

Research Article

An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel-Extreme Learning Machine

Derya Avci¹ and Akif Dogantekin²

¹Department of Electrical and Electronic Engineering, Engineering Faculty, Firat University, 23119 Elazig, Turkey

²Emek Hospital, Gaziantep, Turkey

Correspondence should be addressed to Derya Avci; derya2344@hotmail.com

Received 5 January 2016; Revised 24 March 2016; Accepted 29 March 2016

Academic Editor: Fabrizio Stocchi

Copyright © 2016 D. Avci and A. Dogantekin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Parkinson disease is a major public health problem all around the world. This paper proposes an expert disease diagnosis system for Parkinson disease based on genetic algorithm- (GA-) wavelet kernel- (WK-) Extreme Learning Machines (ELM). The classifier used in this paper is single layer neural network (SLNN) and it is trained by the ELM learning method. The Parkinson disease datasets are obtained from the UCI machine learning database. In wavelet kernel-Extreme Learning Machine (WK-ELM) structure, there are three adjustable parameters of wavelet kernel. These parameters and the numbers of hidden neurons play a major role in the performance of ELM. In this study, the optimum values of these parameters and the numbers of hidden neurons of ELM were obtained by using a genetic algorithm (GA). The performance of the proposed GA-WK-ELM method is evaluated using statical methods such as classification accuracy, sensitivity and specificity analysis, and ROC curves. The calculated highest classification accuracy of the proposed GA-WK-ELM method is found as 96.81%.

1. Introduction

Parkinson disease (PD) is a degenerative disorder of the central nervous system. It results from the death of dopamine generating cells in the substantia nigra, a region of the mid-brain. This disease affects about 1% of the world population over the age of 55 [1, 2].

In advanced stages of the disease, nonmotor features, such as dementia and dysautonomia, occur frequently [3]. PD is diagnosed in case of presence of two or more cardinal motor features such as rest tremor, bradykinesia, or rigidity [4]. Functional neuroimaging holds the promise of improved diagnosis and allows assessment in early disease [5].

The main symptoms of PD are bradykinesia, tremor, rigidity, and postural instability. When all of these symptoms are seen, then the person can be diagnosed by doctors with Parkinson disease. The dysphonia is considered to be one of the most difficult aspects of Parkinson disease by many patients and their families. Nearly 9 out of 10 people with PD have a speech or voice disorder. Dysphonic symptoms typically include reduced loudness, breathiness, roughness,

decreased energy in the higher parts of the harmonic spectrum, and exaggerated vocal tremor and these symptoms can be detected using many different vocal tests [6]. In the design of an automatic diagnosis system for PD, it is more suitable to use the voice data because it is one of the most common symptoms. In literature, there are many studies on speech measurement for general voice disorders [6–12]. In these studies, the speech signals are recorded and then these signals are detected by means of different methods certain properties of these signals. Then, a classifier is used to diagnose patients with PD from certain properties of signal. The classifier is the heart of the automatic diagnosis system. The reliable classifier should diagnose the disease at as high accuracy as possible even though there are many uncontrolled variations. In literature, different classifiers have been proposed for automatic diagnosis of PD. The NNs and adaptive neurofuzzy classifier with linguistic hedges (ANFIS-LH) are investigated for automatic diagnosis of PD in [13]. The performance of probabilistic neural network (PNN) for automatic diagnosis of PD is evaluated in [14]. SVM classifier is also investigated for the same goal in [15]. NNs have some drawbacks such

as need of long training time and uncertainties in activation function to be used in hidden layer, number of cells in hidden layer, and the number of hidden layer. In case of SVM, type of kernel function and penalty constant and so forth affects the classification performance. If these parameters are not appropriately selected, the classification performance of SVM degrades. Similarly, the performance of ANFIS depends on type and parameters of membership function and output linear parameters.

Among these classifiers, NNs have been widely used in pattern recognition and regression. The NNs are commonly trained by backpropagation based on a gradient-based learning rule [16]. Up to now, the gradient-based learning methods have been widely applied for learning of NNs [17, 18]. However, they have several shortcomings such as difficult setting of learning parameters, slow convergence, training failures due to local minima, and repetitive learning to improve performance of NNs. Also, it is clear that gradient descent-based learning methods are generally very slow [20].

Recently, the Extreme Learning Machine (ELM) proposed by Huang et al. has been widely used in classification regression problems because of its properties fast learning capability and fast learning. Although output weights are analytically calculated, there is no rule in determination of number of hidden neurons, activation function. The ELM may not provide high classification performance because of cases mentioned above.

In [6], the GA is used for selection of feature subset for the input of ANN. The proposed method is not suitable for real-time implementation. Besides, the feature vector is randomly reduced to a lower dimension in [3, 6, 7, 9]. ANFIS structure might not have a good performance if a huge amount of data exists.

Recently, a new learning algorithm called Extreme Learning Machine (ELM) which randomly selected all the hidden nodes parameters of generalized single-hidden layer feed-forward networks (SLFNs) and analytically determines the output weights of SLFNs is proposed in [18–20]. Although output weights are analytically calculated, there is no rule in determination of number of hidden neurons and type of the kernel function. To obtain a good classification performance of ELM, these parameters should be determined properly.

