Computer-Aided Diagnosis of Parkinson's Disease Using Complex-Valued Neural Networks and mRMR Feature Selection Algorithm

Musa Peker[1], Baha Şen[2]* and Dursun Delen[3]

[1]Department of Information Technology, Samandira Technical and Vocational High Schools, Sancaktepe, Istanbul, Turkey
[2]Yıldırım Beyazıt University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Ankara.
[3]Department of Management Science and Information Systems, Oklahoma State University, Stillwater, Oklahoma, USA

ABSTRACT

Parkinson's disease (PD) is a neurological disorder which has a significant social and economic impact. PD is diagnosed by clinical observation and evaluations, coupled with a PD rating scale. However, these methods may be insufficient, especially in the initial phase of the disease. The processes are tedious and time-consuming, and hence systems that can automatically offer a diagnosis are needed. In this study, a novel method for the diagnosis of PD is proposed. Biomedical sound measurements obtained from continuous phonation samples were used as attributes. First, a minimum redundancy maximum relevance (mRMR) attribute selection algorithm was applied for the identification of the effective attributes. After conversion to a complex number, the resulting attributes are presented as input data to the complex-valued artificial neural network (CVANN). The proposed novel system might be a powerful tool for effective diagnosis of PD.

Keywords: computer aided diagnosis, Parkinson's disease, complex-valued neural network, classification, mRMR feature selection method

1. INTRODUCTION

Parkinson's disease (PD) is a central nervous system disorder which causes partial or complete passivation of motor reflexes, speech, and other vital functions [1]. PD affects a significant portion of the world population, and impacts on approximately 1% of those over 50 years of age in 2005; this percentage is expected to increase as people live longer [2].

---

*Corresponding author: Baha Şen, University of Yıldırım Beyazıt, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Çankırı Caddesi, Çiçek Sokak No:3 Aındağ, Ulus/Ankara. Turkey. E-mail: bsen@ybu.edu.tr. Other authors' emails: pekermusa@gmail.com; dursun.delen@okstate.edu.

The symptoms and effects of PD deteriorate with time. With pharmacological or surgical intervention, it is possible to reduce some of the symptoms and increase the longevity of those suffering from PD. This is especially the case in the early stages of the disease.

Although it is possible to reduce the effects of its symptoms with medical interventions and drug therapy, there is no cure (i.e., medical treatment) for PD [3, 4]. In measuring the severity level of PD, doctors use different measurement methods including the Hoehn-Yahr Scale and the Combined Parkinson's Disease Rating Scale [5]. These measurements are based on the historical progression of the disease, and are usually helpful in detecting the presence and severity of the spectrum of symptoms. The processes of these measurement regiments are known to be both time and effort consuming [6]. Therefore, developing a system that could help in diagnosing PD would be very useful and practical for medical professionals.

In recent years, the number of computer-based system and related studies to provide diagnostic support to medical professionals has increased considerably [7, 8]. Among these studies, some of them specifically dealt with the diagnosis of PD. A review of the literature revealed the existence of an interesting relationship between speech disorders and PD [9-11]. Research also showed progressively declining performance of sound usage with the patients diagnosed with PD [12, 13]. Therefore, it is postulated that the speech/sound samples of the PD patient can be used as an input to a decision support system for diagnosing the disease [1, 4, 12, 14, 15]. Because collection of these data samples is a non-invasive process, it can be easily obtained from the patient and used as the input to the diagnostic support system.

Using the speech samples for the diagnosis of PD has been the subject of several investigations. For instance, Shahbaba et al. [16] used a non-linear model based on Dirichlet mixtures for the diagnosis of PD. An 87.7% classification accuracy was obtained with this method. Das [17] carried out a comparative study of artificial neural networks (ANN), DMneural, regression and decision trees for the diagnosis of the PD using speech samples. The experimental results showed that the ANN method achieved a 92.9% general classification performance. Guo et al. [18] proposed a hybrid model based on expectation maximization (EM) and a genetic algorithm (GA), and obtained a 93.1% classification accuracy. Luukka [19] proposed a new method that used fuzzy entropy measures to combine with similarity classifiers, and achieved an average accuracy rate of 85.03%. Li et al. [20] introduced a fuzzy-based non-linear transformation approach with a support vector machine (SVM) and achieved a 93.47% classification accuracy. Aström et al. [21] obtained a 91.2% classification accuracy using a parallel neural network model, while Spadoto et al. [22] obtained an 84.01% classification accuracy using an optimum-path forest (OPF) classifier. Polat et al. [23] applied a fuzzy c-means (FCM) clustering-based feature weighting (FCMFW) algorithm with a k-nearest neighbor (KNN) classifier, and achieved a 97.93% classification accuracy. Daliri et al. [24] proposed a chi-square distance kernel based SVM and obtained a 91.2% classification accuracy result. Zuo et al. [25] presented a diagnostic system based on particle swarm optimization (PSO)-enhanced fuzzy k-nearest neighbor algorithm, and achieved a 97.47% average classification accuracy. Sakar and Kursun [26] proposed a hybrid mutual information-based feature selection

method combined with SVM-based classifier, and achieved an accuracy rate of 92.75%. Ozcift and Gulten [27] proposed a method that combined 30 machine learning algorithms with a rotation forest (RF) group classifier. In that study, where a correlation-based feature selection (CFS) algorithm was used as a feature selection algorithm, an 87.13% classification accuracy was obtained. Chen et al. [28] proposed a system for the detection of PD using a fuzzy k-nearest neighbor approach and principal component analysis (PCA), and achieved a 96.07% accuracy. Rouzbahani and Daliri [29] used an SVM-based feature selection method for the diagnosis of PD from sound signals. SVM, KNN and some discrimination-function-based (DBF) algorithms were used as the classification algorithms. The highest accuracy rate was obtained with KNN algorithm with 93.82%. Ma et al. [30] used a kernel-based extreme learning machine with a subtractive clustering features weighting approach and obtained high accuracy rates. Some of the previous studies have also been carried out using gait variability extracted from the sound recordings. For instance, Khorasani and Daliri [31] developed a method based on hidden Markov model (HMM) with Gaussian mixtures using the raw gait data for diagnosis of PD. In this study, they obtained a 90.3% classification accuracy. Daliri [32] also used gait dynamics along with SVM for the automatic diagnosis of PD (a neuro-degenerative diseases) and obtained a classification accuracy of 89.33%.

In this study, a new hybrid approach consisting of feature selection and complex valued neural networks is proposed. For model building and PD diagnosis, a rich dataset consisting of features extracted from speech sound samples, is used. To develop the prediction models, a minimum redundancy maximum relevance (mRMR)-based feature selection algorithm is first applied to the raw data to determine the effective features. With the elimination of the irrelevant and redundant features, the aim was to improve the prediction accuracy of the classification algorithm, and also to reduce the computational burden. The mRMR algorithm was preferred because of its superior performance reported in a large number of previous studies where it was compared to other feature selection methods. After the feature selection, the attributes/features consisting of real values were converted to complex values. These complex valued attributes were then presented to the complex-valued neural network as the input vector. The proposed method was named/entitled mRMR + CVANN (complex-valued artificial neural networks). In the final stage, the classification results obtained from the proposed method were compared to those of the results obtained from previous studies found in the literature.

