fundamental frequency in variation	0.00627	0.4504	0.2265	0.0834
Spread 2		1.423	3.6711	2.3818	0.3827
PPE	Pitch period entropy	0.04453	0.5273	0.2065	0.0901

between data points. Clustering algorithms are not only used for classification but are also used for data compression, feature weighting, and data reduction. The most commonly preferred clustering methods in terms of frequency of use are k-means clustering [39], fuzzy c-means clustering [40], mountain clustering [41], and subtractive clustering [42]. In this study, a data weighting process has been carried out using the k-means clustering (KMC) algorithm which is the most widely preferred in the literature.

In KMCFW, initially, the clusters of each feature are found using KMC. The distance between its cluster and the mean value of that feature is calculated. Features are weighted in accordance with the calculated distance [43].

The aim of the feature weighting method is to map the features according to their distributions in a dataset and also to transform them from nonlinearly separable datasets to linearly separable ones [43]. The feature weighting method works upon the principle that it decreases the variance in features forming the dataset. By means of this, data displaying the same features are gathered together, and the differentiation ability of the classifier is increased.

The k-means algorithm determines the cluster centers, based on minimizing the squared error-based cost function. The purpose of this algorithm is to locate the cluster centers as far away as possible from each other and to associate each data point with the nearest cluster center [44]. Euclidean distance is often used as a measure of

uniqueness in a KMC algorithm. The Euclidean distance (J) is defined as in Eq. 1:

$$J = \sum_{i=1}^{K} \left(\sum_{k} \|x_k - c_i\|^2 \right) \quad (1)$$

where k indicates the number of clusters, c_i indicates the center of the clusters, and x_k indicates the kth pattern in the ith cluster. This pattern is a member of the closest cluster center, and accordingly, the elements of binary membership matrix (u) are defined as in Eq. 2:

$$u_{ij} = \begin{cases} 1, & \|x_j - c_i\|^2 \leq \|x_j - c_t\|^2, \forall t \neq i \\ 0, & \text{others} \end{cases} \quad (2)$$

where u_{ij} indicates whether or not the jth pattern belongs to the ith cluster. Each cluster center minimizing the cost function c_i is defined as in Eq. 3:

$$c_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}} \quad (3)$$

where N indicates the number of patterns.

The working of the KMC algorithm can be summarized as follows:

1. k units are selected randomly as initial cluster centers.
2. Units without cluster centers are assigned in accordance with the defined distance measure to the clusters that the initial cluster centers belong to.
3. New cluster centers are created by averaging the variables in k initial sets that were created.
4. Units are assigned to the closest clusters which are the newly created cluster centers. Distances are calculated.
5. The distances to the previous cluster centers are compared with the distances to the newly created cluster centers.
6. If the distances reduce reasonably, return to step 4.
7. If a fundamental change is not in question, the iteration is finalized and the algorithm is ended.

Briefly, the KMC feature weighting works as follows [43]: At first, the cluster centers are calculated using the KMC method. After calculating the centers of features, the ratios of means of these features to their center are calculated, and these ratios are multiplied by the data point of each feature. Figure 1 shows the flowchart of KMCFW. Figure 2 shows the pseudo-code of the feature weighting.

2.3 Complex-valued artificial neural network (CVANN)

In a complex-valued neural network algorithm, input signals, weights, threshold values, and output signals are all complex numbers (Fig. 3). Recently, the use of
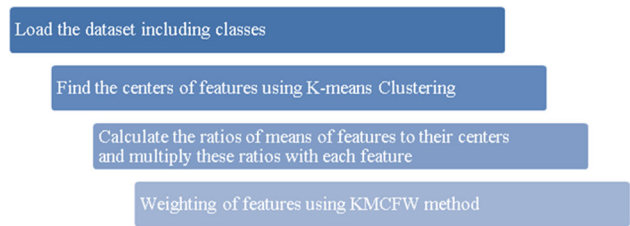

Fig. 1 Flowchart of KMCFW

complex-valued classifiers has increased for the solution of different classification problems [45–49].