This paper proposes an expert Parkinson diseases (PD) diagnosis system based on genetic algorithm- (GA-) wavelet kernel- (WK-) Extreme Learning Machines (ELM). The classifier used in this paper is single layer neural network (SLNN) and it is trained by the ELM learning method. In wavelet kernel-Extreme Learning Machine (WK-ELM) structure, there are three adjustable parameters of wavelet kernel. These parameters and the numbers of hidden neurons play a major role in the performance of ELM. Therefore, values of these parameters and numbers of hidden neurons should be tuned carefully based on the solved problem. In this study, the optimum values of these parameters and the numbers of hidden neurons of ELM were obtained by using a genetic algorithm (GA). The hepatitis disease datasets are obtained from the UCI machine learning database. The performance of the proposed GA-WK-ELM method is evaluated through statical methods such as classification accuracy,

sensitivity and specificity analysis, and ROC curves. In here, the numbers of hidden neurons of ELM and parameters of wavelet kernel function are optimized by GA. In GA structure, an individual is composed of a total of 20 bits. These are as follows:

- (i) The first four bits (1st, 2nd, 3rd, and 4th bits) of each of these individuals represent the w parameter values (between 1 and 16) of the wavelet kernel functions.
- (ii) The second four bits (5th, 6th, 7th, and 8th bits) of each of these individuals represent the x parameter values (between 1 and 16) of the wavelet kernel functions.
- (iii) The third four bits (9th, 10th, 11th, and 12th bits) of each of these individuals represent the y parameter values (between 1 and 16) of the wavelet kernel functions.
- (iv) The rest of the 20 bits represent the number of hidden neurons (between 5 and 260).

The 40 number of these individuals is randomly chosen for the initial population. Thus, it is purposed to obtain the best possible performance from ELM classifier. The training and testing dataset for the proposed method is obtained from the UCI dataset. This dataset is composed of 192 pieces of data. The randomly selected 128 of 192 pieces of data are used for training of classifier whereas the remaining data is used for testing of classifier. For different kernel function and number of hidden neurons, the results of the proposed method are given. Further, a comparison is performed with previous studies to show the validity of the proposed method. From results, the proposed method is a quite powerful tool for automatic diagnosis of hepatitis and may work in real-time systems.

The paper is organized as follows. Section 2 presents pattern recognition for the diagnosis of Parkinson disease. In Section 3, wavelet kernel-Extreme Learning Machines and in Section 4 genetic algorithms are briefly presented, respectively. In Section 5, application of GA-WK-ELM for the diagnosis of Parkinson disease is explained. The obtained results are given in Section 6. Finally, Section 7 provides the discussion and conclusion of this study.

2. Pattern Recognition for Diagnosis of Parkinson Disease

The pattern recognition for diagnosis of disease is commonly composed of two stages. They are feature extraction and classification stages. In the feature extraction stage, the useful information of data is extracted by a feature extractor. The feature extraction not only reduces the computational burden of the classifier but also improves classification performance. In classification stage, extracted features from data are given as input to the classifier. Depending on classification problem, the data is separated into two or more classes by the classifier.

The pattern recognition concept used in this study is given in Figure 1. The proposed concept consists of three

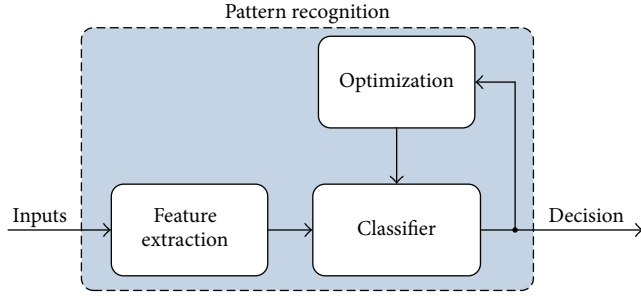


FIGURE 1: The block diagram of the pattern diagnosis concept.

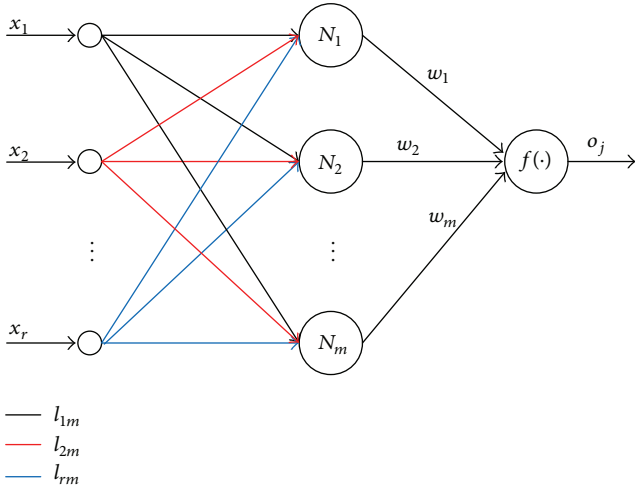


FIGURE 2: The structure of a single-hidden layer feedforward network.

stages including feature extraction, classification, and optimization of classifier's parameters. These stages are explained in detail as follows.