The rest of the paper is organized as follows. In section 2, a brief description of the individual methods used in this study is provided. In section 3, the details of the proposed diagnosis system is given. In section 4, the experimental design is described in detail. In section 5, the experimental results are summarized and the comparative analysis of these results are presented. In section 6, the findings are summarized and the paper is concluded with final remarks.

2. METHODS

This section provides a short description of the feature selection algorithms used to process the sound data samples. The details of CVANN architecture developed and used to build the prediction/classification models are then explained.

2.1. Minimum Redundancy Maximum Relevance (mRMR) Algorithm

In classification type applications, features that are extracted from the original/raw dataset are used as the inputs to the classification method. In some applications, the number of features may be limited to just a few, while in others the number of features may be too many. The features that are extracted for each application are stored in an attribute matrix. Thus, both number of rows (sample size) and the number of features define the size of the data table and affect the processing time [33]. Features that can distinguish between classes more effectively are called high-level features and are particularly important in terms of the performance of a classifier [34, 35]. Instead of using all features, using only the high-level features (a subset of the total number of features) can reduce processing time and potentially can improve the prediction performance.

The mRMR feature selection algorithm is employed in the study. The mRMR is essentially a filtering algorithm that tries to select the features that are most relevant to the class labels and to filter out the rest. While identifying the most relevant features, the algorithm also tries to minimize the redundancy among the selected/relevant features [36]. Specifically, the mRMR algorithm treats each feature and the class vector (response variable or output variable) as a discrete random variable. To measure the similarity between two features or between one attribute and the class vector, it uses mutual information measure ($I(x, y)$). Mutual information is defined as:

$$I\left(\mathrm{X}, Y\right) = \sum_{i,j} p(x_i, \, y_j) \log \frac{p(x_i, \, y_j)}{p(x_i) p(y_j)} \tag{1}$$

For x and y features, $p(x_i)$ and $p(y_j)$ are marginal probability functions, and $p(x_i, y_j)$ is the joint probability distribution. The mutual information value is 0 where two random variables are completely independent [37], and this value is symmetric and cannot be negative ($I(X, Y) \geq 0$, $I(X, Y) = I(Y, X)$).

Let S be the desired feature set to be selected, while $|S|$ denotes the number of elements of this set. According to the mRMR algorithm, two conditions have to be met in order to select the attribute. The first one is the maximum relevance, *MaxMR*:

$$max \, MR, \;\; MR = \frac{1}{|S|} \sum_{i \in S} I(h, \, i) \tag{2}$$

The second one is minimum redundancy, *minMV*:

$$min \, MV, \qquad MV = \frac{1}{|S|^2} \sum_{i, \, j \in S} I(i, \, j) \tag{3}$$

where $h = \{h_1, h_2, \ldots, h_k\}$. This is the class variable of a dataset with K possible classes. Ω indicates the whole feature set, Ω_s indicates all of the features except the selected one $\Omega_s = \Omega - S$). There are two approaches to combine the two conditions mentioned above:

Mutual Information Difference (MID), defined as max(MR-MV), and Mutual Information Quotient (MIQ), defined as max(MR/MV) [38]. In this study, the feature selection is carried out using MID (because of its superior performance).

In this study, a variety of feature selection algorithms are applied to the original dataset and the results are compared to those of the ones obtained with mRMR to determine the effectiveness. These feature selection algorithms included Fisher score, Chi-square, sequential forward selection (SFS), sequential forward floating search (SFFS) and ReliefF. The Fisher score feature selection algorithm uses productive statistical models that can distinguish the most appropriate features [39]. The Chi-square feature selection algorithm is also one of the most commonly used methods in determining the effective features. With this method, the information value of a feature is measured by calculating the statistical value of chi-square [40]. In the SFS method, the feature selection process begins with an empty subset; for each step thereon, the feature that maximizes the classification accuracy is added to the current feature set. This process is repeated until all the features have been tested and thoroughly evaluated. The subset that maximizes classification accuracy is selected as the best feature set [41]. In the SFFS algorithm, subsets are evaluated using a forward-and-backward motion. Specifically, if a subset produces better results than the previous one, one back step is applied. If performance is not improved, the back step does not apply. In this way, the reverse direction tracking is carried out without the need for dynamic parameter settings [42]. ReliefF is a simple yet effective algorithm that estimates the value of features by measuring the interdependencies. Specifically, this algorithm changes the weight of feature conformity/value using the nearest neighbor algorithm [43].

2.2. Complex Valued Neural Network (CVANN)

CVANN is a type of artificial neural network architecture that has its network parameters in the form of complex numbers. These network parameters include weights, the threshold values as well as inputs and outputs. There are a number of studies in the literature emphasizing the advantages of CVANNs compared to the ordinary real-valued ANNs [44-48]. The use of complex valued input/output, weight and activation functions makes CVANN capable of boosting the functionality (and the resulting performance) of a single neuron and the network of neurons (the neural network), and can also decrease the model building/training time [49, 50]. Figure 1 graphically illustrates a simple comparison between ordinary and complex valued neural network. This example shows that a 2-input ordinary neural network can be reduced to a single input by using a complex valued neural network. Simplifying and using the input values in this way was first proposed by Chen et al. [51]. According to Chen et al. [51], this type of input representation provides a significant reduction in complexity of larger networks and hence faster training and mode building opportunities. The main reason behind the advantages of this method is largely credited to the use of complex numbers that have both real and imaginary parts, and hence, have the ability to contain and pass along two-dimensional information as a single dimension. As mentioned above, this innovative representation leads to downsizing of the network and faster training of the prediction/classification model.
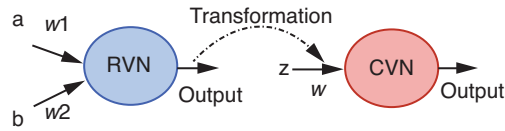
Figure 1.   The representation of neural network with one input and one weight in the complex value (right part of the figure) which is normally realized with 2 inputs and 2 weights in real values (left part of the figure). RVN: Real-valued neuron, CVN: Complex-valued neuron. $z = a + ib$ and $w = w1 + iw2$.
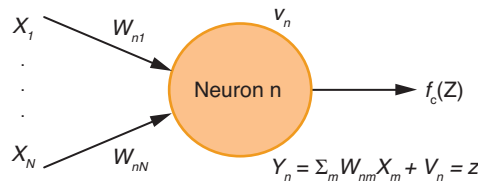

Figure 2.   A simple representation of a single neuron used in CBP.

