Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier

Mehmet Fatih CAGLAR[1], Bayram CETISLI[2], Inayet Burcu TOPRAK[3]

[1]Department of Electronics and Communication Engineering,
[2]Department of Computer Engineering,
Faculty of Engineering and Architecture, Süleyman Demirel University
[3]Biomedical Equipment Technology Program,
Vocational School of Technical Sciences, Akdeniz University
mfcaglar@mmf.sdu.edu.tr, bcetisli@mmf.sdu.edu.tr , ibmutlu@akdeniz.edu.tr

Abstract- Neurological disorders contain Parkinson's disease (PD), epilepsy and Alzheimer's; influence the lives of patients and their families. PD creates cognitive and state of mind disturbances. Generally, the diagnosis is based on medical history and neurological inspection conducted by interviewing and observing the patient in person using the Unified Parkinson's Disease Rating Scale (UPDRS).

In this study, we aimed to discriminate between healthy people and people with PD. For that reason, Parkinson dataset that contains biomedical voice of human is used. Artificial Neural Networks (ANN) are widely used in biomedical field for modeling, data analysis, and diagnostic classification. Two types of the ANNs were used for classification: Multilayer Perceptrons (MLP) and Radial Basis Function (RBF) Networks. The other method is Adaptive Neuro-Fuzzy Classifier (ANFC) with linguistic hedges. This method is also used for feature selection from the dataset. Adaptive Neuro-Fuzzy Classifier with linguistic hedges gave the best recognition results with %95.38 training and %94.72 testing classifying performance indeed.

Keywords- Parkinson's Disease, Sustained Vowel Phonation, Classification, Adaptive Neuro-Fuzzy Classifier with Linguistic Hedges, Artificial Neural Networks.

Yapay Sinir Ağları ve Adaptif Sinir-Bulanık Sınıflayıcı Kullanarak Sürdürülmüş Fonasyon Testlerinden Parkinson Hastalığının Otomatik Tanınması

Özet- Parkinson hastalığı, epilepsi ve Alzheimer hastalığı hasta ve ailesinin yaşamını etkileyen nörolojik bozukluklardır. Parkinson hastalığı, ruhsal durum çöküntülerine ve bilişsel bozukluklara sebep olur. Genellikle, bu hastalığın teşhisinde Parkinson Hastalığı Değerlendirme Ölçeğinin kullanıldığı gözle incelenme ve karşılıklı konuşmaya dayalı nörolojik muayene ve hastanın medikal hikâyesi esas alınır.

Bu çalışmada, sağlıklı kişiler ile Parkinson hastalığı olan kişileri birbirinde ayırt etmek amaçlandı. Bu sebeple, insanın biyomedikal konuşma sesinden (sürdürülmüş fonasyon test kayıtları) elde edilmiş özelliklerini içeren Parkinson veri seti kullanıldı. Yapay Sinir Ağları (YSA) biyomedikal alanda modelleme, veri analizi ve teşhis amaçlı sınıflandırma için geniş uygulama alanı bulmaktadır. Sınıflandırma için iki tip YSA kullanıldı: Çok Katmanlı Algılayıcı (ÇKA) ve Radyal Tabanlı Fonksiyon Ağları (RTFA). Diğer metot olarak ise Dilsel Kuvvetli Adaptif Sinir-Bulanık Sınıflayıcı (DKASBS) kullanıldı. DKASBS aynı zamanda veri setinden özellik seçimi için de kullanıldı. Dilsel Kuvvetli Adaptif Sinir-Bulanık Sınıflayıcı %95.38 eğitim ve %94.72 test başarı oranları ile en iyi tanıma sonuçlarını vermiştir.

Anahtar Kelimeler- Parkinson Hastalığı, Sürdürülmüş Sesli Harf Fonasyonu, Sınıflandırma, Dilsel Kuvvetli Adaptif Sinir-Bulanık Sınıflayıcı, Yapay Sinir Ağları.