3. Wavelet Kernel-Extreme Learning Machines

In literature, the neural networks have been commonly used in pattern recognition and regression problems [20, 21]. The gradient-based learning and backpropagation algorithms are most commonly used methods for neural networks [17, 18]. Moreover, these methods have some drawbacks such as difficult setting of learning parameters, slow convergence, slow learning, and training failures [19].

Because of these disadvantages of classic gradient-based learning and backpropagation neural network algorithms, the Extreme Learning Machine (ELM) algorithm is proposed by Cho et al. [19]. In the Extreme Machine Learning algorithm, the output weights of a single-hidden layer feedforward network (SLFN) are analytically calculated by using the Moore-Penrose (MP) generalized inverse instead of iterative learning scheme [20]. In Figure 2, the structure of a single-hidden layer feedforward network using Extreme Learning Machine algorithm is given. In here, l_{1m} , l_{2m} , and l_{rm} are weights vector connecting the k th hidden neuron and the input neurons, w is the weight vector connecting the k th

hidden neuron and output neuron, and $f(\cdot)$ is the activation function.

The most significant features of ELM are ordered as below:

- (i) In ELM structure, the learning speed is very fast. Because of this, single-hidden layer feedforward network can be trained by using ELM. Thus, an ELM learning method, which is faster than other classical learning methods, is obtained.
- (ii) The obtaining of the less training error and the fewer norms of weights are aimed at by using ELM, because the ELM learning algorithm has good performance for neural networks.
- (iii) In the structure of single-hidden layer feedforward network, the ELM learning algorithm is used with nondifferentiable activation functions.
- (iv) The easy solutions are tried to get in the ELM structure [19].

The outputs of an ELM with m neurons and f activation function are given as below:

$$o_j = \sum_{i=1}^m \beta_i f(l_i x_r + b_i). \tag{1}$$

The ELM learning algorithm has faster learning speed than classic neural networks. Moreover, it has better generalization performance than them. Nowadays, the number of researchers, who work in ELM science topic, has increased [19–23]. In ELM learning algorithm, the initial parameters of the hidden layer need not be tuned. In ELM algorithm, all nonlinear piecewise continuous functions are used as the hidden neurons. Therefore, for M optional various samples $\{(r_j, m_j) \mid r_j \in Q^l, m_j \in Q^k, j = 1, \dots, M\}$, the output function in ELM by using K hidden neurons is

$$u_K(r) = \sum_{j=1}^K S_j v_j(r) = v(r) S, \tag{2}$$

where $v = [v_1(r), v_2(r), \dots, v_K(r)]$ is the output vector of the hidden layer with respect to the input r . $S = [S_1, S_2, \dots, S_K]$ is the vector of the output weights between the hidden layer of K neurons and the output neuron. v vector changes the data from input space to the ELM feature space [19–23]. The training error and the output weights should be synchronously minimized for decreasing the training error in ELM algorithm. So, generalization performance of neural networks increases:

$$\begin{aligned} \text{minimize} \quad & \|AS - C\|, \\ & \|S\|. \end{aligned} \tag{3}$$

In here, (3) can be solved by using

$$S = A^T \left(\frac{1}{E} + AA^T \right)^{-1} C, \tag{4}$$

where E is the regulation coefficient, A is the hidden layer output matrix, and C is the expected output matrix of samples, respectively. So, the output function of the ELM learning algorithm can be given as follows:

$$u(r) = v(r) A^T \left(\frac{1}{E} + AA^T \right)^{-1} C. \quad (5)$$

If the feature vector $v(r)$ is unknown, the kernel matrix of ELM based on Mercer's conditions can be computed as follows:

$$D = AA^T : k_{jz} = v(r_j) v(r_z) = b(r_j, r_z). \quad (6)$$

In this way, the output function $u(r)$ of the wavelet kernel-Extreme Learning Machine (WK-ELM) can be given as below:

$$u(r) = [b(r, r_1), \dots, b(r, r_M)] \left(\frac{1}{E} + D \right)^{-1} C. \quad (7)$$

In there, $D = AA^T$ and $b(r, g)$ is the kernel function of Extreme Learning Machine. There are some kernel functions, which are linear kernel, polynomial kernel, Gaussian kernel, and exponential kernel, appropriate for the Mercer condition in ELM literature. The readers can find more details in [21, 22]. In this study, wavelet kernel function is used for simulation and performance analysis of WK-ELM:

$$b(r, g) = \cos \left(w \frac{\|r - g\|}{x} \right) \exp \left(-\frac{\|r - g\|^2}{y} \right). \quad (8)$$

In the result of these application studies, it was observed that the training and testing performance of the wavelet kernel function shown in (8) is better than the performances of linear kernel, polynomial kernel, Gaussian kernel, and exponential classical kernel functions, respectively. The values of adjustable parameters w , x , and y are important for training performance of ELM. Because of this, values of these parameters should be attentively tuned for solving the problem. However, the hidden layer feature mapping need not be known and the number of hidden neurons need not be chosen in WK-ELM algorithms. Moreover, the WK-ELM learning algorithm has better generalization performance than classic ELM learning algorithm. At the same time, it was shown that WK-ELM is more stable than classic ELM and is faster than Support Vector Machine (SVM) [24].