There are many studies in the literature emphasizing the advantages of complex-valued ANNs compared to real-valued ANNs [45–47, 50]. These advantages are high-level functionality, better plasticity, and greater flexibility. Additionally, they learn faster and arrive at better generalizations [51]. Neurons in a complex-valued neural network have the ability to learn without generating higher degree inputs and progress to a higher-dimensional space. In addition, the study by Nitta et al. [50] can be examined to see the advantages of CVANN more clearly. This study shows that the XOR problem, which cannot be solved by using two-layered real-valued neural networks, can be easily solved using two-layered CVANN.

2.3.1 The mathematical model of the CVANN algorithm

The mathematical model of complex-valued neural networks is as presented below [52, 53]. The active value of the n neuron Y_n can be defined as follows:

$$Y_n = \sum_{m} W_{nm} I_m + \theta_n \quad (4)$$

In Eq. 4, W_{nm} is a complex-valued connection weight between the n neuron and the m neuron. I_m is a complex-valued input signal of the m neuron, and θ_n is a complex-valued threshold value of the n neuron. To obtain the complex-valued output signal Y_n, the active value is converted into two components in the form of real and imaginary parts as shown below:

$$Y_n = x + iy = z \quad (5)$$

Here, i stands for the value of $\sqrt{-1}$. Considering the various output functions of each neuron, the output function can be defined using the following equation:

$$f_c(z) = f_R(x) + i.f_R(y) \quad (6)$$

$f_R(x)$ and $f_R(y)$ are expressed as the activation function of the neural network. Suppose that the sigmoid function is selected as the activation function. In this case, $f_R(u) = 1/(1 + \exp(-u)), u \in R$ (R specifies the set of real

Fig. 2 Pseudo-code of the feature weighting

> **Initialize** total values (inputs, outputs, and distances) to zero
> **Load** the A dataset including n features and m output
> **For** each feature in dataset
> Step 1. Calculate the cluster centers using the k-means clustering algorithm
> Step 2. Compute the ratios of the means of features to their centers and multiply these ratios with each feature
> Step 3. Multiply these ratios with each data in features (weighting process)
> **End For**
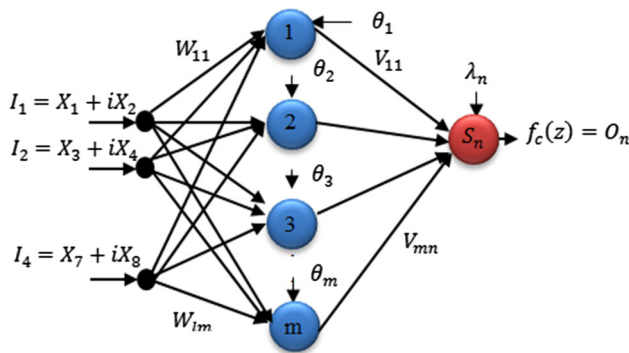
Fig. 3 A three-layer complex-valued neural network [47]. It comprises input, output, and hidden layers. *Each circle* represents a single neuron. $I_N, W_{lm}, \theta_m, O_n, z$ and $f_c(z)$ are all complex numbers

numbers), the real and imaginary parts of an output of a neuron mean the sigmoid functions of the real part x and the imaginary part y of the net input z to the neuron, respectively.

Figure 3 presents the three-layered (input, hidden, and output) CVANN structure used in the study. W_{ml} is the weight between the input layer neuron l and the hidden layer neuron m; V_{nm} is the weight between the hidden layer neuron m and the output layer neuron n; θ_m indicates the threshold value for the hidden layer neuron m; and γ_n indicates the threshold value for the output layer neuron n. I_l, H_m, and O_n indicate the input layer neuron l, the hidden layer neuron m, and the output layer neuron n, respectively. Similarly, U_m and S_n indicate the active values of the hidden layer neuron m and the output layer neuron n, respectively.

