DIAGNOSIS OF THE PARKINSON DISEASE BY USING DEEP NEURAL NETWORK CLASSIFIER

Abdullah CALISKAN[1], Hasan BADEM[2,3], Alper BAŞTÜRK[2], Mehmet Emin YÜKSEL[1]

[1]Biomedical Engineering, Erciyes University, Kayseri, Turkey
[2]Computer Engineering, Erciyes University, Kayseri, Turkey
[3]Computer Engineering, Sutcu Imam University, Kahramanmaras, Turkey
acaliskan@erciyes.edu.tr, hbadem@erciyes.edu.tr, ab@erciyes.edu.tr, yuksel@erciyes.edu.tr

Abstract: *Parkinson disease occurs when certain clusters of brain cells are unable to generate dopamine which is needed to regulate the number of the motor and non-motor activity of the human body. Besides, contributing to speech, visual, movement, urinary problems, Parkinson disease also increases the risks of depression, anxiety, and panic attacks, disturbances of sleep. Parkinson disease diagnosis via proper interpretation of the vocal and speech data is an important classification problem. In this paper, a Parkinson disease diagnosis is realized by using the speech impairments, which is one of the earliest indicator for Parkinson disease. For this purpose, a deep neural network classifier, which contains a stacked autoencoder and a softmax classifier, is proposed. The several simulations are performed over two databases to demonstrate the effectiveness of the deep neural network classifier. The results of the proposed classifier are compared with the results of the state-of-art classification method. The experimental results and statistical analyses are showed that the deep neural network classifier is very efficient classifier for Parkinson disease diagnosis.*
Keywords: *Parkinson disease, deep learning, deep neural network, stacked autoencoder*

1. Introduction

Parkinson's disease (PD) is a serious health problem in both industrial and developing countries, over 10 million people around the world have PD according to The American Parkinson Disease Association (APDA) [1]. It is yet unknown whether the cause of PD is environmental or genetic factors. However, it is known that the symptoms are caused by loss of certain clusters of brain cells, which have the ability to produce neurotransmitters including dopamine, acetylcholine, serotonin and norepinephrine [1, 2]. The loss of neurotransmitters, particularly dopamine, causes a number of symptoms such as speech, visual, movement, urinary problems, weight loss, depression, anxiety, and panic attacks, disturbances of sleep etc. [1-3]. Currently, there is no cure or medication that reduces or stops the progression of PD. However, it is possible to suppress or reduce the symptoms of disease especially at the early stages of the disease [4].

The requisite physical visits to the clinic for monitoring and treatment are difficult and time consuming for both the *people with Parkinson* (PWP) and physicians. Widening access to the improved communication methods and developed technology can offer the remote monitoring of PWP with reducing medical expenses and unnecessary physical visits [5]. However, reliable clinical monitoring tools must be employed to use these facilities for PWP. Studies have shown that about 90 percent of PWP have vocal impairment and speech problems [6, 7], which are one of the earliest indicator for PD [8]. PWP suffer from vocal and speech impairments such as dysphonia, hypophonia, monotone and dysarthria [9]. Therefore, analyzing the voice of the PWP with advanced signal processing techniques not only allows provides the diagnosis of the PD but also allows the tracking of the progression of the PD.

The diagnosis of the PD consists of three main steps including, preprocessing, feature extraction and classification [9, 10]. During the preprocessing step, the speech signals are filtered to eliminate noises and segmented with time-windows. From each segment, several features are extracted during the feature extraction step, which is a very sufficient step to diagnose the PD efficiently by analyzing the speech of the PWP. The performance of the classification method is dependent directly on the capabilities of feature extraction method. Therefore, another important issue that needs to be addressed in order to diagnose the PD from the speech disorders is the choice of the classification method, which is handled in this study.

Conventional classification methods including the support vector machine (SVM), naive Bayes (NB) and decision tree (DT), etc. [10, 11] produce satisfying results about the diagnosis of the PD. However, deep neural networks (DNNs) may offer a potentially superior classifier for the speech of the PWP over the conventional methods. In contrast to the conventional methods, DNNs not only reduce the dimension of the features by using autoencoders (AEs),

Received on: 01.03.2017
Accepted on: 14.03.2017

but also classify the samples by the softmax layer. DNNs have been successfully used in various medical applications [12-17]. Recent advances in the field of deep neural networks have made them attractive for classification problems [12]. The application of deep neural networks has opened a new area for complex classification problems not efficiently resolvable by other classification techniques [12].

DNN classifiers have recently shown their superiority over other classical classifier approaches based on feature vector classification [12, 18]. In this paper, we propose a DNN classifier to address the aforementioned classification problem for the diagnosis of the PD. Proposed DNN classifier can accurately diagnose PD by using the speech signals generated by the patients. Proposed DNN has the ability to learn features by using AE and design robust classifiers by using softmax layer.