4. Genetic Algorithms

To solve a problem, an evolutionary process is used in the structures of genetic algorithms [25]. A genetic algorithm begins with a set of solutions which are represented by individuals. This set of solutions is known as a population. Each population is a solution set and new solutions are selected according to their fitness values. In the genetic algorithm, the iterative process is repeated as long as the new population is better than the old one. The higher the fitness value of an individual is, the more likely this individual is reproduced for the next population. The iterative process is

finished when some conditions (e.g., number of individuals in the population) are satisfied [26].

The stages of the genetic algorithm are given as below.

Stage 1. A random population of n individuals is created. These individuals are a suitable solution to the problem. In here, the value of n is 20.

Stage 2. The fitness $f(x)$ of each individual x is calculated in the population [25]. In these experimental studies, each of the individuals in the population is randomly formed.

Stage 3. Two parental individuals from among the individuals are selected. These individuals have the higher fitness value in the population. Then, the crossover operator is realized to these parental individuals. The aim of the crossover operator is creating the varied individuals. These have higher fitness values than former individuals.

Stage 4. In this stage, a crossover probability is used for crossover operating to form the new individuals. If crossover is not done, the individual will be the exact copy of the parents.

Stage 5. In this stage, each new individual is obtained by mutating with a mutation probability. This mutation process is realized by using any one or more bits of the individual.

Stage 6. In this stage, the new individuals are obtained from the new population.

Stage 7. In this stage, if the end conditions are satisfied, the genetic algorithm is stopped. It is returned to the best solution in the current population.

Stage 8. In this stage, it is returned to Stage 2. Then, the new generated population is used for further algorithm.

5. Application of GA-WK-ELM for Diagnosis of Parkinson Disease

The Parkinson dataset used in this study is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson disease (PD), and it includes a total of 192 voice recordings from individuals. In addition, these biomedical voice measurements have different attribute information given in Table 1 [6–12].

The essential aim of processing the data is to discriminate healthy people from those with PD, according to the "status" attribute which is set to 1 for healthy people and 0 for people with PD, which is a two-decision classification problem.

The block diagram of the proposed method is given in Figure 3. As shown in the figure, the feature vector obtained from the PD dataset is applied to WK-ELM optimized with GA. The Parkinson dataset used in this study is taken from the University of California at Irvine (UCI) machine learning repository [6–12]. It was used for training and testing of the proposed GA-WK-ELM method.

TABLE 1: The attributes of biomedical voice measurements.

Number	Attributes	Explanation
1	MDVP:Fo (Hz)	Average vocal fundamental frequency
2	MDVP:Fhi (Hz)	Maximum vocal fundamental frequency
3	MDVP:Flo (Hz)	Minimum vocal fundamental frequency
4	MDVP:Jitter (%)	Several measures of variation in fundamental frequency
5	MDVP:Jitter (abs)	
6	MDVP:RAP	
7	MDVP:PPQ	
8	Jitter:DDP	Several measures of variation in amplitude
9	MDVP:Shimmer	
10	MDVP:Shimmer (dB)	
11	Shimmer:APQ3	
12	Shimmer:APQ5	Two nonlinear dynamical complexity measures
13	MDVP:APQ	
14	Shimmer:DDA	
15	RPDE	
16	D2	The measure of ratio of noise to tonal components in the voice status
17	NHR	
18	HNR	
19	DFA	
20	spread1	Three nonlinear measures of fundamental frequency variation
21	spread2	
22	PPE	

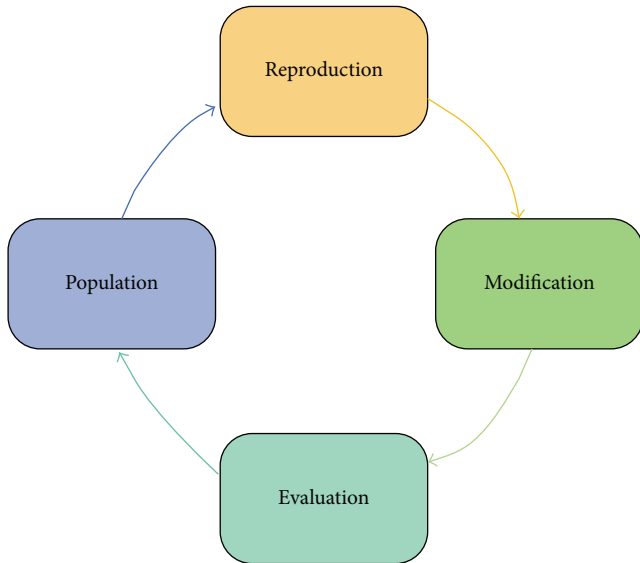


FIGURE 3: Components of a GA.