$$U_m = \sum_l W_{lm} I_l + \theta_m \tag{7}$$

$$H_m = f_c(U_m) \tag{8}$$

$$S_n = \sum_m V_{mn} H_m + \lambda_n \tag{9}$$

$$O_n = f_c(S_n) \tag{10}$$

In this study, the square error function was preferred. It is expressed as shown in Eq. 11 for p pattern:

$$E_p = (1/2) \sum_{n=1}^{N} |T_n - O_n|^2 = (1/2) \sum_{n=1}^{N} |\delta^n|^2 \tag{11}$$

where N is the number of neurons in the output layer. $(\delta^n = T_n - O_n)$ is the error between O_n, obtained by the n output layer neuron, and T_n, the target output. The square error can also be rewritten as:

$$E_p = (1/2) \sum_{n=1}^{N} \left\{ |Re(T_n) - Re(O_n)|^2 + |Im(T_n) - Im(O_n)|^2 \right\} \tag{12}$$

In order to minimize the square error E_p, the learning rule for the complex-valued back-propagation model is described below [54]. Configuration of weights and threshold values is done according to the following equations (where $\eta > 0$, η is a small learning constant):

$$\Delta V_{nm} = -\eta . \frac{\partial E_p}{\partial Re[V_{nm}]} - i.\eta \frac{\partial E_p}{\partial Im[V_{nm}]} \tag{13}$$

$$\Delta \lambda_n = -\eta . \frac{\partial E_p}{\partial Re[\lambda_n]} - i.\eta \frac{\partial E_p}{\partial Im[\lambda_n]} \tag{14}$$

$$\Delta W_{ml} = -\eta . \frac{\partial E_p}{\partial Re[W_{ml}]} - i.\eta \frac{\partial E_p}{\partial Im[W_{ml}]} \tag{15}$$

$$\Delta \theta_m = -\eta . \frac{\partial E_p}{\partial Re[\theta_m]} - i.\eta \frac{\partial E_p}{\partial Im[\theta_m]} \tag{16}$$

Expressions given from Eqs. 13 to 16 can be rewritten as follows:

$$\Delta V_{nm} = \overline{H_m} \Delta \lambda_n \tag{17}$$

$$\Delta \lambda_n = \eta (Re[\delta^n](1 - Re[O_n])Re[O_n] + i.Im[\delta^n](1 - Im[O_n])Im[O_n]) \tag{18}$$

$$\Delta W_{ml} = \overline{I_l} \Delta \theta_m \tag{19}$$

$$\Delta\theta_m = \eta \left[x\sum_n \begin{pmatrix} (1-Re[H_m])Re[H_m] \\ \begin{pmatrix} Re[\delta^n](1-Re[O_n]) \\ Re[O_n]Re[V_{nm}] \\ +Im[\delta^n](1-Im[O_n]) \\ Im[O_n]Im[V_{nm}]) \end{pmatrix} \end{pmatrix} \right]$$

$$- i\eta \left[x\sum_n \begin{pmatrix} (1-Im[H_m])Im[H_m] \\ \begin{pmatrix} Re[\delta^n](1-Re[O_n]) \\ Re[O_n]Im[V_{nm}] \\ -Im[\delta^n](1-Im[O_n]) \\ Im[O_n]Re[V_{nm}]) \end{pmatrix} \end{pmatrix} \right]$$

(20)

2.3.2 Summary of the CVANN algorithm

- Initialization: Assign all weight and threshold values as numbers.
- Submission of inputs and outputs (the target): Providing complex-valued input vectors $I_1, I_2, I_3, …, I_N$ and corresponding complex-valued output (target) vectors $T_1, T_2, T_3, …, T_N$ to the network. N is the number of patterns to be used in training.
- Calculation of actual output: Calculate the *actual output* (Y_n). Actual output is calculated using Eq. 10.
- Determining the error value: Calculating the error value depends on the obtained output and the target output value according to Eq. 11.
- Changing the weight and threshold values: Update the weight and threshold values using the formulas in Eqs. 17–20. Continue this process until the error is minimized.