1. INTRODUCTION

PD is a degenerative disorder of the central nervous system. It was first described in 1817 by James Parkinson, a British physician who published a paper on what he called "the shaking palsy." Researchers believe that at least 500,000 people in the United States currently have PD, although some estimates are much higher. Society pays an enormous price for PD. The total cost to the nation is estimated to exceed $6 billion annually. The risk of PD increases with age, so analysts expect the financial and public health impact of this disease to increase as the population gets older [1].

The symptoms of Parkinson's disease result of the loss of dopamine which is a brain chemical (neurotransmitter) involved in controlling movement. The shortage of this brain chemical occurs when nerve cells in a part of the brain (substantia nigra) that produces dopamine fail and deteriorate. The exact cause of this deterioration is not known [2]. But scientists are doing a lot of research to look for the answer. They are studying many possible causes, including aging and poisons in the environment. Abnormal genes seem to lead to Parkinson's disease in some people. But so far, there is not enough proof to show that it is always inherited [3].

The four main symptoms of Parkinson's are tremor, which means shaking or trembling. Tremor may affect your hands, arms, legs, or head: stiff muscles, slow movement and problems with balance or walking. Other symptoms may include depression and other emotional changes; difficulty in swallowing, chewing, and speaking; urinary problems or constipation; skin problems; and sleep disruptions [1].

PD usually affects people over the age of 50. There are currently no blood or laboratory tests that have been proven to help in diagnosing sporadic PD. Therefore the diagnosis is based on medical history and a neurological examination. The disease can be difficult to diagnose accurately. Doctors may sometimes request brain scans or laboratory tests in order to rule out other diseases [1]. Particularly in early stages, the disease can be difficult to diagnose accurately. Owing to symptom overlap with other diseases only 75% of clinical diagnoses of PD are confirmed to be idiopathic PD at postmortem [4]. So, automatic techniques based on Artificial Intelligence are needed to increase the diagnosis accuracy and to help physicians make better decisions [5]. At present, there is no cure for PD. But there are several types of medicines that can control the symptoms and make the disease easier to live with [1], although %1 of people over 65 years-old in the society have PD [6].

This paper deals with the application of ANN and Adaptive Neuro-Fuzzy to a medical dataset concerning PD with the aim of automatically classify patients in PD or non-PD depending on their medical attributes.

2. MATERIALS AND METHODS

2.1 Parkinson Dataset

The data for this study consists of 195 sustained vowel phonations from 31 people, of which 23 were diagnosed with PD [5, 7, 8]. 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 non-PD for healthy and PD for people with Parkinson's disease, which is a two-decision classification problem. The features of dataset are given in Table 1.

Table 1. The features of Parkinson dataset

No	Feature	No	Feature	No	Feature
1	MDVP:Fo(Hz)	**9**	MDVP:Shimmer	**17**	RPDE
2	MDVP:Fhi(Hz)	**10**	MDVP:Shimmer(dB)	**18**	DFA
3	MDVP:Flo(Hz)	**11**	Shimmer:APQ3	**19**	spread1
4	MDVP:Jitter(%)	**12**	Shimmer:APQ5	**20**	spread2
5	MDVP:Jitter(Abs)	**13**	MDVP:APQ	**21**	D2
6	MDVP:RAP	**14**	Shimmer:DDA	**22**	PPE
7	MDVP:PPQ	**15**	NHR	**23**	Class Label
8	Jitter:DDP	**16**	HNR		

2.2 Neural Networks

Neural networks are used as a powerful means in engineering area after development especially, in computer technology. The fundamental characteristic of the neural networks is an adaptive, non-algorithmic and parallel distributed [9]. ANNs are widely used in biomedical field for modeling, data analysis, and diagnostic classification.

In this study, two types of the ANNs are used: MLP, and RBF networks. The MLPNNs, which have features such as the ability to learn and generalize, smaller training set requirements, fast operation, ease of implementation and

therefore most commonly used neural network architectures [9]. The MLPNNs are a non-parametric technique for performing a wide variety of detection and estimation tasks.