The dataset has 22 relevant features as given in Table 1 and includes a total of 192 cases. Thus, it is a matrix with dimensions of 192×22 . Training of the GA-ELM is carried out with dataset of 128 and the remaining data is used for testing of GA-WK-ELM. To optimize the parameters of WK-ELM, GA is used. The fitness function of the GA is training error of WK-ELM classifier.

This GA-WK-ELM method for diagnosis of PD includes three layers. In the first layer of GA-WK-ELM, the Parkinson data is obtained from the UCI machine learning database mentioned in Section 5. In the second layer of GA-WK-ELM, the numbers of hidden neurons of WK-ELM and parameters of wavelet kernel function are optimized by the GA. In the GA structure, an individual has a total of 20 bits. These bits can be ordered as below:

- (i) *The First Four Bits of GA-WK-ELM.* In this structure of GA-WK-ELM, the 1st, 2nd, 3rd, and 4th bits of each of these individuals symbolize the w parameter values (between 1 and 16) of the wavelet kernel functions.
- (ii) *The Second Four Bits of GA-WK-ELM.* In this structure of GA-WK-ELM, the 5th, 6th, 7th, and 8th bits of each of these individuals symbolize the x parameter values (between 1 and 16) of the wavelet kernel functions.
- (iii) *The Third Four Bits of GA-WK-ELM.* In this structure of GA-WK-ELM, the 9th, 10th, 11th, and 12th bits of each of these individuals symbolize the y parameter values (between 1 and 16) of the wavelet kernel functions.
- (iv) *The Rest of the Bits of GA-WK-ELM.* In this structure of GA-WK-ELM, the rest of the 20 bits symbolize the number of hidden neurons (between 5 and 260).

The 40 number of these individuals is randomly selected for the initial population. So, this GA structure is purposed to obtain the best possible performance from the WK-ELM

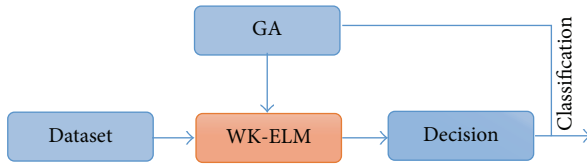


FIGURE 4: The block diagram of GA-WK-ELM based optimal PD diagnosis system.

classifier. The training and testing dataset for the proposed GA-WK-ELM method is obtained from the UCI dataset. This Parkinson disease dataset includes 192 pieces of data. The randomly selected 100 of 192 pieces of PD data are used for training of WK-ELM classifier. The rest of this PD data is used for testing of the WK-ELM classifier. In here, the results of the proposed GA-WK-ELM method are given for the optimum parameters values of wavelet kernel function and number of hidden neurons of WK-ELM. In here, a comparison is performed with previous studies to show the validity of the proposed GA-WK-ELM method. From results, the proposed GA-WK-ELM method is a quite powerful tool for automatic diagnosis of hepatitis and may work in real-time systems.

The block diagram of the proposed GA-WK-ELM method is given in Figure 4. In these applications, a 3-fold cross-validation schema was applied where the two-fifths data were used for training the proposed GA-WK-ELM method and the other remaining data were used as the test dataset. This method was repeated for three times for obtaining the average classification rates. Thus, the correct diagnosis performance of the suggested GA-WK-ELM method is computed.

In here, the maximum training accuracy value of the WK-ELM classifier was used as the fitness function of GA. This maximum training accuracy was calculated from the result of training of WK-ELM for each of the individuals by using parameters represented by these individuals. The w , x , and y parameters values of wavelet kernel function and the number of hidden neurons of the WK-ELM classifier are optimized by GA. The PD dataset has 22 relevant features. These features were obtained from 192 patients. So, dimensions of the features vector are 192×22 . Here, 40 random individuals are selected as the initial population. Each of these individuals has 20 bits. In Tables 2 and 3, coding for parameters of wavelet kernel function and the number of hidden neurons are given, respectively.

An example for individuals of the population is shown in Figure 5. The 1st, 2nd, 3rd, and 4th bits of this individual symbolize the w parameter values of the wavelet kernel functions, which are between 1 and 16. The 5th, 6th, 7th, and 8th bits of this individual symbolize the x parameter values of the wavelet kernel functions, which are between 1 and 16. The 9th, 10th, 11th, and 12th bits of this individual symbolize the y parameter values of the wavelet kernel functions, which are between 1 and 16. The rest of the 20 bits of this individual symbolize the number of hidden neurons, which are between 5 and 260.

The correct diagnosis performance of the suggested GA-WK-ELM method for PD dataset is computed by three

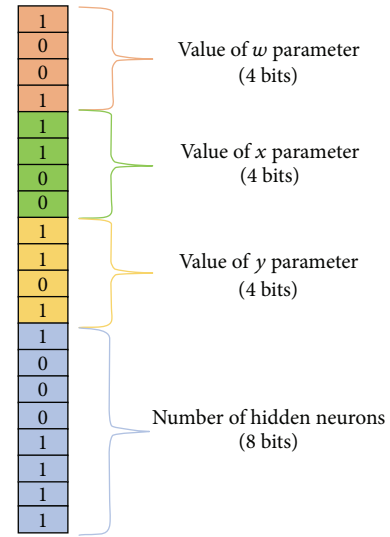


FIGURE 5: An example for individuals of the population.