2.4 Application and experimental results

In this study, a new hybrid model is proposed for PD diagnosis. As shown in Fig. 4, the proposed method consists of two steps: In the first step, features in the PD dataset were weighted using the KMCFW method. The aim of this method is to map the features according to their distributions in a dataset and to transform from linearly non-separable space to linearly separable space. Using this method, similar data in the same feature are gathered together. This will substantially help to improve the differentiation ability of the classifiers [31, 43]. In the next step, an input set was created by obtaining a complex value from two real values for CVANN input. For example, the first feature value is X_1 and the second one is X_2. These two feature values are converted into the complex number format as $X_1 + iX_2$.

In this way, 11 complex-valued features were obtained from 22 feature values. The feature values obtained in the last step are classified using the CVANN algorithm. The block diagram of the proposed system is shown in Fig. 4.

Figure 5 shows a box graph representation of the original and weighted PD dataset with all 22 features. Figures 6 and 7 show the 3D distribution of two classes of the original and weighted 195 samples formed by the best three principal components obtained using the PCA algorithm. From Figs. 5, 6, and 7, it can be seen that the differentiation ability of the original PD dataset has been improved substantially using the KMCFW approach. After the data preprocessing step, the classification algorithm has been used and has differentiated the weighted PD dataset.

In the classification stage, the CVANN algorithm was preferred. The neural network architecture gives the highest accuracy rate, and its parameters were found empirically. Accordingly, the optimal network structure (input–hidden–output) has been identified as 11-10-1. The learning coefficient was determined as 0.9, and the maximum number of iterations was determined as 1000. Complex sigmoid was selected as the activation function.

The prediction performance of the KMCFW–CVANN method was tested using five different performance evaluation criteria, the formulations of which are given below. These criteria are accuracy, sensitivity, specificity, measure, and kappa statistic value, respectively.
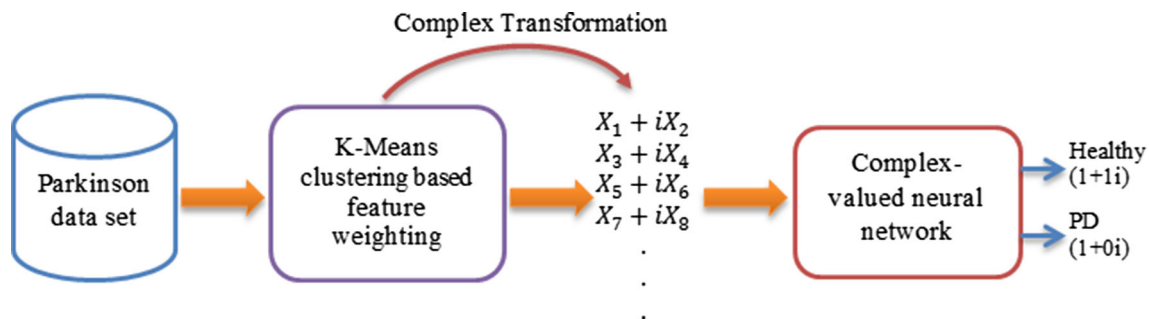

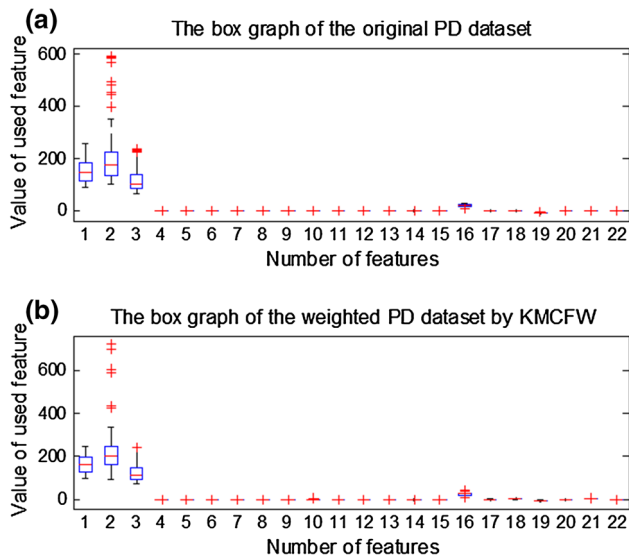**Fig. 4** *Block diagram* of the proposed system for the diagnosis of PD
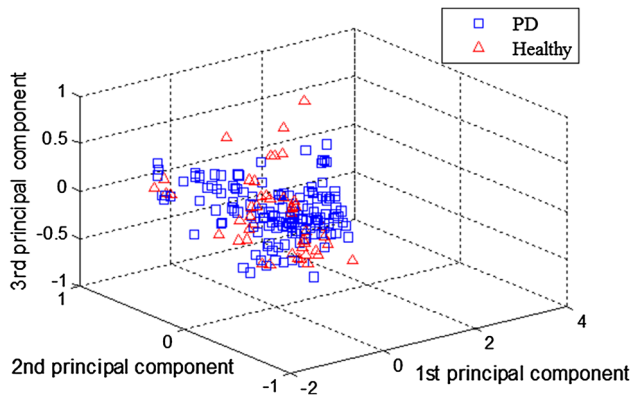
Fig. 5 Original and weighted PD dataset