There are number of training algorithms used to train an MLPNN and a frequently used one is called the backpropagation training algorithm. The backpropagation algorithm, which is based on searching an error surface using gradient descent for points with minimum error, is relatively easy to implement. However, the backpropagation has some problems for many applications. The algorithm is not guaranteed to find the global minimum of the error function since gradient descent may get stuck in local minima, where it may remain indefinitely. In addition to this, long training sessions are often required in order to find an acceptable weight solution because of the well-known difficulties inherent in gradient descent optimization. Therefore, a lot of variations to improve the convergence of backpropagation were proposed. Optimization methods such as second-order methods (conjugate gradient, quasi-Newton, Levenberg-Marquardt) have also been used for ANN training in recent years [9].

A radial basis function network is an artificial neural network that uses radial basis functions as activation functions. It is a linear combination of radial basis functions. They are used in function approximation, time series prediction, control, and classification. Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer [9].

2.3 Adaptive Neuro-Fuzzy Classifier with Linguistic Hedges

Fuzzy classification systems, which are based on fuzzy rules, have been successfully applied to various classification tasks [9]. The fuzzy systems can be constituted with neural networks, and such systems are called as neuro-fuzzy systems [9]. The neuro-fuzzy classifiers define the class distributions and show the input-output relations [9], whereas the fuzzy systems describe the systems using natural language. Neural networks are employed for tuning or training the system parameters in neuro-fuzzy applications. An ANFC consist of input, membership function, fuzzification, defuzzification, normalization and output layers [9, 10].

The LHs that are constituted by the power of fuzzy sets introduce the importance of the fuzzy sets for fuzzy rules. They can also change the primary meaning of fuzzy membership functions to secondary meaning.

Let A be a continuous linguistic term for input variable x with MF $\mu_A(x)$. Then A^s is interpreted as a modified version of the original linguistic term expressed as

$$A^s := \left\{ \left(x, \left(\mu_A(x) \right)^p \right) | x \in X \right\}, \quad (1)$$

where p denotes the linguistic hedge value of the linguistic term A.

To improve the meaning of fuzzy rules and classification accuracy, a layer, which defines the adaptive linguistic hedges, is added into the proposed classifier network [10]. The LHs are trained with other network parameters by scaled conjugate gradient (SCG) training algorithm. The tuned LH values of fuzzy sets improve the flexibility of fuzzy sets; this property of LH can improve the distinguishability rates of overlapped classes.

The ANFC with LHs is based on fuzzy rules. A fuzzy classification rule that has two inputs $\{x_1, x_2\}$ and one output y is defined with LHs as;

IF x_1 is A_1 with p_1 hedge **AND** x_2 is A_2 with p_2 hedge **THEN** y is C_1 class,

where A_1 and A_2 denote linguistic terms that are defined on X_1 and X_2 feature space; p_1 and p_2 denote linguistic hedges, respectively; C_1 denotes the class label of the output y.

The values of LHs can be used to show the importance degree of fuzzy sets. When this property is used for classification problems, and every class is defined by a fuzzy classification rule, the LHs of every fuzzy set denote the importance degree of input features. If the LHs values of features are close to concentration values, these features are more important or relevant, and can be selected. On the contrary, if the LH values of features are close to dilation values, these features are not important, and can be eliminated. According to the LHs value of features, the redundant, noisily features can be eliminated, and significant features can be selected [10].

In some problems, a lot of features can be used.If irrelevant features are used in combination with good features, the classifier will not perform well as it would with only good features. Therefore, the goal should be aimed at choosing a discriminative subset of features. ANFC with LHs can also be used for feature selection.

3. EXPERIMENTAL STUDIES

Before the classification of Parkinson dataset, the important features of dataset should be determined to diagnosis of the disease easily. For this aim, the ANFC with LHs are used to select the relevant features. The selected features MDVP:Fhi(Hz)-spread1, and status-D2 are illustrated in Fig. 1 and Fig. 2, respectively.