The effectiveness of the DNN classifier is evaluated on real Parkinson data sets which are taken from UCI [19]. The proposed method is applied to the classification of the speech impairments. As one of the earliest indicator of PD, the speech impairments may enable us to monitor and diagnose the PWD in vivo and discover reliable biomarker for identifying PD at an early stage. In this study, we have also compared the proposed DNN classifier with other widely used methods including SVM, DT and NB on two data sets: one is Oxford Parkinson's Disease Detection (OPD) database [20] which is a tracer of the PD and normal control (NC) subjects; the other contains PWP and NC subjects in the Parkinson Speech Dataset with Multiple Types of Sound Recordings (PSD) database [8]. For both OPD and PSD, diagnosis of the PD is performed by the DNN and classical classifiers including SVM, NB and DT. Experimental results show that when using the DNN on classification of the PD, we can achieve significantly better classification performance than the both compared methods and presented methods in the literature. The experimental results indicate that the proposed classifier provides an effective way for the diagnosis and classification of the PD, thanks to its capability of generating new features from raw features.

The rest of the paper is organized as follows. In Section II, a DNN based on an autoencoder and a softmax layer is introduced and formulated for classification problem of the PD. In Section III, results of the classification experiments are reported and different aspects of the proposed classifier are discussed. Section IV presents the conclusions.

2. Deep Neural Network

Deep learning methods emerge as a highly effective method because they have a structure that allows extracting attributes from data without pre-processing. Extracting attributes from data with classical methods is an extremely tiring process. However, Deep learning methods that do this automatically can produce more effective results. Deep learning techniques are trying to imitate the working mechanism of the human brain [12,21].

The DNN consists of many simple structures that are organized to form a stack. Almost all of these simple structures perform non-linear operations, changing the data size to represent the data in a different space, helping to reveal hidden features in the data [12,21,22]. The proposed DNN has two main part stacked autoencoder (sAE) and softmax classifier, which are cascaded to each other. The desired number of autoencoder join together to form sAE [24], which will be defined below

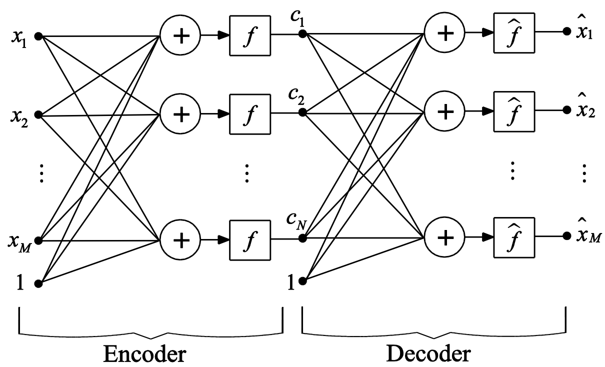

Figure 1. The Autoencoder Network

2.1. Autoencoder

The AE is a feedforward neural network, which consists of three layer including input layer, hidden layer and output layer [22]. The AE attempt to generate its own input as the output of the network, that may create different representation of inputs. Therefore, the AE is trained with an unsupervised manner to tune its weights $\mathbf{W}$ and biases $\mathbf{b}$ and to reduce the error between input $\mathbf{x}$ and its output $\hat{\mathbf{x}}$ as much as possible [22-24].

As demonstrated in Figure 1, the leftmost side of AE called encoder generates new features for second AE or softmax layer. The rightmost of the AE contains the decoder, which is employed for training of the AE. The dimension of input M of the AE is always the same with the dimension of output the AE as can be seen from Figure 1. The dimension of the hidden layer N of the AE is generally less than the dimension of input of the AE to reduce the dimension of the feature vector. However, the dimension of the hidden layer of the AE is chosen rarely greater than the dimension of input of the AE to extract hidden and interesting features from raw data set.

The objective function of AE is defined by the following function [23, 24], which consists of three part including, the mean square error E_M, regularization E_R, Kullback-Leibler divergence E_S.

$$E_T = E_M + E_R + E_S \tag{1}$$

The first part E_M in the objective function is the mean square error which is evaluated as follows:

$$E_M = \frac{1}{S}\sum_{k=1}^{S} e_k^2 \tag{2}$$

where $e_k = \left\|\boldsymbol{x}^{(k)} - \hat{\boldsymbol{x}}^{(k)}\right\|$ for $k = 1 \ldots S$ and S is the number of the instances.

The second part is given as:

$$E_R = \frac{\lambda}{2}\|\mathbf{W}\|_2^2 \quad (3)$$

Regularization term λ is employed to prevent the overfitting of the objective function in the above equation [24].