Fig. 6 3D distribution of two classes in the *original* feature space
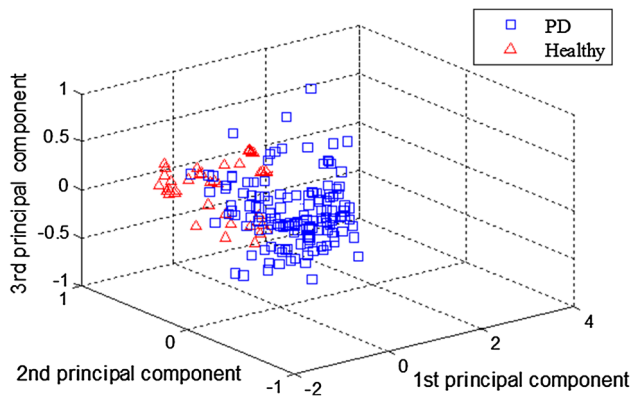

Fig. 7 3D distribution of two classes in the *weighted* feature space

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\,\% \qquad (21)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\,\% \qquad (22)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \times 100\,\% \qquad (23)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (24)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (25)$$

$$\text{f-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (26)$$

where the f-measure is composed of precision and recall values. TP is the number of true positives, which represents the fact that some cases within the PD class are correctly classified as having PD. FN is the number of false negatives, which represents that some cases within the PD class are classified as being healthy. TN is the number of true negatives, which represents that some cases within the healthy class are correctly classified as healthy, and FP is the number of false positives, which represents that some cases within the healthy class are classified as PD.

Kappa statistics is an alternative way for evaluating the accuracy of classifiers. Initially, it was introduced as a measure for measuring the degree of consistency between two observers [55]. Since then, it has been used in a variety of disciplines. In the field of machine learning, this measure is used to compare the accuracy of a classifier with the accuracy of a random classifier which estimates by chance. This measure is defined as:

$$\text{KS} = \frac{P_0 - P_c}{1 - P_c} \qquad (27)$$

P_0 is the accuracy of the classifier, while P_c is the accuracy obtained by random guessing with regard to the same dataset. The Kappa statistic values can be between -1 and 1. -1 indicates complete inconsistency (completely wrong classification), while 1 indicates perfect consistency (completely correct classification). The results obtained according to the said performance evaluation criteria are presented in Table 2. In addition, the results obtained by the application of the ANN method to the same feature values are added to the table. To make an equal comparison with the results obtained by different researchers, both k-fold cross-validation and 50–50 % holdout methods are preferred as data distribution methods. The experiment was repeated 10 times to determine the reliability and stability of the results, and the average values of the obtained values were selected. When we analyze Table 2, it can be seen that the CVANN method gives much better results compared to the real-valued ANN.