TABLE 2: Coding for parameters of wavelet kernel function.

Values of w , x , and y parameters	Coding
1	0 0 0 0
2	0 0 0 1
3	0 0 1 0
4	0 0 1 1
5	0 1 0 0
6	0 1 0 1
7	0 1 1 0
8	0 1 1 1
9	1 0 0 0
10	1 0 0 1
11	1 0 1 0
12	1 0 1 1
13	1 1 0 0
14	1 1 0 1
15	1 1 1 0
16	1 1 1 1

TABLE 3: Coding for number of hidden neurons.

The number of hidden neurons	Coding
5	0 0 0 0 0 0 0 0
6	0 0 0 0 0 0 0 1
7	0 0 0 0 0 0 1 0
8	0 0 0 0 0 0 1 1
⋮	⋮
260	1 1 1 1 1 1 1 1

evaluation methods as classification accuracy, sensitivity and specificity analysis, and Receiver Operating Characteristic (ROC) curves.

TABLE 4: The correct Parkinson diseases diagnosis performance comparing of the GA-WK-ELM method with classic ELM classifiers, which have different types of kernel function and the number of hidden neurons.

Used method	Type of the kernel function	Value of w wavelet kernel parameter	Value of x wavelet kernel parameter	Value of y wavelet kernel parameter	The number of hidden neurons	Accuracy (%)
GA-WK-ELM	Wavelet	6	5	12	74	96.32
GA-WK-ELM	Wavelet	15	3	10	86	96.81
GA-WK-ELM	Wavelet	9	4	17	23	95.46
GA-WK-ELM	Wavelet	5	4	12	42	95.28
Classic ELM	Poly	—	—	—	76	89.31
Classic ELM	Hard limit	—	—	—	142	83.22
Classic ELM	Tangent sigmoid	—	—	—	164	91.75
Classic ELM	Tangent sigmoid	—	—	—	242	91.64
Classic ELM	Poly	—	—	—	56	83.85
Classic ELM	Radial basis	—	—	—	108	87.62
Classic ELM	Sigmoid	—	—	—	265	92.28
Classic ELM	Radial basis	—	—	—	356	91.54
Classic ELM	Radial basis	—	—	—	462	93.45

The performance of the proposed method is evaluated by the *Sensitivity Analysis* (SEA) and *Specificity Analysis* (SPA) and classification accuracies, which are obtained from statistical methods in given equations ((9)–(11)), respectively, are presented in Table 6:

SEA

$$= \frac{\text{the number of correctly classified persons with PD}}{\text{the number of total PD cases}}, \quad (9)$$

SPA

$$= \frac{\text{the number of persons correctly classified as healthy}}{\text{the number of total healthy persons}}. \quad (10)$$

The overall classification correct ratio of the proposed method (OC) is calculated as (7):

$$\text{OC} = \frac{\text{the number of correct classifications}}{\text{the number of total cases}}. \quad (11)$$

In this experimental study, a genetic algorithm structure was designed for deciding the w , x , y parameters values of wavelet kernel function and the number of hidden neurons of the WK-ELM classifier. A total of 20 bits are used for each of the individuals in the initial population. In this GA structure, the 1st, 2nd, 3rd, and 4th bits of the individual give the w parameter value, the 5th, 6th, 7th, and 8th bits of this individual give the x parameter value, and the 9th, 10th, 11th, and 12th bits of the individual give the y parameter value, respectively. The remainder of the 20 bits of the individual give the number of hidden neurons, which are between 5 and 260.

6. Obtained Results

In these experimental studies, an expert diagnosis system for PD based on the GA-WK-ELM method is introduced.

Then, the correct PD diagnosis performance of the suggested GA-WK-ELM method is also evaluated by classification accuracy, sensitivity and specificity analysis, and ROC curve, respectively.

The suggested GA-WK-ELM method is used for finding the optimum values of the wavelet kernel function w , x , and y parameters and the number of ELM classifier hidden neurons in these experimental studies. The comparing results by using the GA-WK-ELM method and the classic ELM classifier by using the same PD database can be given in Table 4. In these classic ELM classifiers, each of sigmoid, tangent sigmoid, triangular basis, radial basis, hard limit, and polykernel functions is used as the kernel function, respectively. The readers can find more detailed information about these kernel functions in [21, 22]. As shown in this table, the best correct diagnosis rate of the suggested GA-WK-ELM method is found as 96.81% by using 15, 3, 10, and 86 values for the w , x , y wavelet kernel function parameters and the number of hidden neurons, respectively.