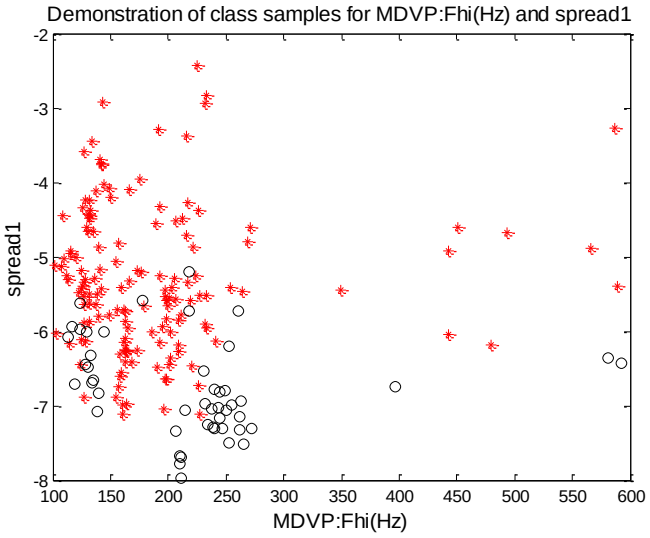

Figure 1. Demonstration of MDVP:Fhi(Hz)-spread1 dataset

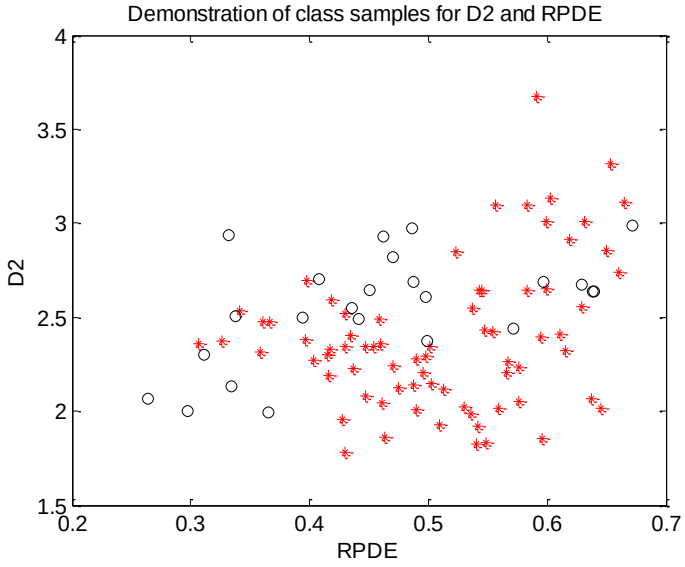

Figure 2. Demonstration of D2-RPDE dataset

After using ANFC with LHs for determining powerful features, training and testing procedures are applied with them. Each column of the data has 195 different properties which divided in 2 parts, %50 for both training and testing. Each person on the database has 6 or 7 different records. If a person has got enough "true" checks like as 4 of 6/7 records, he/she will be healthy or has PD.

A MLP network which has two logarithmic-sigmoidal function activated hidden layers and a linear output layer is trained with scaled-conjugated gradient (SCG) back propagation algorithm for rapid convergence. The structure of the MLPNN is 4-40-20-2. The RBF network has a Radial-Basis layer with 90 neurons and a linear output layer. The spread of the RBFN is set 300 for best recognition. The ANFC with LHs which also is trained with SCG back propagation algorithm for rapid convergence has 2 classes with 2 clusters. All networks are run 10 times each and average training and testing performances are obtained.

The selected features and the obtained classification results with ANFC-LH are given in Table 2.

Table 2. The classification results of Parkinson Dataset with selected features.

Selected Features	Number of Features	Recognition Rate (%)
All (23)	**23**	**94.35**
1-2-10-17-20-21-22	**7**	**96.41**
2-17-19-21	**4**	**95.38**

These results that are obtained by ANFC-LH are compared with the other methods. MLPNN and RBFNN and support vector machines (SVM) (7) are used for comparison. The comparison results are given in Table 3.

Table 3. The comparison results of different methods for Parkinson dataset classification.

Method	No of Features	Features	Recognition Rate (%)	
SVM (7)	4	16-17-18-22	**91.40±4.4**	
			Training (%)	**Testing (%)**
MLPNN	4	2-17-19-21	**93.88**	**89.69**
RBFNN	4	2-17-19-21	**91.84**	**87.63**
ANFC-LH	4	2-17-19-21	**95.38**	**94.72**

According to Table 3, all of methods compared with same number of features, and ANFC-LH gives the best classification result. The classification true/false counts are listed in Table 4.