A sparsity constraint is imposed in the last part where the AE reveals hidden features from hidden layer of the AE. The last part is defined as follows:

$$E_S = \beta \sum_{j=1}^{N} KL(\rho||\hat{\rho}_j) \quad (4)$$

where β is *the weight of the sparsity penalty term* which controls the sparsity constraint.
In the last part, $KL(\rho||\hat{\rho}_j)$ is the *Kullback-Leibler divergence* which is defined as follows [24]:

$$KL(\rho||\hat{\rho}_j) = \rho \, log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \, log \frac{1-\rho}{1-\hat{\rho}_j} \quad (5)$$

Here, ρ named sparsity parameter controls the activation of the weights. The sparsity parameter is user-supplied and $\hat{\rho}_j$ evaluated below is the mean activation value of j^{th} neuron in the hidden layer of the autoencoder [23, 24].

$$\hat{\rho}_j = \frac{1}{S}\sum_{i=1}^{S} f_j(\mathbf{x}^{(i)}) \quad (6)$$

where, the activation function of the j^{th} neuron of the hidden layer is f_j.

2.2. Stacked Autoencoder

Desired number of the encoder part of the trained AE is combined to construct the stacked autoencoder (sAE). The output of the hidden layer of the trained AE is connected to the second trained AE whose hidden layer of output is connected to the input of the third trained AE. The same pattern (fourth trained AE, fifth trained AE, etc) is maintained as desired to construct the sAE. The output of the sAE is given to softmax classifier explained bellow section.

2.3. The Softmax Classifier

A softmax classifier is a supervised layer of the deep classifier [25] which is based on softmax function defined as follows:

$$v_j = \frac{e^{u_j}}{\sum_{k=1}^{K} e^{u_k}} \quad (7)$$

where $j = 1 \ldots K$.

The softmax function attempt to embed a K-dimensional vector of arbitrary real values u_j into another K-dimensional vector of real values v_j, which are normalized between zero and one.

The softmax classifier inspired by the softmax function, for data classification maps high-dimensional data samples to a lower dimensional domain while increasing the separation between different classes. A neural layer and a normalization layer combined to construct the softmax classifier shown in Figure 2.

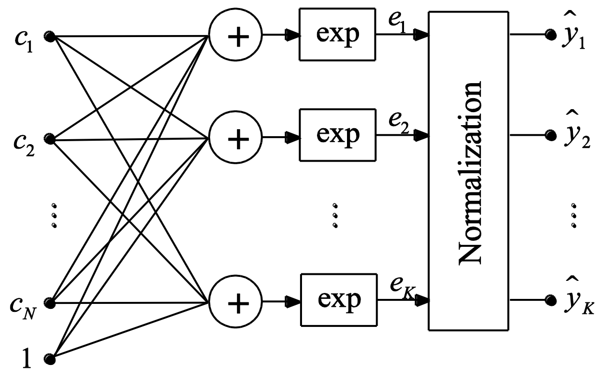

Figure 2. The softmax classifier

The input layer of the softmax classifier and the encoding section of an autoencoder are structurally very similar to each other. The only difference is that the neuron activation function here is the exponential function.

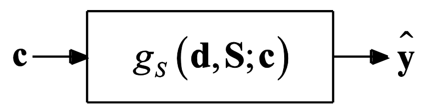

Figure 3. The block diagram of the softmax classifier

A softmax classifier attempt to embed a N-dimensional vector into another K-dimensional classes. The relationship between the input and output of the neural layer of a softmax classifier is evaluated as follows:

$$\mathbf{r} = e^{\mathbf{d}+\mathbf{S}^{\mathbf{T}}\mathbf{c}} \quad (8)$$

where $\mathbf{r} = [r_1 \, r_2 \, \ldots \, r_K]^T$, $\mathbf{d} = [d_1 \, d_2 \, \ldots \, d_K]^T$
$\mathbf{S} = [\mathbf{s_1} \, \mathbf{s_2} \, \ldots \, \mathbf{s_K}]^T$ and $\mathbf{c} = [c_1 \, c_2 \, \ldots \, c_K]^T$
Here, the elements of the **d** vector are the biases of the network. The weights of the network matrix is the **S** matrix, which has the columns defined as follows:

$$\mathbf{s_k} = [s_{k1} \, s_{k2} \, \ldots \, s_{kN}]^T \quad (9)$$

for $k = 1 \ldots K$.

The output layer of the softmax is the normalization layer which is employed for normalizing the output values of the neural layer of the softmax classifier:

$$y_j = \frac{r_j}{\sum_{k=1}^{K} r_j} \quad (10)$$

which may also be defined as follows:

$$y_j = \frac{e^{s_j^T \mathbf{c}}}{\sum_{k=1}^{K} e^{s_j^T \mathbf{c}}} \quad (11)$$

for $j = 1 \ldots K$.