As shown in Table 4, the highest correct PD diagnosis rate is obtained as 96.81% by using the suggested GA-WK-ELM method, because the optimum values of the WK-ELM w , x , y parameters and the numbers of hidden neurons of WK-ELM were obtained by using a genetic algorithm (GA) in these experimental studies. In this study, after finding the optimum parameters, we do not need to use GA and then the WK-ELM can be directly used.

In Table 5, to show the validity of the suggested GA-WK-ELM method, compared results with previous studies using the same dataset [6–12] are also given. From this table, the highest diagnosis rate is calculated as 94.72 by [13] by using the neurofuzzy classifier with linguistic hedges (ANFIS-LH). In here, training times have not been given in these studies. Moreover, the feature vector is randomly reduced to a lower dimension in [13]. The suggested GA-WK-ELM method in this study shows a correct diagnosis

TABLE 5: The comparison results of the proposed GA-WK-ELM method and previous studies.

Studies	Method	The number of features	Training		Testing
			Time	Accuracy (%)	Accuracy (%)
Ref [13]	ANFIS-LH	4	—	95.38	94.72
	MLPNN	4	—	93.88	89.69
	RBFNN	4	—	91.84	87.63
Ref [14]	PNN	22	—	81.74	81.28
In this study	GA-WK-ELM	22	0.21 μ s	99.42	96.81

TABLE 6: The obtained PD diagnosis accuracy by statistical methods.

Method	Correct diagnosis rate
Sensitivity Analysis	95.45
Specificity Analysis	98.17
Average	96.81

performance even though the feature vector is directly used without reduction. However, the training time of WK-ELM is extremely short.

The obtained PD diagnosis accuracies by statistical evaluation criteria are given in Table 6.

In this study, ROC curves and AUC values are calculated by using TP, TN, FP, and FN, which are true positives, true negatives, false positives, and false negatives, respectively [27]. The used ROC curve in this study is a graphical plot. It shows the performance of a binary classifier system as its discrimination threshold is varied. The ROC curve is formed by plotting the true-positive rate against the false-positive rate at various threshold settings. In here, the true-positive rate is also known as sensitivity in biomedical informatics or recall in machine learning. The false-positive rate is also known as the fallout. It can be calculated as $1 - \text{specificity}$. So, the ROC curve is the sensitivity as a function of the fallout.

ROC analysis supplies tools to choose the optimal models. Moreover, it eliminates the suboptimal ones independently from the class distribution or the cost context. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision-making. In here, the ROC curve of GA-WK-ELM is given by using the obtained best TP, TN, FP, and FN values in Figure 6. The obtained AUC value of ROC curves by using the GA-WK-ELM classifier can be given as 0,9576.

7. Discussion and Conclusion

This paper suggests an expert PD diagnosis system based on GA-WK-ELM. The proposed GA-WK-ELM PD diagnosis system has advantages such as finding of the optimal w , x , and y parameters combination of wavelet kernel, direct using of feature vector, fast training and testing time, and generalization capability over conventional neural networks with backpropagation. The suggested GA-WK-ELM method is formed from two stages as WK-ELM classification and optimization of WK-ELM classifier's parameters. The feature vector from Parkinson dataset is used as input to the

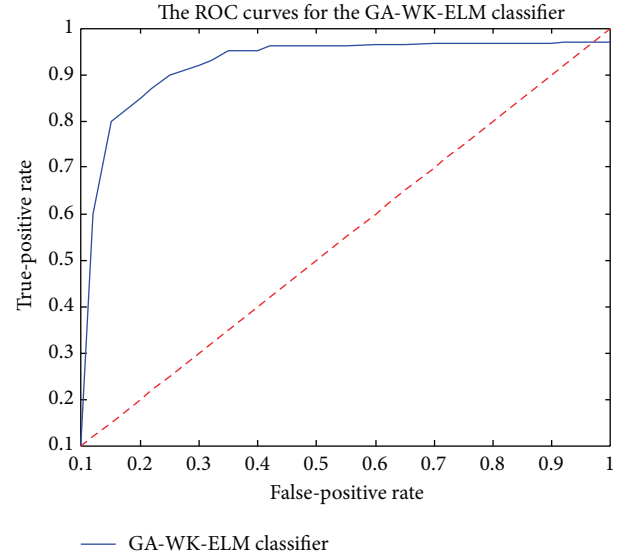


FIGURE 6: The ROC curve of the suggested GA-WK-ELM method for expert PD diagnosis.