In addition to that mentioned above, CVANNs also have other advantages compared to real-valued neural networks, including high level functionality, better plasticity and greater flexibility. They tend to learn faster and achieve better generalizations [52]. The capability of a single neuron in a complex-valued neural network is enhanced with its flexibility—it can learn complex and nonlinear input/output mappings at both linear and nonlinear levels. That is, these complex-valued neurons have the ability to learn without generating higher degree inputs and progressing to a higher dimensional space. In a comparative study, Nitta et al. [53] showed that the XOR problem, which cannot be solved using two-layer, real-valued neural networks, can easily be solved using a two-layer CVANN.

In this study, a complex-valued, back-propagation (CBP), feed-forward learning algorithm is used to train the CVANN models. A simple representation of a single neuron used in the CBP algorithm is shown in Figure 2.

Y_n, the activity value of the neuron n can be defined as follows:

$$Y_n = \sum_m W_{nm} X_m + V_n \tag{4}$$

where W_{nm} is the complex valued connection weight between n and m neurons, X_m is the complex-valued input signal of m neuron, and V_n is the complex-valued threshold value of n neuron. To obtain a complex-valued output signal, Y_n activity value is converted into two components in the form of real and imaginary parts, as shown below:

$$Y_n = x + iy = z \tag{5}$$

where x and y indicate real and imaginary part of Y_n value, respectively; i represents $\sqrt{-1}$. Considering the various output functions of each neuron, the overall output functions can be defined using the following equation, where f_c and f_R represent complex and real-valued functions, respectively:

$$f_c(z) = f_R(x) + i.f_R(y) \tag{6}$$

One of the difficulties encountered in CBP applications is the selection of the most appropriate activation function. It is necessary for the activation function to be suitable to the practical applications of complex multilayer perceptron. Detailed information about the features that the complex activation function needs to have can be found in [54]. In the literature, there are several activation functions proposed for CBP. In this study, the preferred complex activation function is a superposition of real and imaginary logarithmic sigmoid [55]. This function is expressed as a complex sigmoid activation function. A complex sigmoid activation function can be defined using the following equation:

$$f_c(z) = f_R(x) + i.f_R(y) = \frac{1}{1 + \exp(-x)} + i\frac{1}{1 + \exp(-y)} \tag{7}$$

The CVANN used in this study has three layers (input, hidden and output). Figure 3 illustrates the three-layered CVANN structure used in this study. A detailed description of the underlying mathematical model of three-layered CVANN can be found in [56, 57].

In Figure 3, W_{ml} is the weight between the input layer neuron I 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 is the threshold value for the hidden layer neuron m, and γ_n represents the
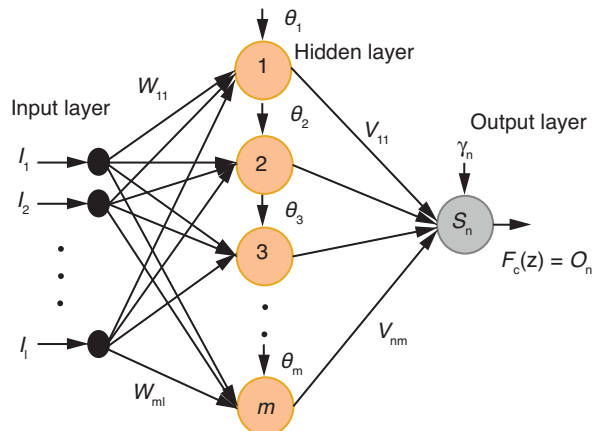

Figure 3.   Complex-valued neural network.

threshold value for the output layer neuron n. I_l, H_m, O_n represent the output values of the input neuron l, the hidden neuron m, and the output neuron n, respectively. Similarly, U_m and S_n, are the internal potentials of the hidden neuron m and the output neuron n, respectively.

$$U_m = \sum_l W_{ml} I_l + \theta_m \tag{8}$$

$$S_n = \sum_m V_{nm} H_m + \gamma_n \tag{9}$$

$$H_m = f_c(U_m) \tag{10}$$

$$O_n = f_c(S_n) \tag{11}$$

This study chose to use a square error function, which can be expressed as:

$$E_p = \left(\tfrac{1}{2}\right) \sum_{n=1}^{N} |T_n - O_n|^2 = \left(\tfrac{1}{2}\right) \sum_{n=1}^{N} |\delta^n|^2 \tag{12}$$

where N is the number of neurons in the output layer, and $\delta^n = T_n - O_n$ is the error between the actual pattern O_n and the target pattern T_n of output neuron n.

The learning rule for the complex-valued back-propagation model is performed using the equations given in [58]. The goal here is to minimize the squared error (E_p). The weights and threshold values are determined using the following equations (where $\eta > 0$, η is a learning rate):

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

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

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

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

The expression from eqns. 13 to 16 can be rewritten as follows:

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

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

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

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

$$- i\eta \left[x \sum_n \begin{pmatrix} \big(1 - Im[H_m]\big) Im[H_m] \\ \\ Re\big[\delta^n\big]\big(1 - Re[O_n]\big) \\ Re[O_n] Im[V_{nm}] \\ - Im\big[\delta^n\big]\big(1 - Im[O_n]\big) \\ Im[O_n] Re[V_{nm}] \end{pmatrix} \right] \tag{20}$$

Summary of the CVANN algorithm:

1. ***Initialization:*** Assign all the weights and threshold values as small complex-valued numbers (greater than zero).
2. ***Presenting inputs and outputs (target):*** Present the complex-valued input vectors $(I_1, I_2, I_3, \ldots, I_N)$ and corresponding complex-valued output vectors (target) $(T_1, T_2, T_3, \ldots, T_N)$ to the network, where N is the number of patterns to be used in training.
3. ***Calculating the actual output (Y_n):*** The actual output is calculated using eqn. 5.
4. ***Calculating the stopping criterion*** according to eqn. 21: The algorithm is stopped when the condition in the equation is met.

$$\sqrt{\sum_{p}\sum_{n=1}^{N}\left|T_n^{(p)}-O_n^{(p)}\right|^2} = 10^{-3} \tag{21}$$

where $T_n^{(p)}$ and $O_n^{(p)}$ are complex numbers and indicate the target and output values of n neuron for p pattern, respectively; N indicates the number of neurons in the output layer.

5. ***Changing the weight and threshold values:*** Update the weight and threshold values using the formulas in eqns. 17-20. Training of CVANN is stopped when the target goal (i.e., the minimum expected error rate) is achieved.

3. THE PROPOSED mRMR + CVANN DIAGNOSIS SYSTEM

This study proposes a novel hybrid method for PD diagnosis. A block diagram of the proposed system is presented in Figure 4. As shown in the block diagram, first, the attribute set containing the sound measurements of people (both healthy and PD) is presented to the system. Next, a normalization process is applied to the data to scale the variable valued between 0 and 1 to make the classification process unbiased towards any variable and to make the learning process more efficient. The min-max method, arguably the most preferred method, is adopted for variable normalization. Specifically, eqn. 22 was used to convert the variable values to the 0-1 value-range as per the min-max method.m

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{22}$$

In this equation, x' represents the normalized value; x_i represents the input value; x_{min} represents the smallest number within the input set; x_{max} represents the maximum number within the input set.