The input-output relationship of the softmax classifier may shortly defined as follows:

$$\mathbf{y} = g_s(\mathbf{d}, \mathbf{S}; \mathbf{c}) \tag{12}$$

where $\mathbf{y} = [r_1\ r_2\ \dots\ r_K]^T$. The block diagram of the softmax classifier is demonstrated in Figure 3.

2.3. The proposed Deep Neural Network Classifier

The classification of the PD is achieved by using the DNN classifier, which combines the sAE network and softmax classifier. The sAE contains two encoder part of the trained AE. The structure diagram of the proposed DNN is illustrated in Figure 4. The weights of the DNN are optimized by an appropriate optimization algorithm. Limited Memory BFGS [26] optimization algorithm is one of the most suitable optimization algorithm employed for training of the DNN in this study. The input $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \dots \mathbf{x}^{(S)}\}$ of the DNN classifier is the features of the speach signals. The output of the DNN classifier $\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \dots \mathbf{y}^{(S)}\}$ is the labelled with PD and control grup which are represented with $1, 0$ respectively. The training procedure of the DNN is very complex and is summarized as follows:

1. The first AE is trained with $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \dots \mathbf{x}^{(S)}\}$ data set, which is send to both the input of the AE and output of the AE. The output of the hidden layer of the trained AE is $\{\mathbf{c}^{1,(1)}, \mathbf{c}^{1,(2)} \dots \mathbf{c}^{1,(S)}\}$, which is utilized to train the second AE. This training process shown in Figure 4-a is completely unsupervised.
2. The second AE is trained with $\{\mathbf{c}^{1,(1)}, \mathbf{c}^{1,(2)} \dots \mathbf{c}^{1,(S)}\}$ data obtained from first AE. The training of the second AE illustrated in Figure 4-b is repeated as it is in the first AE.
3. The output of the hidden layer of the second AE is $\{\mathbf{c}^{2,(1)}, \mathbf{c}^{2,(2)} \dots \mathbf{c}^{2,(S)}\}$ data, which is the input of the softmax classifier. The softmax classifier is trained to minimize the error between the label $\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \dots \mathbf{y}^{(S)}\}$ and output of the softmax classifer. This training procedure demonstated in Figure 4-c is supervised.
4. The encoder part of the trained AEs are combine to construct the sAE. The decoder part of the trained AEs are not used. The sAE and trained softmax layer are combined to construct the DNN. The weights of the DNN are tuned one more time to complete the training process shown in Figure 4-d.

