An Efficient Diagnosis System for Parkinson's Disease Using Deep Belief Network

Ali H. Al-Fatlawi
Computing and Networks Research lab,
University of Kufa, Najaf, Iraq
Ali.H.Mahdi@uokufa.edu.iq

Mohammed H Jabardi
Computing and Networks Research lab,
University of Kufa, Najaf, Iraq
Mohammed.Naji@uokufa.edu.iq

Sai Ho Ling, *SMIEEE*
Centre for Health Technologies,
University of Technology Sydney,
Sydney, Australia
Steve.Ling@uts.edu.au

Abstract*— In this paper, a deep belief network (DBN) has been adopted as an efficient technique to diagnosis the Parkinson's disease (PD). This diagnosis has been established based on the speech signal of the patients. Through the distinguishing and analyzing of the speech signal, the DBN has the ability to diagnose Parkinson's disease. To realize the diagnosis of Parkinson's disease by using DBN, the proposed system has been trained and tested with voices from a number of patients and healthy people. A feature extraction process has been prepared to be inputted to the deep belief network (DBN) which is used to create a template matching of the voices. In this paper, DBN is used to classify the Parkinson's disease which composes two stacked Restricted Boltzmann Machines (RBMs) and one output layer. Two stages of learning need to be applied to optimize the networks' parameters. The first stage is unsupervised learning which uses RBMs to overcome the problem that can cause because of the random value of the initial weights. Secondly, backpropagation algorithm is used as a supervised learning for the fine tuning. To illustrate the effectiveness of the proposed system, the experimental results are compared with different approaches and related works. The overall testing accuracy of the proposed system is 94% which is better than all of the compared methods. In short, the DBN is an effective way to diagnosis Parkinson's disease by using the speech signal.

Keywords— *Parkinson's disease, Deep Belief Network, Restricted Boltzmann Machine.*

I. Introduction

A. The Parkinson's disease

Parkinson's disease (PD) is one of the serious diseases that impacts the central nervous system of the humans and affects most of its functions. Mostly, it targets people after the age of 60 years [1], and it may be subject to genetic and environmental factors [2]. Due to the presence of a significant number of the Parkinson's patients around the world, it is vital to analyze and investigate this disease deeply. In North America, for example, there are about one million people are suffering from PD [3]. There are four primary symptoms of PD which are bradykinesia (slowness of movement), rigidity, tremor, and poor balance [4],[5],[6],[7]. Thus, it is classified within the motor system disorders. Clinically, it breakdowns many of the brain's cells that producing the Dopamine which is the chemical neurotransmitter carrying information from a nerve cell to another. Dopamine is a chemical that allow the brain to interact efficiently in controlling feeling, behavior, awareness, body movement and the speech ability of the humans [8]. In this regard, any breakdown in these nerve cells can lead to an imbalance in vital functions of the Parkinson's patients, including problems with their speeches. Thus, their voices can be soft, rapid, and illegible. Subsequently, vocal impairment is one of the early signs and symptoms of Parkinson's disease [9] because 90% of Parkinson's patients suffer from vocal impairment [10]. Thus, voice measurement is an important factor to diagnosis and follow the Parkinson's patients [11], [12].

When Parkinson's patients performed a set of tests in verbal glibness, major problems noticed in their pronunciation and voice [13]. Thus, one of the obvious signs of this disease is unclear, not understandable speech, and difficult to pronounce the words. Although it has many signs that can classify a Parkinson's patient, there is no definitive test to diagnosis the Parkinson's patients.It may be difficult to diagnose, especially in its early stages. Furthermore, the diagnosis of Parkinson's disease can be based on the patient's medical history, neurological examination capacity and symptoms suffered by the patient [1]

B. Related Works

Analyzing and classification of the patients' speech signals are considered to be an early detection of Parkinson's disease through discriminating characteristic and features of the patients' voices [14]. Recently, many technologies and algorithms have been adopted to diagnosis the disease, including neural networks (NNs) and data-mining algorithms [15]. In this direction, many researches and studies have developed methods to classify the patients' voices and decide whether they are: healthy or sick. [16] addresses that the problem of classifying voices of Parkinson's patients using data-mining methods such as Random Forest, Ada-Boost and

K-NN. It concluded that the K-NN is the best one of these three methods with an overall accuracy of 90.26%. Beside, Support Vector Machine (SVM) is another model that can be used for the classification and the regression problems. It has been used in diagnosing of the Parkinson's disease and examining the people to decide their health status [16]. [17] and [18] compared the results of the SVM with the performance of the NN in handling this issue. [17] found out that the two methods (SVM and NN) have the same classification accuracy (94.8%) while [18] highlights that the SVM can work more efficiency and better accuracy (93%) comparing with NN (86%). Moreover, other researchers have interested in investigating these data to classify the regression of the illness. In that, [19] reproduces the results from [17] and [18] and compares the two methods (SVM and ANN) with AdaBoost model by using binary classification data set. It argues that the last one has a better performance in processing Parkinson's disease, especially in the regressing tasks. The SVM and Genetic Algorithm can work jointly in diagnosing of PD [20]. First, selecting the impacted feature can be achieved by the Genetic Algorithm, and then Support Vector Machine (SVM) has been adopted to classify the patterns [20].