The comparative analysis of the diagnosis of PD performed in this study in terms of previously performed

Table 2 Results obtained according to the performance evaluation criteria

Method	Performance metrics	Tenfold cross-validation				50–50 % holdout method			
		Mean	SD	Max	Min	Mean	SD	Max	Min
ANN	ACC	**95.85**	4.26	98.66	88.05	**94.98**	4.79	97.96	86.79
	Sensitivity	**96.03**	3.85	98.55	90.59	**95.97**	4.18	98.15	88.65
	Specificity	**95.41**	4.05	97.99	89.85	**95.34**	3.08	96.09	91.80
	f-Measure	**0.9549**				**0.9479**			
	Kappa	**0.8896**				**0.8705**			
CVANN	ACC	**99.52**	0.24	100	98.05	**99.39**	0.45	100	97.95
	Sensitivity	**100**	0	100	100	**100**	0	100	100
	Specificity	**99.47**	0.27	100	98.02	**99.42**	0.39	100	98.01
	f-Measure	**0.9978**				**0.9917**			
	Kappa	**0.9899**				**0.9885**			

Table 3 Comparison of the results with the literature

Study	Method	Data selection method	Accuracy (%)
Shahbaba and Neal [25]	Dirichlet process mixtures	(Fivefold CV)	87.70
Das [26]	ANN	(Holdout)	92.90
Sakar and Kursun [35]	Mutual information + SVM	(Bootstrap with 50 replicates)	92.75
Guo et al. [27]	GA-EM	(Tenfold CV)	93.10
Luukka [7]	Fuzzy entropy measures + similarity	(Holdout)	85.03
Li et al. [28]	Fuzzy-based nonlinear transformation + SVM	(Holdout)	93.47
Ozcift et al. [30]	Correlation-based feature selection-rotation forest	(Tenfold CV)	87.10
Spadoto et al. [31]	PSO + OPF harmony search + OPF gravitational search + OPF	(Holdout)	84.01
Daliri et al. [33]	SVM with Chi-square distance kernel	(50–50 % Training–testing)	91.20
Astrom and Koker [56]	Parallel ANN	(Holdout)	91.20
Polat et al. [32]	FCMFW + KNN	(50–50 % Training–testing)	97.93
Chen et al. [57]	PCA-fuzzy k-nearest neighbor	(Average tenfold CV)	96.07
Zuo et al. [34]	PSO-fuzzy k-nearest neighbor	(Tenfold CV)	97.47
Proposed method	KMCFW–CVANN	(Tenfold CV)	99.52
Proposed method	KMCFW–CVANN	(50–50 % training–testing)	99.39

studies is given in Table 3. As shown, our proposed method obtains better classification results than all the methods proposed in previous studies. The accuracy rates obtained by other researchers vary between 85 and 97 %. The proposed method gives better result with an accuracy of 99.52 %. In an important issue such as medical diagnostics and diagnostic systems, even a 0.1 % increase can be very important. Consequently, the proposed method is expected to make an important contribution to this field.

There is no significant difference between the proposed method and the methods presented in Table 3 in terms of simplicity and computational load, the proposed method having two steps combining feature obtaining and classification. Using CVANN in the classification stage does not lead to an additional computational load. When analyzing the computation time, it can be seen from Table 3 that a complex-valued classifier allows faster classification compared to a real-valued classifier. As a result, the proposed method is fast and with a light computational load.

3 Conclusion

The paper presents an automated diagnostic system supporting the neurologist in the diagnosis of PD. The main novelty of this paper lies in the proposed system, which is entitled KMCFW–CVANN, that integrates an effective clustering features weighting method and a fast classifier. It allows the diagnosis of PD in an efficient and fast manner.

In this study, a Parkinson's dataset comprising the features obtained from speech and sound samples was used. In the proposed method, KMCFW was used as a data

preprocessing tool, with the aim of decreasing the variance in features of the PD dataset in order to further improve the diagnostic accuracy of the CVANN classifier.

It can be seen from the experiments that the *complex-valued* ANN method gives a much better result compared to *real-valued* ANN. The prediction performance of the KMCFW–CVANN hybrid method was tested with five different performance evaluation criteria. These are accuracy, sensitivity, specificity, f-measure, and kappa statistic value. The proposed method gave better results with an accuracy value of 99.52 %. With this value, it is clear that the proposed system outperforms other methods proposed in the literature.