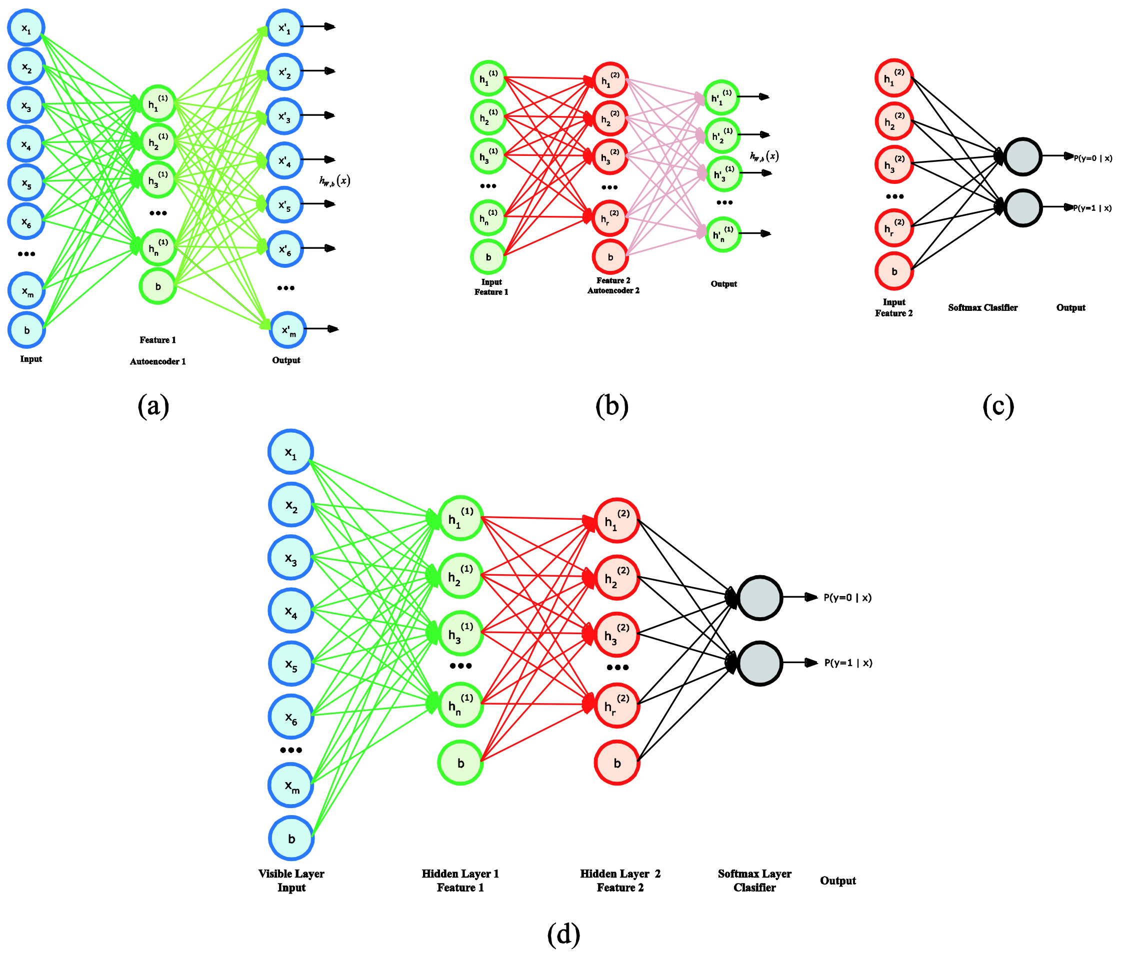

Figure 4. The training procedure of DNN network

3. Experimental Results

In this study, a DNN classifier is proposed for the diagnosis of the PD. The proposed DNN is compared with the state-art-methods including the SVM, NB and DT classifiers over OPD and PSD datasets. All methods are run for 30 different 10-fold cross-validation techniques and compared on the obtained results. All runs were performed on a computer with 3.4 GHz Intel i7 2600 CPU and 12 GB RAM.

3.1. Datasets

The main aim of "Oxford Parkinson's Disease Detection (OPD)" and "Parkinson Speech Dataset with Multiple Types of Sound Recordings (PSD)" dataset are to discriminate healthy people from PD. These datasets are obtained from Data Mining Repository of the University of California, Irvine (UCI) [19].

The OPD dataset is composed of a range of biomedical voice measurements from 31 people, 23 with PD. The data set contains 23 attributes and 195 instances obtained from 31 patients [20].

The PSD dataset was created by Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University. The PSD is collected from 40 people, of which 20 patients were healthy and the remaining 20 patients were with PD. The dataset contains multi types of sound recordings and includes 1040 samples for training set and 168 samples for testing set [9]. The training and testing set of OPD dataset are merged for 10-fold cross validation. Therefore, this dataset is redesigned so as to contain 1208 instances and 26 attributes.

Table 1. The specific parameters of the proposed DNN

<table>
<tr><td></td><td></td><td></td><td>OPD</td><td>PSD</td></tr>
<tr><td rowspan="13">Pre-learning</td><td rowspan="5">AE 1</td><td>Num. of Neuron</td><td>4</td><td>4</td></tr>
<tr><td>ρ</td><td>0.15</td><td>0.15</td></tr>
<tr><td>β</td><td>2</td><td>4</td></tr>
<tr><td>λ</td><td colspan="2">0.003</td></tr>
<tr><td>Max iter.</td><td colspan="2">400</td></tr>
<tr><td rowspan="5">AE 2</td><td>Num. of Neuron</td><td>4</td><td>4</td></tr>
<tr><td>ρ</td><td>0.25</td><td>0.5</td></tr>
<tr><td>β</td><td>2</td><td>2</td></tr>
<tr><td>λ</td><td colspan="2">0.003</td></tr>
<tr><td>Max iter.</td><td colspan="2">400</td></tr>
<tr><td rowspan="3">SM</td><td>Class</td><td colspan="2">2</td></tr>
<tr><td>λ</td><td colspan="2">0.003</td></tr>
<tr><td>Max iter.</td><td colspan="2">400</td></tr>
<tr><td rowspan="3">FT</td><td rowspan="3">Back-propagation</td><td>Class</td><td colspan="2">2</td></tr>
<tr><td>λ</td><td colspan="2">0.003</td></tr>
<tr><td>Max iter.</td><td colspan="2">400</td></tr>
</table>

3.2. Simulation Results

Specific tuning parameters of the DNN must be determined for developing an efficient DNN classifier. However, there is no analytical strategy to choose the best values for specific parameters. Therefore, the values of these parameters are heuristically chosen and experimentally validated for the simulation. The specific parameters of the proposed DNN are listed in Table 1.

In order to evaluate and compare the classification achievement of the proposed DNN, the simulations are realized with two different setups. The first setup is performed with 10-fold cross validation to compare the state art methods and to validate the performance of the DNN with statistical analyses. The other setup is also run with %70 training set and %30 testing set of used dataset to compare the performances of the DNN with the performance of methods in the literature. Both runs are performed 30 times with different initializing.