The performance of MLPNN, RBFN and ANFC with LHs is assessed by the following measures and the statistical results are listed in Table 5.

Table 4. The results of classification as PD/healthy of 31 persons.

Output/desired	Result (PD)			Result (Healthy)		
Network Type	MLPNN	RBF	ANFC-LH	MLPNN	RBF	ANFC-LH
Result (PD)	**22**	**22**	**23**	**1**	**1**	**0**
Result (Healthy)	**1**	**1**	**1**	**7**	**7**	**7**

$$\text{Specificity}=\frac{\text{number of correct classified PD}}{\text{number of total PD}} \quad (2)$$

$$\text{Sensitivity}=\frac{\text{number of correct classified healthy}}{\text{number of total healthy}} \quad (3)$$

$$\text{Total classification accuracy}=\frac{\text{number of correct classified persons}}{\text{number of total persons}} \quad (4)$$

Table 5. Statistical results.

Statistical parameters	Values (%)		
	MLPNN	RBF	ANFC-LH
Specificity	**95.65**	**95.65**	**100.00**
Sensitivity	**87.50**	**87.50**	**87.50**
Total classification accuracy	**93.55**	**93.55**	**96.77**

4. CONCLUSION AND DISCUSSION

Experimental results show that ANFC-LH classifier has enough satisfaction with higher accuracy for classification of Parkinson dataset which collected from sustained vowel phonations. Also, ANFC-LH classifier is used for the feature selection. Another important item is that the selected features are different from the M.A Little et al. study (7). They used HNR, RPDE, DFA and PPE features of dataset. However, we used completely different features, such as MDVP: Fhi(Hz), RPDE, spread1 and D2. M.A Little et al. used uncorrelated measurement method based on filtering for selection of features. So, in this work we take into consideration that different and novel feature selection and classifying techniques presented, valuable for diagnosing hidden PD in early ages, may give greater accurate solutions.

ACKNOWLEDGEMENT

The dataset was created by Max Little of the University of Oxford in collaboration with the National Centre for Voice and Speech, Denver, Colorado. We thank to all them.

5. REFERENCES

[1] National Institute of Neurological Disorders and Stroke, http://www.ninds.nih.gov/disorders/parkinsons_disease/detail_parkinsons_disease.htm, (2009.06.21)

[2] WebMD, http://www.webmd.com/parkinsons-disease/guide/parkinsons-disease-cause , (2009.06.21)

[3] WebMD, http://www.webmd.com/parkinsons-disease/guide/parkinsons-disease-topic-overview?page=2 , (2009.06.21)

[4] Gelb, D., Oliver, E., Gilman S, "Diagnostic criteria for Parkinson disease". *Arch Neurol* 56 (1): 33–9, doi:10.1001/archneur.56.1.33. PMID 9923759, 1999.

[5] Ene, M., "Neural network-based approach to discriminate healthy people from those with Parkinson' s disease", Annals of University of Craiova, *Math. Comp. Sci. Ser.*, Vol. 35, 112-116, 2008.

[6] Apaydın H., Özekmekci, S., **Parkinson Hastalığı: Hasta ve Yakınları için El Kitabı**, Parkinson Hastalığı Derneği, İstanbul, 2008.

[7] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O., 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', *Nature Precedings*, hdl:10101/npre.2008.2298.1, 2008.

[8] Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A.E., Moroz, I.M., "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection", *BioMedical Engineering*, OnLine 6:23, 2007.

[9] Jang, J.S.R., Sun, C.T. & Mizutani, E., **Neuro-fuzzy and soft computing**', Upper Saddle River :Prentice Hall, 1997.

[10] Cetisli, B., "Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1" *Expert Systems with Applications*, doi:10.1016/j.eswa.2010.02.108, 2010.

[11] Toprak I. B., Caglar M. F., Merdan M.,"High Accuracy Classification of EEG Signals Using Wavelet Transform and Neural Networks for Diagnosing Epilepsy", **INISTA-2007 (Innovations in Intelligent Systems and Applications),** Istanbul, 2007.