After the normalization process, the mRMR algorithm was used for the determination/selection of the most effective attributes. After being converted to complex numbers, the resulting attributes are presented to the CVANN as the input dataset.
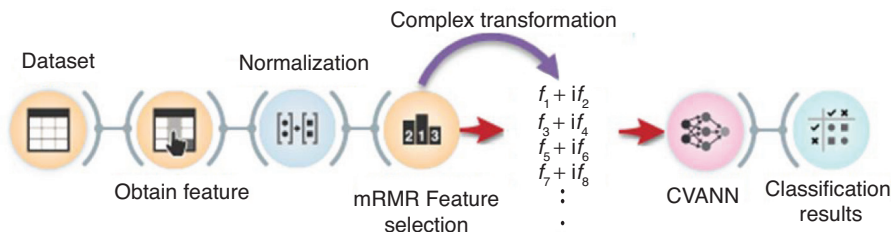

Figure 4.   A block diagram for the proposed PD diagnosis method.

4. EXPERIMENTAL DESIGN
4.1. Data Description
The voice/speech dataset used in this study was originally obtained by Max Little from the University of Oxford in cooperation with the National Voice and Speech Center, Denver, Colorado. The data consists of the speech samples (continued phonation test records) obtained from people with and without PD [59]. The complete dataset consisted of 195 biomedical sound measurements of 23 PD patients and eight non-PD/healthy people. The dataset also contained a status column defined as 1 for PD patients, and 0 for non-PD/healthy people. Table 1 presents the statistical values of all variables in the dataset [3, 4].

Table 1. Attributes in PD dataset.

Description	Feature Label	Minimum value	Maximum value	Average value	SD
Average vocal fundamental frequency (**Hz**)	f_1 = MDVP: Fo	88.33	260.11	154.22	41.39
Maximum vocal fundamental frequency (**Hz**)	f_2 = MDVP: Fhi	102.14	592.03	197.10	91.49
Minimum vocal fundamental frequency (**Hz**)	f_3 = MDVP: Flo	65.48	239.17	116.32	43.52
	f_4 = MDVP: Jitter (%)	0.0017	0.0332	0.0062	0.0048
Several measures of	f_5 = MDVP: Jitter (Abs)	0.00001	0.00026	0.00004	0.00003
variation in	f_6 = MDVP: RAP	0.0007	0.0214	0.0033	0.0030
fundamental frequency	f_7 = MDVP: PPQ	0.0009	0.0196	0.0034	0.0028
	f_8 = Jitter: DDP	0.0020	0.0643	0.0099	0.0089
	f_9 = MDVP: Shimmer	0.0095	0.1191	0.0297	0.0189
	f_{10} = MDVP: Shimmer (dB)	0.0850	1.3020	0.2822	0.1949
Several measures of	f_{11} = Shimmer: APQ 3	0.0045	0.0564	0.0156	0.0102
variation in amplitude	f_{12} = Shimmer: APQ 5	0.0057	0.0794	0.0178	0.0120
	f_{13} = MDVP: APQ	0.0072	0.1377	0.0240	0.0169
	f_{14} = Shimmer: DDA	0.0136	0.1694	0.0469	0.0305
Two measures of ratio of noise to tonal components	f_{15} = NHR	0.0007	0.3148	0.0248	0.0404
in the voice	f_{16} = HNR	8.441	33.047	21.885	4.426
Two non-linear dynamical	f_{17} = RPDE	0.2565	0.6851	0.4985	0.1039
complexity measures	f_{18} = D2	0.574	0.825	0.718	0.055
Signal fractal scaling exponent	f_{19} = DFA	−7.965	−2.434	−5.684	1.090
Three non-linear measures	f_{20} = Spread 1	0.0063	0.4504	0.2265	0.0834
of fundamental	f_{21} = Spread 2	1.4230	3.6711	2.3818	0.3827
frequency variation	f_{22} = PPE	0.0445	0.5273	0.2065	0.0901

4.2. Experimental Setup

All of the experiments were conducted within the MATLAB environment using a PC with Intel Core i7-2670 QM (2.2 GHz) microprocessor and 8 GB RAM. For all varieties of experiments, the selection of training and testing data samples was performed using a 10-fold cross-validation (CV) methodology. With 10-fold CV, the data splitting process is carried out as follows. First, the complete dataset is randomly divided into 10 disjoint subsets, of which, nine subsets are used for training while the remaining one subset is used for testing the trained prediction model. This process repeated 10 times; each time a different subset was used for testing while the remaining nine were used for training. The prediction results of all 10 trials are then combined to determine the true accuracy of the prediction model. Compared to simple split with one training and one test dataset, this CV methodology tends to provide less biased measure of accuracy with a certain degree of reliability and validity. In the dataset used for this study, there were 195 sound measurement samples. In each of the 10 folds, 174-175 samples were selected as training dataset and the remaining 19-20 samples were selected as test dataset. The process is repeated 10 times with different fold as the test set, and the results are collected and aggregated. The data in the training and testing datasets are also stratified for the output variable to maintain the proportional representation of PD and non-PD samples.

In order to assess the performance of the proposed method, four different statistical accuracy measures were evaluated: the accuracy (i.e., hit rate), sensitivity, specificity, F-measure and Kappa coefficient. Formulas for these parameters are shown in eqns. 23-27.

$$Accuracy\ (ACC) = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{23}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{24}$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \tag{25}$$

$$F\ measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{26}$$

where TP (true positive) is the number of PD patients that are accurately classified as PD, TN (true negative) is the number of non-PD patients that are accurately classified as non-PD, FN (false negative) is the number of PD patients that are inaccurately classified as non-PD, and FP (false negative) is the number of non-PD patients that are inaccurately classified as PD.

The F-measure, calculated based on the harmonic mean of the precision and recall, is often used as a complementary performance evaluation metric to assess classification methods. The F-measure takes numerical values in the range of 0 to 1, where F-measure values close to one denote the higher classification performance.

Kappa coefficient (KC) is another alternative to the ordinary classification performance metrics. Generally speaking, KC is used to measure the degree of consistency between two observers [60]. In the field of machine learning, this criterion is used to compare the accuracy of a classifier with the accuracy of a random classifier (i.e., random chance) [61]. This measure is algorithmically defined as:

$$KC = \frac{P_0 - P_c}{1 - P_c} \tag{27}$$

where P_0 is the accuracy of the classifier, and P_c is the accuracy obtained with random estimation/chance on the same dataset. Kappa statistics produces values in the range of -1 to 1, where values close to -1 indicate low level of consistency (higher rate of misclassification) while values close to 1 indicate high level of consistency (higher rate of accurate classification).

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the execution of our proposed methodology, the most effective features were determined by applying six different feature selection algorithms: mRMR, Fisher score, Chi-square, SFS, SFFS, and ReliefF. The best feature sets obtained using each of these selection algorithms are presented in Table 2.