Table 2. The Accuracy rate, sensitivity, and specificity of the OPD data sets for 30 differently 10-fold cross runs

		Methods			
		DNN	SVM	DT	NB
AR	Mean	**86.095**	85.780	84.371	69.644
	Std	**0.476**	0.560	1.175	0.599
Sens.	Mean	58.27	47.639	69.014	91.526
	Std	3.004	3.888	4.550	2.368
Spec.	Mean	95.387	98.643	89.766	62.536
	Std	0.675	0.577	1.539	0.763

Table 3. The Accuracy rate, sensitivity, and specificity of the PSD data sets for 30 differently 10-fold cross runs

		Methods			
		DNN	SVM	DT	NB
AR	Mean	**65.549**	65.450	64.520	59.890
	Std	**0.213**	0.221	0.825	0.352
Sens.	Mean	39.943	40.823	59.238	40.890
	Std	1.524	0.456	1.504	0.590
Spec.	Mean	84.998	84.224	68.640	74.289
	Std	1.000	0.338	1.677	0.408

Table 4. The statistical comparison results of Mann Whitney U test for 30 differently 10-fold cross runs over OPD data set

Comparison	Mean Dif.	Z-value	p-value	Sig. ($p<0.05$)
DNN-SVM	0.310	-2.368	0.018	**DNN**
DNN-DT	1.720	-6.013	0.000	**DNN**
DNN-NB	16.450	-6.656	0.000	**DNN**

Table 5. The statistical comparison results of Mann Whitney U test for 30 differently 10-fold cross runs over PSD data set

Comparison	Mean Dif.	Z-value	p-value	Sig. ($p<0.05$)
DNN-SVM	0.100	-1.900	0.057	-
DNN-DT	1.030	-6.003	0.000	**DNN**
DNN-NB	5.660	-6.654	0.000	**DNN**

The evaluation and comparison of the proposed DNN and the-state-art-methods such as SVM, NB and DT are performed for first run setup and the means and standard deviations of their accuracy rate (AR), sensitivity (Sens) and specificity (Spec) are reported for the OPD dataset in Table 2 and the PSD dataset in

Table 3. Moreover, the obtained 30 mean ARs of used methods are sorted and illustrated in Figure 5 and Figure 6 for OPD and PSD datasets, respectively.

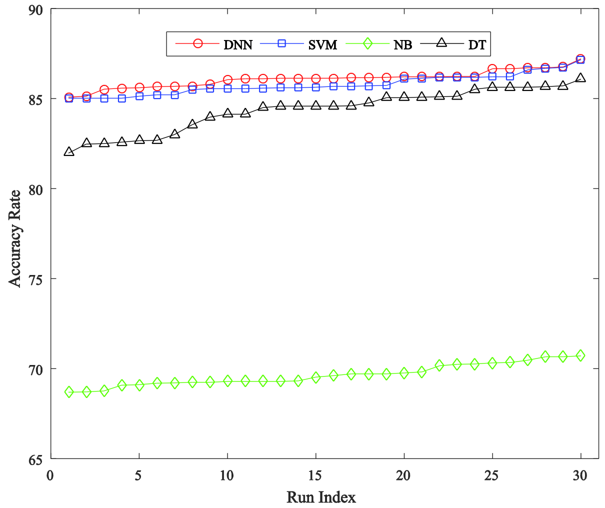

Figure 5. Accuracy rate graphics of 30 differently 10-fold cross runs for OPD data set

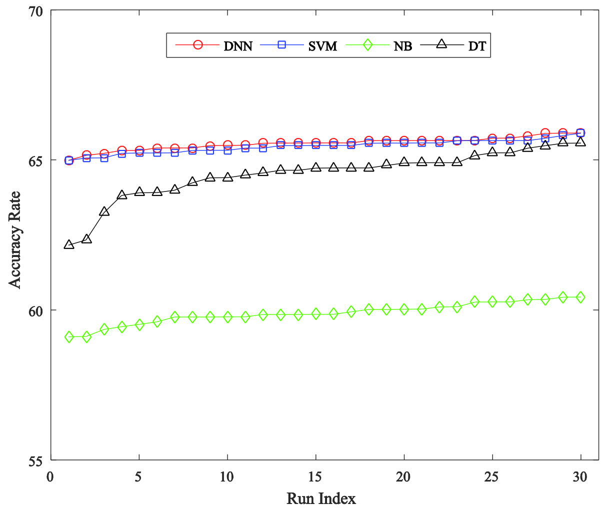

Figure 6. Accuracy rate graphics of 30 differently 10-fold cross runs for PSD data set

When Table 2 and Table 3 are analyzed, it is seen that the proposed DNN, SVM and DT exhibit almost similar performances regarding their mean accuracy rates. Besides, while the performance of the proposed DNN is better than those of SVM and DT, the performance of the NB is worse than the others.