All of this points to the fact that the proposed system using complex-valued classifiers can be shown to have a positive impact in terms of providing an accurate and rapid diagnosis of PD. It is projected that such high accuracy rates with regard to prediction can also be obtained in different medical diagnosis situations.

References

1. Jankovic J (2007) Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 79(4):368–376
2. Khorasani A, Daliri MR (2014) HMM for classification of Parkinson's disease based on the raw gait data. J Med Syst 38(12):1–6
3. Langston JW (2002) Parkinson's disease: current and future challenges. NeuroToxicology 23(4–5):443–450
4. Pahwa R, Lyons KE (2013) Handbook of Parkinson's disease, 5th edn. Informa Healthc, USA
5. Lang AE, Lozano AM (1998) Parkinson's disease—first of two parts. N Engl J Med 339:1044–1053
6. Schrag A, Ben-Schlomo Y, Quinn N (2002) How valid is the clinical diagnosis of Parkinson's disease in the community? J Neurol Neurosurg Psychiatry 7:529–535
7. Moore DJ, West AB, Dawson VL, Dawson TM (2005) Molecular pathology of Parkinson's disease. Annu Rev Neurosci 28:57–87
8. Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA (2002) Risk tables for parkinsonism and Parkinson's disease. J Clin Epidemiol 55:25–31
9. Ramaker C, Marinus J, Stiggelbout AM, van Hilten BJ (2002) Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord 17(5):867–876
10. Ozcift A (2012) SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. J Med Syst 36(4):2141–2147
11. Sakar BE (2014) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings, Ph.D. Thesis, Istanbul University, Turkey
12. Darley FL, Aronson AE, Brown JR (1969) Differential diagnostic patterns of dysarthria. J Speech Hear Res 12:246–269
13. Gamboa J, Jimenez-Jimenez FJ, Nieto A, Montojo J, Orti-Pareja M, Molina JA, García-Albea E, Cobeta I (1997) Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs. J Voice 11:314–320
14. Ho A, Bradshaw JL, Iansek R (2008) For better or for worse: the effect of Levodopa on Speech in Parkinson's disease. Mov Disord 23(4):574–580
15. Harel B, Cannizzaro M, Snyder PJ (2004) Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn 56:24–29
16. Skodda S, Rinsche H, Schlegel U (2009) Progression of dysprosody in Parkinson's disease over time—a longitudinal study. Mov Disord 24(5):716–722
17. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 56(4):1010–1022
18. Harel B, Cannizzaro M, Snyder PJ (2004) Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn 56:24–29
19. Seera M, Lim CP, Tan SC, Loo CK (2015) A hybrid FAM–CART model and its application to medical data classification. Neural Comput Appl. doi:10.1007/s00521-015-1852-9
20. Sapir S, Ramig L, Spielman J, Fox C (2010) Formant centralization ratio (FCR): a proposal for a new acoustic measure of dysarthric speech. J Speech Lang Hear Res 53:114–125
21. Cnockaert L, Schoentgen J, Auzou P, Ozsancak C, Defebve L, Grenez F (2008) Low frequency vocal modulations in vowels produced by Parkinsonian subjects. Speech Commun 50:288–300
22. Erdogdu Sakar B, Isenkul M, Sakar CO, Sertbas A, Gurgen F, Delil S, Apaydin H, Kursun O (2013) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform 17(4):828–834
23. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 56(4):1015–1022
24. Shahbaba B, Neal R (2009) Nonlinear models using Dirichlet process mixtures. J Mach Learn Res 10:1829–1850
25. Das R (2010) A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Syst Appl 37(2):1568–1572
26. Guo PF, Bhattacharya P, Kharma N (2010) Advances in detecting Parkinson's disease. Med Biom 6165:306–314
27. Luukka P (2011) Feature selection using fuzzy entropy measures with similarity classifier. Expert Syst Appl 38(4):4600–4607
28. Li DC, Liu CW, Hu SC (2011) A fuzzy-based data transformation for feature extraction to increase classification performance with small medical datasets. Artif Intell Med 52(1):45–52
29. Ozcift A, Gulten A (2011) Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Comput Methods Progr Biomed 104(3):443–451
30. Spadoto AA, Guido RC, Carnevali FL, Pagnin AF, Falcao AX, Papa JP (2011) Improving Parkinson's disease identification through evolutionary-based feature selection. In: Proceedings of the annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11), pp 7857–7860
31. Polat K (2012) Classification of Parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering. Int J Syst Sci 43(4):597–609
32. Zuo WL, Wang ZY, Liu T, Chen HL (2013) Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. Biomed Signal Process Control 8(4):364–373
33. Sakar CO, Kursun O (2010) Telediagnosis of Parkinson's disease using measurements of dysphonia. J Med Syst 34(4):591–599
34. Chen HL, Huang CC, Yu XG, Xu X, Sun X, Wang G, Wang SJ (2013) An efficient diagnosis system for detection of Parkinson's