Using the specified order, attribute values are converted into complex number format before being submitted to the classifier as an inputs. Accordingly, an input set was created by obtaining 1 complex value from 2 real values. Complex-valued attributes given as input to the CVANN are shown in Table 3. The example given in the table were prepared using the mRMR ranked attributes.
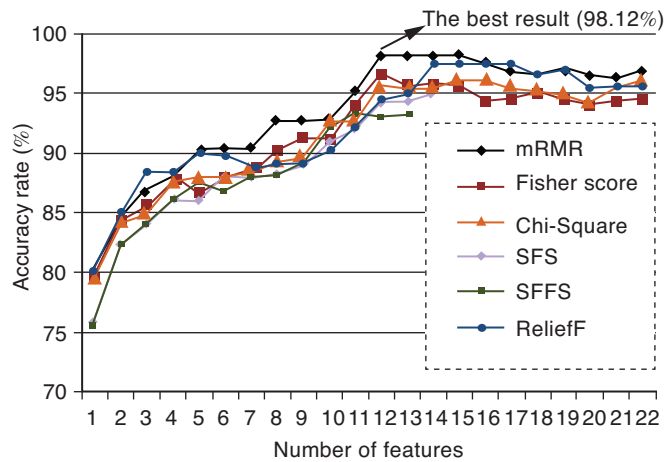
In order to achieve a high level of efficiency for the CVANN algorithm, the required values for all parameters were identified using an in-depth trial-and-error methodology. Specifically, the most effective parameter values were determined via a 10-fold

Table 2. Effective feature rankings obtained by different feature selection algorithms.

Feature Selection Algorithm	Feature Ranking
mRMR	$f_{20}, f_1, f_{11}, f_{18}, f_{19}, f_3, f_{21}, f_6, f_{22}, f_2, f_{17}, f_{15}, f_{13}, f_{16}, f_8, f_{12}, f_5, f_7, f_{10}, f_4, f_{14}, f_9.$
Fisher Score	$f_{20}, f_{22}, f_3, f_1, f_{21}, f_{16}, f_9, f_{18}, f_{13}, f_{10}, f_{12}, f_{11}, f_{14}, f_5, f_{17}, f_7, f_4, f_6, f_8, f_{19}, f_{15}, f_2.$
Chi-Square	$f_{20}, f_{22}, f_5, f_1, f_3, f_8, f_{13}, f_{21}, f_6, f_{12}, f_2, f_{15}, f_7, f_9, f_4, f_{11}, f_{10}, f_{14}, f_{16}, f_{17}, f_{18}, f_{19}.$
SFS	$f_{17}, f_{22}, f_{20}, f_1, f_4, f_5, f_8, f_{21}, f_6, f_2, f_{19}, f_{18}, f_7, f_3.$
SFFS	$f_{17}, f_{22}, f_5, f_{20}, f_1, f_4, f_{21}, f_8, f_3, f_6, f_7, f_{19}, f_2.$
ReliefF	$f_{20}, f_1, f_{22}, f_{11}, f_{19}, f_{14}, f_9, f_{12}, f_{17}, f_{21}, f_{10}, f_{16}, f_{18}, f_3, f_{13}, f_7, f_2, f_6, f_5, f_8, f_4, f_{15}.$

Table 3. Complex-valued attributes given as input to the CVANN.

Feature No	1	2	3	4	5	6	…	22
Feature	$f_{20}+i.0$	$f_{20}+if_1$	$f_{20}+if_1,$ $f_{11}+i.0$	$f_{20}+if_1$ $f_{11}+if_{18}$	$f_{20}+if_1$ $f_{11}+if_{18}$ $f_{19}+i.0$	$f_{20}+if_1$ $f_{11}+if_{18}$ $f_{19}+if_3$	…	$f_{20}+if_1$ $f_{11}+if_{18}$ $f_{19}+if_3$ $f_{21}+if_6$ $f_{22}+if_2$ $f_{17}+if_{15}$ $f_{13}+if_{16}$ f_8+if_{12} f_5+if_7 $f_{10}+if_4$ $f_{14}+if_9$

Figure 5. Changes in classification accuracy with the increase in the number of attributes selected using the feature selection algorithm.

cross-validation-based experimentation method from the training data, and these parameter values were used during the testing phase. For a fair comparison, the same set of parameter values were used in all experiments. Accordingly, an optimal network structure (input-hidden-output) was determined as [the number of input variables]-10-2 (representing the optimal number of neurons to use in input-hidden-output layers). The learning coefficient was determined as 0.9, and eqn. 21 was used as the stopping criterion. A complex sigmoid function was selected as the activation function.

Figure 5 illustrates the relationship between the number of features and the classification accuracy, showing that the classification accuracy increases with the

number of attributes selected/ranked by each of the six feature selection algorithms. The figure also shows that the classification accuracy stabilizes beyond 12 features, and among all six feature selection methods, mRMR produces the best prediction accuracy at the rate of 98.12% with only 12 features.

The best complex feature combinations obtained using the six different feature selection algorithms, and the best accuracy rates obtained with that feature combinations are given in Table 4. According to Figure 5 and Table 4, the best results were obtained using the mRMR + CVANN method. The ReliefF + mRMR hybrid method produced the second best results. In general, the lowest accuracy rates were produced with the feature sets obtained using the SFFS algorithm. It is somewhat surprising to see that SFFS produced lower accuracy rates compared to SFS as SFFS is presumably the improved version of SFS. According to the literature, depending on the dataset used and properties set, it is possible in rare cases where SFS performs better, while in most other cases, SFFS produces better accuracy results [62-65].

The results obtained in terms of the performance evaluation criteria mentioned above are presented in Table 5. Also included in this table are the results obtained using all the features presented to the classifier. In addition, the results obtained using ANN are also presented for a direct comparison with CVANN.

As shown in Table 5, the mRMR + CVANN model achieved the highest accuracy results of 98.12%, 99.24% and 98.96% in terms of Accuracy, sensitivity and specificity, respectively, and obtained the highest F-measure of 0.9905 and Kappa statistic value of 0.9896. Compared to the CVANN without feature selection, mRMR + CVANN improved the average performance by 3.77%, 3.91%, and 7.85% in terms of ACC, sensitivity and specificity, respectively. Also, the CVANN algorithm produced better results compared to the traditional ANN. When mRMR + CVANN and mRMR + ANN

Table 4. Feature rankings and complex combinations obtained using different feature selection algorithms.