Although the proposed DNN produces better accuracy results than the others, these results should be supported with statistical analyses. Therefore, the Mann Whitney U test is conducted to compare the significance of classification methods to validate this information. The results of the statistical Mann-Whitney U test is reported for the OPD dataset in Table 4 and the PSD dataset in Table 5. The columns of the mean difference and p-value show which one is better among two algorithms in these tables.

When Table 4 is analyzed in terms of statistical significance, it is clearly seen that it has been found statistical significance between compared two algorithms in favor of the proposed DNN for OPD dataset ($p \leq 0.05$). At the same time, Table 5 shows that there is a statistical significance between the proposed DNN and DT, also between the proposed DNN and NB in favor of the proposed DNN for the OPD dataset ($p \leq 0.05$). However, no statistical significance has been found between the proposed DNN and SVM for the PSD dataset ($p > 0.05$).

The second setup is run with %70 training set and %30 testing set for the comparison of the proposed DNN with the previously presented methods in the literature. The AR of the proposed DNN and compared methods are given in Table 6 for the OPD dataset also they are presented in Table 7 for the PSD dataset. These results show that the DNN has superior classification performance, compared with the previous study, which handled the classification problem of the PD over OPD and PSD data sets.

Table 6. The accuracy results of second run setup for the OPD data set

Method	Mean of AR	Method	Mean of AR
The Proposed DNN	**93.79**	DT [27]	84.30
MLP NN [27]	92.90	DES-CS [28]	91.26
DMneural [27]	84.30	SVM [29]	92.75
Regression [27]	88.60	KNN [30]	73.19

Table 7. The accuracy results of second run setup for the PSD data set

Method	Mean of AR	Method	Mean of AR
The Proposed DNN	**68.05**	KNN (k=7) [9]	57.50
KNN (k=1) [9]	55.00	SVM (Linear) [9]	67.50
KNN (k=3) [9]	55.00	SVM (RBF) [9]	65.00
KNN (k=5) [9]	55.00		

4. Conclusion

In this paper, a DNN classifier is proposed for the detection of the speech impairments in PWP for improving the diagnosis of the PD. The results show that the proposed classifier outperforms the other methods in both OPD and PDS databases. The DNN classifier can reduce the dimension of the data with AEs to make efficient classification. The advantages of the proposed classifier can be summarized as follows:

1. The proposed DNN classifier has the ability to extract hidden features, which considerably increases the performance of the classifier.
2. The PD can be remotely diagnosed and monitored using the DNN classifier. Therefore, PWP rarely need to make physical visits to the clinic.
3. As one of the earliest indicators of the PD, the speech impairments may enable us to monitor and diagnose the PWD in vivo and discover reliable biomarkers for identifying the PD at an early stage.
4. The DNN classifier can be used as a reliable classifier for the PD thanks to its efficient specificity and sensitivity accuracy rates

6. References

[1] D. G. Standaert, M. H. Saint-Hilaire, C. A. Thomas "Parkinson's Disease Handbook", American Parkinson Disease Association, New York, USA, 2015.

[2] J. Jankovic, "Parkinson's disease: clinical features and diagnosis", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368-376, 2008.

[3] D. J. Gelb, E. Oliver, S. Gilman, "Diagnostic criteria for Parkinson disease", *Archives of Neurology*, vol. 56, no. 1, pp. 33-39, 1999.

[4] N. Singh, V. Pillay, Y. E. Choonara, "Advances in the treatment of Parkinson's disease", *Progr. Neurobiol.*, vol. 81, pp. 29–44, 2007.

[5] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, 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, April 2009.

[6] A. K. Ho, R. Iansek, C.Marigliani, J. L. Bradshaw, S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease", *Behav. Neurol.*, vol. 11, pp. 131–137, 1998.

[7] J. A. Logemann, H. B. Fisher, B. Boshes, E. R. Blonsky, "Frequency and co-occurrence of vocal-tract dysfunctions in speech of a large sample of Parkinson patients", *J. Speech. Hear. Disord.*, vol. 43, pp. 47–57, 1978.

[8] J. R. Duffy, "Motor Speech Disorders: Substrates, Differential Diagnosis, and Management", Elsevier eBook, 2013.

[9] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, O. Kursun, "Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings", *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828-834, July 2013.

[10] A. Tsanas, M. A. Little, P. E. McSharry, L. O. Ramig, "Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests", IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 884-893, April 2010.

[11] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman and L. O. Ramig, "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264-1271, May 2012.

[12] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", *Nature*, vol. 521, pp. 436-444, 2015.

[13] A. Caliskan, H. Badem, A. Basturk, M. E. Yuksel, "Classification and Diagnosis of Cardiac Arrhythmia Disease by Deep Learning", *International Conference on Artificial Intelligence and Data Processing (IDAP16)*, Malatya, Turkey, 2016, pp. 291-293.