disease using fuzzy k-nearest neighbor approach. Expert Syst Appl 40(1):263–271

35. Ma C, Ouyang J, Chen HL, Zhao XH (2014) An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. Comput Math Methods Med. doi:10.1155/2014/985789

36. Parkinsons Dataset. https://archive.ics.uci.edu/ml/datasets/Parkinsons. Accessed 10 Sept 2014

37. Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA (2002) Risk tables for Parkinsonism and Parkinson's disease. J Clin Epidemiol 55:25–31

38. Little MA, McSharry PE, Hunter EJ, Ramig LO (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 56:1015–1022

39. MacQueen JB (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

40. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York

41. Yager RR, Filev DP (1994) Generation of fuzzy rules by mountain clustering. IEEE Trans Syst Man Cybern 24:209–219

42. Chiu SL (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2:267–278

43. Gunes S, Polat K, Yosunkaya S (2010) Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. Expert Syst Appl 37(12):7729–7736

44. Moftah HM, Azar AT, Al-Shammari ET, Ghali NI, Hassanien AE, Shoman M (2014) Adaptive k-means clustering algorithm for MR breast image segmentation. Neural Comput Appl 24(7–8):1917–1928

45. Hirose A, Shotaro Y (2013) Relationship between phase and amplitude generalization errors in complex and real-valued feed-forward neural networks. Neural Comput Appl 22(7–8):1357–1366

46. Ceylan R, Ceylan M, Ozbay Y, Kara S (2011) Fuzzy clustering complex-valued neural network to diagnose cirrhosis disease. Expert Syst Appl 38(8):9744–9751

47. Peker M, Sen B, Delen D (2015) A novel method for automated diagnosis of epilepsy using complex-valued classifiers. IEEE J Biomed Health Inf. doi:10.1109/JBHI.2014.23877952015

48. Sivachitraa M, Savithab R, Sureshb S, Vijayachitrac S (2015) A fully complex-valued fast learning classifier (FC-FLC) for real-valued classification problems. Neurocomputing 149:198–206

49. Shogo O, Arima Y, Hirose A (2014) Millimeter-wave security imaging using complex-valued self-organizing map for visualization of moving targets. Neurocomputing 134:247–253

50. Nitta T (2004) Orthogonality of decision boundaries in complex-valued neural networks. Neural Comput 16(1):73–97

51. Aizenberg I (2011) Complex-valued neural networks with multi-valued neurons. Springer, Heidelberg, pp 264–265

52. Nitta T (1997) An extension of the back-propagation algorithm to complex numbers. Neural Network 10:1391–1415

53. Ozbay Y, Kara S, Latifoglu F, Ceylan R, Ceylan M (2007) Complex-valued wavelet artificial neural network for Doppler signals classifying. Artif Intell Med 40(2):143–156

54. Nitta T (1993) A back-propagation algorithm for complex numbered neural networks. In: Proceedings of 1993 international joint conference on neural networks, pp 1649–1652

55. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Measur 20(1):37–46

56. Astrom F, Koker R (2011) A parallel neural network approach to prediction of Parkinson's disease. Expert Syst Appl 38(10):12470–12474

57. Chen HL, Huang CC, Yu XG, Xuc X, Sund X, Wangd G, Wangd SJ (2013) An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Syst Appl 40(1):263–271