Method	Best complex combination	Accuracy
mRMR +CVANN	$f_{20}+if_1,f_{11}+if_{18},f_{19}+if_3,f_{21}+if_6,f_{22}+if_2,f_{17}+if_{15}$	**98.12**
Fisher score +CVANN	$f_{20}+if_{22},f_3+if_1,f_{21}+if_{16},f_9+if_{18},f_{13}+if_{10},f_{12}+if_{11}$	96.50
Chi-square +CVANN	$f_{20}+if_{22},f_5+if_1,\ f_3+if_8,f_{13}+if_{21},f_6+if_{12},f_2+if_{15},f_7+if_9,f_4+if_{11}$	96.10
SFS +CVANN	$f_{17}+if_{22},f_{20}+if_1,f_4+if_5,f_8+if_{21},f_6+if_2,f_{19}+if_{18}$	95.05
SFFS +CVANN	$f_{17}+if_{22},f_5+if_{20},f_1+if_4,f_{21}+if_8,f_3+if_6,f_7+i.0$	93.25
ReliefF +CVANN	$f_{20}+if_1,f_{22}+if_{11},f_{19}+if_{14},f_9+if_{12},f_{17}+if_{21},f_{10}+if_{16},f_{18}+if_3$	97.55

Table 5. The results of the performance evaluation criteria.

Algorithm	Performance Metrics	All feature				After feature selection			
		Mean	SD	Max	Min	Mean	SD	Max	Min
	ACC	92.95	4.11	96.52	86.52	94.28	4.81	98.55	86.88
	Sensitivity	94.05	4.45	98.41	87.65	95.30	4.24	98.97	88.01
	Specificity	90.26	5.68	98.65	85.21	91.02	5.55	97.99	85.39
ANN	*F* measure	**0.9389**				**0.9513**			
	Kappa	**0.8560**				**0.8899**			
	Computation time (seconds)	8.9 s				7.5 s			
	ACC	94.4	4.1	99.7	87.9	98.12	0.81	100.00	97.81
	Sensitivity	95.3	3.7	99.2	88.6	99.24	0.93	100.00	97.32
mRMR +	Specificity	91.1	5.4	98.9	83.2	98.96	1.05	100.00	97.05
CVANN	*F* measure	**0.9599**				**0.9905**			
	Kappa	**0.8926**				**0.9896**			
	Computation time (seconds)	7.8 s				6.6 s			

are compared, mRMR + CVANN has improved the average performance by 3.84%, 3.94%, and 7.94% in terms of ACC, sensitivity, and specificity, respectively. Better results were also obtained with mRMR + CVANN for Kappa and F-measure values. The standard deviation of the mRMR + CVANN method was lower than that of ANN. This shows that the proposed method is more robust and more reliable than the other methods mentioned above.

Table 5 also shows that the feature selection method and the CVANN produces better results in terms of computation time requirement. After the application of the feature selection method, the computing time decreases. As a result, the proposed method is deemed to be a fast, accurate and reliable prediction method for this application domain.

The classification accuracy rates obtained in this study and in the previous studies on the same dataset are compared, and the results are exhibited in Table 6. Only the studies that used the same dataset were compared for a reliable and fair comparison. As shown in Table 6, previous prediction methods provided fairly good results, with accuracy levels ranging between 80% and 97%. Our proposed method produced a better prediction performance with 98.12% overall accuracy on text dataset, compared to previous studies. The two methods with accuracy results closest to the present method were those proposed in [23] and [25]. Polat et al. [23] adopted 50%-50% training-test data selection for cross validation. They obtained a classification accuracy of 97.93% for the diagnosis of PD. In order to perform a fair comparison with their method, the proposed method of the current study was re-run with 50% -50% training-test data split. In this re-run, the classification accuracy of our method came out as 98.25%, which is slightly better (by 0.32%) than Polat et al.

Table 6. A comparative analysis of the present and published studies for PD diagnosis.

Study	Method	Data Selection Method	Accuracy
Shahbaba and Neal [16]	Dirichlet process mixtures	5-fold CV	87.70
Das [17]	ANN	hold out	92.90
Guo et al. [18]	GA-EM	10-fold CV	93.10
Luukka [19]	Fuzzy entropy measures + Similarity	hold out	85.03
Li et al. [20]	Fuzzy-based non-linear transformation + SVM	hold out	93.47
Aström and Koker [21]	Parallel ANN	hold out	91.20
Spadoto et al. [22]	PSO + OPF Harmony search + OPF Gravitational search + OPF	hold out	84.01
Polat et al. [23]	FCMFW + KNN	50-50%training-testing	97.93
Zuo et al. [25]	PSO+ Fuzzy k-nearest neighbor	10-fold CV	97.47
Sakar and Kursun [26]	Mutual information + SVM	bootstrap with 50 replicates	92.75
Ozcift et al. [27]	CFS-RF	10-fold CV	87.13
Chen et al. [28]	PCA-FKNN	10-fold CV	96.07
Psorakis et al. [67]	Improved multiclass multi-kernal relevance vector machines	10-fold CV	89.47
Proposed Method	mRMR+CVANN	10-fold CV	98.12

In previous studies, we observed that CVANN produces higher accuracy dataset compared to traditional real-valued ANN applied to the same problem and the same dataset [50]. Especially for the systems that naturally work with complex values, CVANN provides significantly better prediction results [44, 66]. For important applications such as critical diagnostics and diagnostic systems in medicine, even a slight increase in accuracy rate makes a significant difference. The present study further corroborates that CVANN is a viable (and perhaps superior) alternative tool for building and deploying highly accurate medical diagnostic systems.

There are a number of possible reasons behind the success of CVANN, such as the following:

- Mapping capability of CVANN: A neuron has two main functions to perform: an aggregation function and an activation function. The aggregation function maps a multidimensional input space into the neuron's net-input space, which is one dimensional for a real-valued network and two dimensional for a complex-valued network [68]. The activation function allocates net input space into discrete clusters that represent different classes using a threshold operation on the output provided by the activation function collector. In the mapping by the aggregator, each input is multiplied by a connection weight and then the resulted weighted

inputs are added. If we consider $\Im_R$ as the set of all possible mappings for an real-valued network and $\Im_C$ as the set of all possible mappings for a complex-valued networks, it can be seen that $\Im_R \subset \Im_C$. This is because a complex multiplication scales and rotates an input with any optional amount, whereas a real multiplication does a scaling with an optional amount but a rotation of only 0 or π [68]. In other words, the mapping capabilities of a complex-valued network is superior to a real-valued network, and this may be one of the main reasons for its superior performance.

- High functionality is the ability of a single neuron to learn linearly inseparable input/output mappings. Thus, a neuron has the ability to learn these mappings in the initial stage before producing a higher level of input, and transforming to a higher dimensional space, respectively. Studies showed that a single neuron with complex-valued weights can solve linearly inseparable problems such as the exclusive or (XOR) classification problem. This ability suggests that a single CVANN has a higher functionality than a single ANN [52].

- In ANNs, input variables are single values (i.e., real numbers), while in CVANNs, input variables are complex values (complex numbers consisting of real and imaginary parts). Therefore, in CVANN, two-dimensional data inputs are possible. As described in Section 2.2, this multi-dimensional data representation and complex multiplication operations may be among the main factors that improves the accuracy and thus increasing popularity of CVANN.

In summary, the main reason for CVANN to achieve better diagnosis performance than its traditional counterparts is its superior mapping capabilities coupled with efficacy in high functionality.