[14] H. Badem, A. Caliskan, A. Basturk, M. E. Yuksel, "Classification and Diagnosis of the Parkinson Disease by Stacked Autoencoder", *10th International Conference on Electrical and Electronics Engineering*, Bursa, Turkey, 2016, pp. 499-502.

[15] H. Badem, A. Caliskan, A. Basturk, M. E. Yuksel, "Classification of Human Activity by Using a Stacked Autoencoder ", *Medical Technologies National Conference (TIPTEKNO'16)*, Antalya, Turkey, 2016, pp.370-273.

[16] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images", *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119-130, Jan. 2016.

[17] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images", *Neurocomputing*, vol. 191, pp. 214-223, 2016.

[18] D. T. Grozdi, S. T. Jovii, M. Suboti, "Whispered speech recognition using deep denoising autoencoder", *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 15 – 22, 2017.

[19] Lichman, M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.

[20] M. A. Little, P. E. McSharry, S. J. Roberts, D. Costello, I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection", *Biomed. Eng.*, vol. 6, no. 23, 2007.

[21] Y. Bengio, "Practical recommendations for gradient-based trainin of deep architectures", Neural Networks: Tricks of the Trade", *Springer*, pp. pp. 437–478, 2012.

[22] M. Ranzato, C. Poultney, S. Chopra, Y. LeCun, "Efficient Learning of Sparse Representations with an Energy-Based Model", *Proceedings of Neural Information and Processing Systems*, 2006.

[23] A. Ng, "Sparse autoencoder", *CS294A Lecture Notes*, 2011.

[24] Q. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, A. Ng, "On optimization methods for deep learning", *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 265–272.

[25] Y. Zhang, E. Zhang, W. Chen, "Deep neural network for halftone image classification based on sparse auto-encoder", *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 245–255, 2016

[26] D. C. Liu, J. Nocedal, "On the limited memory BFGS method for large scale optimization. *Mathematical programming*, vol. 45 no.1,pp. 503-528. 1989.

[27] Das R., "A comparison of multiple classification methods for diagnosis of Parkinson disease", *Expert Systems with Applications*, vol. 37, no 2, pp.1568-1572, 2010.

[28] Woloszynski T., Kurzynski M., "A probabilistic model of classifier competence for dynamic ensemble selection", *Pattern Recognition*, vol. 44, no.10–11, pp.2656-2668, 2011.

[29] Sakar, C. O., Kursun O.. "Telediagnosis of Parkinson's disease using measurements of dysphonia." *Journal of medical systems*, vol.34, no.4, pp. 591-599, 2010.

[30] Polat K. "Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means

clustering". *International Journal of Systems Science* vol. 43 no.4, pp.597-609, 2012.
[31] Benba A., Jilbab A. Hammouch A., "Hybridization of best acoustic cues for detecting persons with Parkinson's disease", *2014 Second World Conference on Complex Systems (WCCS),* Agadir, 2014, pp. 622-625.

Abdullah CALISKAN received B.Sc. degree in 2011 from the Dept. of Electrical-Electronical Engineering at Gaziantep University. He is currently Ph.D. candidate in Erciyes University, Department of Biomedical Engineering, Kayseri, Turkey. His research interests include machine learning in biomedical application.

Hasan BADEM is the corresponding author of this paper. He received B.Sc. degree in 2009 from the Dept. of Education of Computer and Control Technology at Marmara University and M.Sc. degree in 2012 from Education of Computer, Sutcu Imam University. He is currently Ph.D. candidate in Erciyes University, Department of Computer Engineering, Kayseri, Turkey. His research interests include machine learning in biomedical application, parallel and distributed computation.

Alper BASTURK received his BS degree in electronics engineering from Erciyes University, Kayseri, Turkey, in July-1998. He then joined as a research assistant to the Dept. of Electronics Eng. of Erciyes University. He received his MS and PhD degrees in electronics engineering from Erciyes University in August-2001 and November-2006. He is currently working in the Computer Engineering department. His research areas are digital signal and image processing, neural networks, fuzzy and neuro-fuzzy systems, intelligent optimization and applications of these techniques.

M. Emin YUKSEL received the B.S. degree in electronics and communications engineering from Istanbul Technical University, Istanbul, Turkey, in 1990, and the M.S. and Ph.D. degrees in electronics engineering from Erciyes University, Kayseri, Turkey, in 1993 and 1996, respectively. In 2012, he joined the Department of Biomedical Engineering, Erciyes University, where he is currently a Professor. Between March and December 1995, he was an Academic Visitor to the Signal Processing Section, Department of Electrical Engineering, Imperial College, London, U.K. His current research interests include signal processing, image processing, neural networks, fuzzy systems, deep learning, and applications of these techniques.