WK-ELM classifiers. In wavelet kernel-Extreme Learning Machine (WK-ELM) structure, there are three adjustable w , x , and y parameters of wavelet kernel. These w , x , and y parameters and the numbers of hidden neurons play a major role in the performance of WK-ELM. Because of this, values of these w , x , and y parameters and numbers of hidden neurons should be carefully set based on the solved diagnosis of the PD problem. In this paper, the optimum values of these wavelet kernel parameters and the numbers of hidden neurons of WK-ELM were calculated by using GA. The output of WK-ELM makes decisions about diagnosis of PD. The optimum values of the wavelet kernel w , x , and y parameters and numbers of hidden neurons of the WK-ELM classifier are calculated by a GA to obtain the best possible PD diagnosis performance. The feasibility of the suggested GA-WK-ELM method has been tested by using PD dataset. This dataset has 192 test cases. The suggested GA-WK-ELM method has effective PD diagnosis performance when compared with previous studies depending on direct using of the same feature vector and training time as shown in Tables 4–6 and Figure 6.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] R. Betarbet, T. B. Sherer, and J. T. Greenamyre, "Animal models of Parkinson's disease," *BioEssays*, vol. 24, no. 4, pp. 308–318, 2002.
- [2] G. F. Wooten, L. J. Currie, V. E. Bovbjerg, J. K. Lee, and J. Patrie, "Are men at greater risk for Parkinson's disease than women?" *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 75, no. 4, pp. 637–639, 2004.
- [3] E. Tolosa, G. Wenning, and W. Poewe, "The diagnosis of Parkinson's disease," *Lancet Neurology*, vol. 5, no. 1, pp. 75–86, 2006.
- [4] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinicopathological study of 100 cases," *British Medical Journal*, vol. 55, pp. 181–184, 1992.
- [5] P. P. Piccini and A. Whone, "Functional brain imaging in the differential diagnosis of Parkinson's disease," *The Lancet Neurology*, vol. 3, no. 5, pp. 284–290, 2004.
- [6] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.
- [7] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices: a voice analysis system for the screening and laryngeal diseases," *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, no. 4, pp. 74–82, 1997.
- [8] J. I. Godino-Llorente and P. Gómez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 380–384, 2004.
- [9] S. Hadjitodorov, B. Boyanov, and B. Teston, "Laryngeal pathology detection by means of class-specific neural maps," *IEEE Transactions on Information Technology in Biomedicine*, vol. 4, no. 1, pp. 68–73, 2000.
- [10] D. A. Rahn III, M. Chou, J. J. Jiang, and Y. Zhang, "Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis," *Journal of Voice*, vol. 21, no. 1, pp. 64–71, 2007.
- [11] M. Little, P. McSharry, I. Moroz, and S. Roberts, "Nonlinear biophysically-informed speech pathology detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, Toulouse, France, May 2006.
- [12] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting Nonlinear recurrence and Fractal scaling properties for voice disorder detection," *BioMedical Engineering Online*, vol. 6, article 23, 2007.
- [13] M. F. Caglar, B. Cetisli, and I. B. Toprak, "Automatic recognition of Parkinson's disease from sustained phonation tests using ANN and adaptive neuro-fuzzy classifier," *Journal of Engineering Science and Design*, vol. 1, pp. 59–64, 2010.
- [14] M. Ene, "Neural network based approach to discriminate healthy people from those with Parkinson's disease," *Annals of University of Craiova: Mathematics and Computer Science Series*, vol. 35, pp. 112–116, 2008.
- [15] G. A. David and J. B. Magnus, "Diagnosing Parkinson by using artificial neural networks and support vector machines," *Global Journal of Computer Science and Technology*, vol. 9, no. 4, pp. 63–71, 2009.
- [16] C. O. Sakar and O. Kursun, "Telediagnosis of parkinson's disease using measurements of dysphonia," *Journal of Medical Systems*, vol. 34, no. 4, pp. 591–599, 2010.
- [17] S. Suresh, S. N. Omkar, and V. Mani, "Parallel implementation of back-propagation algorithm in networks of workstations," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 1, pp. 24–34, 2005.
- [18] C. T. Hsu, M. S. Kang, and C. S. Chen, "Design of adaptive load shedding by artificial neural networks," *IEE Proceedings—Generation, Transmission and Distribution*, vol. 152, no. 3, pp. 415–421, 2005.
- [19] J. H. Cho, D. J. Lee, and M. G. Chun, "Parameter optimization of extreme learning machine using bacterial foraging algorithm," *Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 6, pp. 807–812, 2007.
- [20] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 42, no. 2, pp. 513–529, 2012.
- [21] W. Huang, N. Li, Z. Lin et al., "Liver tumor detection and segmentation using kernel-based extreme learning machine," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '13)*, pp. 3662–3665, Osaka, Japan, July 2013.
- [22] S. F. Ding, Y. A. Zhang, X. Z. Xu, and L. N. Bao, "A novel extreme learning machine based on hybrid kernel function," *Journal of Computers*, vol. 8, no. 8, pp. 2110–2117, 2013.
- [23] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '04)*, vol. 2, pp. 985–990, Budapest, Hungary, July 2004.
- [24] B. Li, X. Rong, and Y. Li, "An improved kernel based extreme learning machine for robot execution failures," *The Scientific World Journal*, vol. 2014, Article ID 906546, 7 pages, 2014.
- [25] M. Melanie, *An Introduction to Genetic Algorithms*, A Bradford Book The MIT Press, Cambridge, Mass, USA, 5th edition, 1999.
- [26] D. Whitley, "An executable model of a simple genetic algorithm," *Foundations of Genetic Algorithms*, vol. 2, no. 1519, pp. 45–62, 2014.
- [27] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.