5.1. Limitations and Future Research Directions

As is the case in any developmental research, there are some limitations to the proposed method. First, the usability of the method needs to be improved. In order to use the developed software system, domain experts may need an intuitive, somewhat automated graphical visual interface. In the near future before real-world deployment, we plan to develop a graphical user interface that encapsulates the prediction models and improves user-friendliness. Additionally, in order to increase the applicability, the program's processing time (time to build and calibrate the prediction models and to use the model for diagnosis purposes) needs to be reduced and the efficiency be increased. Experimental results showed that the computation time of the proposed method is longer than desired. A significant improvement in the computational efficiency is possible and is part of our near future development plans. A program capable of processing user's requests at the level of milliseconds will increase the usefulness and adaptation of the prediction system.

In this study, 22 features are selected in the feature selection process. Additional features including socio-demographic and medical/diagnostic characteristics may have significant impacts on accurate diagnosis of PD. A more comprehensive study with a significantly extended feature set is among the future development paths of our research efforts. A database with limited number of features frequently used in the extant

literature is preferred in this study in order to compare and contrast the results of the study with those presented by previous studies. Even though the size of this dataset is small, we propose to apply the same method to a more comprehensive/extensive dataset to develop more robust prediction models.

ANN, generally considered a popular member of the family of black box models, is a biologically inspired mathematical method capable of generating solutions based on historical cases, i.e., previously recorded input and output data. Even though it creates highly efficient models, it is not capable of explaining its inner structure (how and what it does). In other words, ANN cannot explain how the inputs are used to generate results. This black box designation applies not only to ANN but also to CVANN. A variety of research streams are dedicated to shed light to the black box, that is, to better understand the internal structure of the prediction system. Among these streams, sensitivity analysis has received significant interest, where the input variables of a trained neural network is perturbed one variable at a time and the impact of this perturbation is recorded and translated/transformed to a rank ordered variable importance measure for all input variables.

6. CONCLUSION

In this study, we proposed a new approach for accurately diagnosing PD that can help medical personnel to make better and faster decisions. The proposed approach is capable of automatically analyzing data related to PD to develop prediction/diagnostic models with a high degree of accuracy in a relatively short time. The main novelty of the proposed study relates to the use of a hybrid methodology herein referred to as mRMR + CVANN, which integrates an effective feature selection method and a strong classifier. In this methodology, an effective feature set was obtained using an mRMR algorithm. Application of this algorithm resulted in a smaller feature set by eliminating less relevant features. Complex numbered features were then obtained from the optimally selected/reduced feature set. The complex-valued feature combinations produced and used in this study are among the most important contributions/innovations of the proposed method. A CVANN algorithm with high functionality and a very good classification capability was designed and developed during the classification stage of the proposed method. The prediction results obtained were very promising. Thus, a prediction system that can be used as a part of a computer-aided diagnosis system was developed. This system has the capability and potential to help doctors and other medical professionals in the diagnostic related decision processes for different diseases.

CONFLICT OF INTEREST
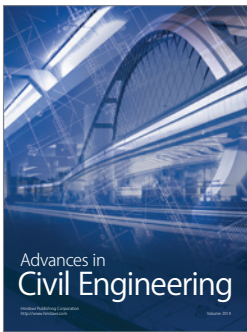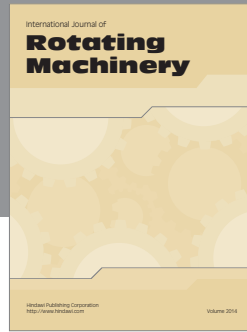The authors indicated no potential conflicts of interest.

REFERENCES
[1]   Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S. Apaydin H, Kursun O. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics,* 2013, 17(4):828–834.

[2] Moore DJ, West AB, Dawson VL, Dawson TM. Molecular pathology of Parkinson's disease. *Annual Review of Neuroscience*, 2005, 28:57–87.

[3] Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA. Risk tables for Parkinsonism and Parkinson's disease. *Journal of clinical epidemiology*, 2002, 55:25–31.

[4] Little MA, McSharry PE, Hunter EJ, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 2009, 56:1015–1022.

[5] Ramaker C, Marinus J, Stiggelbout AM, van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders,* 2002, 17(5):867–876.

[6] Umphred DA, Lazaro RT, Roller M, Burton G (2013). *Neurological rehabilitation*. Elsevier Health Sciences.

[7] Das R, Sengur A. Evaluation of ensemble methods for diagnosing of valvular heart disease. *Expert Systems with Applications*, 2010, 37(7):5110–5115.

[8] Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications*, 2009, 36(4):7675–7680.

[9] Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 1969, 12:246–269.

[10] Gamboa J, Jimenez-Jimenez FJ, Nieto A, Montojo J, Orti-Pareja M, Molina JA, García-Albea E, Cobeta I. Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs. *Journal of Voice*, 1997, 11:314–320.

[11] Ho A, Bradshaw JL, Iansek R. For better or for worse: the effect of Levodopa on speech in Parkinson's disease. *Movement Disorders*, 2008, 23(4):574–580.

[12] Harel B, Cannizzaro M, Snyder PJ. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and Cognition*, 2004, 56:24–29.

[13] Skodda S, Rinsche H, Schlegel U. Progression of dysprosody in Parkinson's disease over time - A longitudinal study. *Movement Disorders*, 2009, 24(5):716–722.

[14] Sapir S, Ramig L, Spielman J, Fox C. Formant centralization ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech Language and Hearing Research*. 2010, 53:114–125.

[15] Cnockaert L, Schoentgen J, Auzou P, Ozsancak C, Defebve L, Grenez F. Low frequency vocal modulations in vowels produced by Parkinsonian subjects. *Speech Communication*, 2008, 50:288–300.

[16] Shahbaba B, Neal R. Nonlinear models using Dirichlet process mixtures. *Journal of Machine Learning Research*, 2009, 10:1829–1850.

[17] Das R. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications,* 2010, 37(2):1568–1572.

[18] Guo PF, Bhattacharya P, Kharma N. Advances in detecting Parkinson's disease. *Medical Biometrics*, 2010, 6165:306–314.

[19] Luukka P. Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, 2011, 38(4): 4600–4607.

[20] Li DC, Liu CW, Hu SC. A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artificial Intelligence in Medicine*, 2011, 52(1):45–52.

[21] Aström F, Koker R. A parallel neural network approach to prediction of Parkinson's disease. *Expert Systems with Applications*, 2011, 38(10), 12470–12474.

[22] Spadoto AA, Guido RC, Carnevali FL, Pagnin AF, Falcao AX, Papa JP. 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), 2011, 7857–7860.

[23] Polat K. Classification of Parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering. *International Journal of Systems Science*, 2012, 43(4):597–609.

[24]   Daliri MR. Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. *Biomedical Signal Processing and Control*, 2013, 8(1):66–70.

[25]   Zuo WL, Wang ZY, Liu T, Chen HL. Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. *Biomedical Signal Processing and Control*, 2013, 8(4):364–373.

[26]   Sakar CO, Kursun O. Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems*, 2010, 34(4):591–599.

[27]   Ozcift A, Gulten A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Computer Methods and Programs in Biomedicine*, 2011, 104(3):443–451.

[28]   Chen HL, Huang CC, Yu XG, Xu X, Sun X, Wang G, Wang SJ. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Systems with Applications*, 2013, 40(1):263–271.

[29]   Rouzbahani HK, Daliri MR. Diagnosis of Parkinson's disease in human using voice signals. *Basic and Clinical Neuroscience*, 2011, 2(3):12–20.

[30]   Ma C, Ouyang J, Chen HL, Zhao XH. An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. *Computational and Mathematical Methods in Medicine*, 2014, doi:10.1155/2014/985789.

[31]   Khorasani A, Daliri MR. HMM for classification of Parkinson's disease based on the raw gait data. *Journal of Medical Systems*, 2014, 38(12):1–6.

[32]   Daliri MR. Automatic diagnosis of neuro-degenerative diseases using gait dynamics. *Measurement*, 2012, 45(7):1729–1734.

[33]   Cao B, Shen D, Sun JT, Yang Q, Chen Z. Feature selection in a kernel space. *24th Annual International Conference on Machine Learning*, 2007, 121–128.

[34]   Hijazi H, Christina C. A classification framework applied to cancer gene expression profiles. *Journal of healthcare engineering,* 2013, 4(2):255–284.

[35]   Daliri MR. Feature selection using binary particle swarm optimization and support vector machines for medical diagnosis. *Biomedizinische Technik/Biomedical Engineering*, 2012, 57(5):395–402.

[36]   Ding C, Peng HC. Minimum redundancy feature selection from microarray gene expression data. *Second IEEE Computational Systems Bioinformatics Conference*, 2003, 523–528.

[37]   Sen B, Peker M, Celebi FV, Cavusoglu A. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *Journal of Medical Systems*, 2014, 38(3):1–21.

[38]   França MV, Zaverucha G, Garcez ASDA. Fast relational learning using bottom clause propositionalization with artificial neural networks. *Machine learning*, 2014, 94(1):81–104.

[39]   Duda RO, Hart PE, Stork DG. *Pattern classification*, 2nd ed., John Wiley & Sons, 2012.

[40]   Jin X, Xu A, Bie R, Guo P. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. *Lecture Notes in Computer Science*, 2006, 3916:106–115.

[41]   Whitney AW. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 1971, 20(9):1100–1103.

[42]   Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. *Pattern Recognition Letters*, 1994, 15(11):1119–1125.

[43]   Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 2003, 53(1-2):23–69.

[44]   Ceylan M, Ceylan R, Ozbay Y, Kara S. Application of complex discrete wavelet transform in classification of Doppler signals using complex-valued artificial neural network. *Artificial Intelligence in Medicine*. 2008, 44(1), 65–76.

[45]   Peker M, Sen B. A new complex-valued intelligent system for automated epilepsy diagnosis using EEG signals. AWERProcedia Information Technology & Computer Science. 2013, 1121–1128.

[46] Ceylan R, Ceylan M, Ozbay Y, Kara S. Fuzzy clustering complex-valued neural network to diagnose cirrhosis disease. *Expert Systems with Applications*. 2011, 38(8): 9744–9751.

[47] Sivachitra M, et al. A fully complex-valued fast learning classifier (FC-FLC) for real-valued classification problems. *Neurocomputing*. 2015, 149:198–206.

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

[49] Karthikeyan B, Gopal S, Vimala M. Conception of complex probabilistic neural network system for classification of partial discharge patterns using multifarious inputs. *Expert Systems with Applications*, 2005, 29(4):953–963.

[50] Peker M, Sen B, Delen D. A novel method for automated diagnosis of epilepsy using complex-valued classifiers. *IEEE Journal of Biomedical and Health Informatics*, 2015, doi: 10.1109/JBHI.2014.23877952015.

[51] Chen X, Tang Z, Variappan C, Li S, Okada T. A modified error backpropagation algorithm for complex-value neural networks. *International Journal of Neural Systems*, 2005, 15(06):435–443.

[52] Aizenberg I. *Complex-valued neural networks with multi-valued neurons*. Heidelberg: Springer, 2011, 264–265.

[53] Nitta T. Orthogonality of decision boundaries in complex-valued neural networks. *Neural Computation*, 2004, 16(1):73–97.

[54] Zimmermann HG, Minin A, Kusherbaeva V. Comparison of the complex valued and real valued neural networks trained with gradient descent and random search algorithms. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2011, 2013–2018.

[55] Benvenuto N, Piazza F. On the complex back propagation algorithm. *IEEE Transactions on Signal Processing*, 1992, 40:967–969.

[56] Nitta T. An extension of the back-propagation algorithm to complex numbers. *Neural Network*, 1997, 10:1391–1415.

[57] Ozbay Y, Kara S, Latifoglu F, Ceylan R, Ceylan M. Complex-valued wavelet artificial neural network for doppler signals classifying. *Artificial Intelligence in Medicine*, 2007, 40(2):143–156.

[58] Nitta T. A back-propagation algorithm for complex numbered neural networks. *Proceedings of 1993 International Joint Conference on Neural Networks*, 1993, 1649–1652.

[59] Parkinsons Data Set, https://archive.ics.uci.edu/ml/datasets/Parkinsons. Accessed September 5, 2014.

[60] Cohen J, A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960, 20(1):37–46.

[61] Zhang Y, Zhang X, Liu W, Luo Y, Yu E, Zou K, Liu X. Automatic sleep staging using multi-dimensional feature extraction and multi-kernel fuzzy support vector machine. *Journal of Healthcare Engineering*, 2014, 5(4):505–520.

[62] Forsati R, Moayedikia A, Jensen R, Shamsfard M, Meybodi MR. Enriched ant colony optimization and its application in feature selection. *Neurocomputing*, 2014, 142:354–371.

[63] Chuang LY, Yang CH, Li JC. Chaotic maps based on binary particle swarm optimization for feature selection. *Applied Soft Computing*, 2011, 11(1): 239–248.

[64] Lian C, Ruan S, Denœux T. An evidential classifier based on feature selection and two-step classification strategy. *Pattern Recognition*, 2015, 48(7):2318–2327.

[65] Moustakidis SP, Theocharis JB. SVM-FuzCoC: A novel SVM-based feature selection method using a fuzzy complementary criterion. *Pattern Recognition*, 2010, 43(11): 3712–3729.

[66] Hänsch R, Hellwich O. Classification of polarimetric SAR data by complex valued neural networks. *In: ISPRS Workshop High-Resolution Earth Imaging for Geospatial Information*, 2009, 4–7.

[67] Psorakis I, Damoulas T, Girolami MA. Multiclass relevance vector machines: sparsity and accuracy. IEEE Transactions on Neural Networks, 2010, 21(10):1588–1598.

[68] Amin MF, Islam MM, Murase K. Ensemble of single-layered complex-valued neural networks for classification tasks. *Neurocomputing*, 2009, 72(10):2227–2234.