Furthermore, [21] and [22] use a dataset created by Max Little from University of Oxford and National Centre for Voice and Speech that is composed of a range of voice measurements from 31 people (23 with Parkinson's disease (PD) and eight healthy) [21], [23], [24]. They classified this dataset by using three different probabilistic neural networks (PNN) to distinguish the people status. These three PNN procedures are hybrid search (HS), Monte Carlo search (MCS) and incremental search (IS). It can be found that the results in term of accuracy for HS is 81.28%, MCS is 80.92% and IS is 79.78%. Furthermore, four models by applying the same data is conduced [22] which are Decision Tree, Regression, Data Mining Neural (DMneural) and Neural Networks. The accuracies that obtained from these four models are 84.3%, 88.6%, 84.3% and 92.9% respectively. Among of these methods, the most efficiency method is the Neural Network. Lastly, [25] applied this dataset on the nine data-mining algorithms: Bayes Net, Naïve Bayes, Logistic, Simple Logistic, KStar, ADTree, J48, LMT and Random Forest. In their experiment, the best algorithm was for the Random Forest with an accuracy of 90.26%, while the worst one is Naïve Bayes with an accuracy of 69.23%. Although [25] applied many methods to handle the data, its accuracy doesn't exceed 90.26% in the best case. In this work, we will use the same of the Oxford's dataset which has been used in [21], [22], [25] and classify the patterns with a deep belief neural network (DBN) and then will compare the our results with the above studies.

C. *Significance and contributions*

There is a significant number of the people are suffering from Parkinson's disease, and there is no definitive test to diagnosis. Therefore, it is necessary to develop reliable techniques to diagnosis this disease. The purpose of this paper is to build an automatic system that can examine the patients' status in their early stages of illness. The system has been developed based on analyzing the characteristics of the patients voices because the affecting speech is an early sign and symptom of Parkinson's disease. In this paper, DBN has been presented as an efficient solution to handle the features of the voices and classify the patient status with high accuracy compared with other studies.

D. *Paper organization*

This paper is organized as follows: In the previous section, a general background and some related works have been presented, in addition to the significance and contributions of the proposed work. In section II, the method of deep belief network is described. Then, the experimental results and discussions are presented. A conclusion is drawn in Section V.

II. METHODOLOGY

Deep Belief Network (DBN) is a type of deep neural networks that uses multiple processing layers to model high-level abstraction in data with complex structure [26]. These processing layers are connected to each other via connection weights but without any connection within the same layer. Therefore, it is a generative graphical model, consisting of multiple layers of hidden units [27]. Two training stages are required to train this kind of networks: supervised and unsupervised learning.

DBN can be designed as a construction of Restricted Boltzmann Machines (RBM) that are stacked on of each other and trained separately. This stage is unsupervised learning and it called the pre-training stage. It divides the network into groups of stacked sub-networks, each of them includes two processing layers. Basically, this operation is performed to solve the problems that are associated with selecting random values in initializing the connection weights and providing the network with pre-trained weights. It can use the Greedy Layer-Wise unsupervised training algorithm [28] for this function.

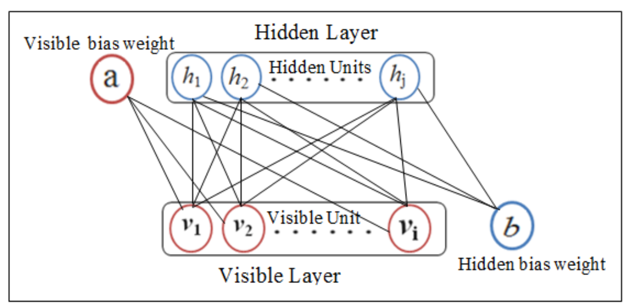

Figure 1: General architecture of RBM

RBM is a generative stochastic of neural network that can be learned based on a probability model by using unsupervised learning technique. As shown in Figure 1, the network of the RBM composes two processing layers: visible and hidden layer. These layers are connected together to allow the construction and reconstruction processes with no connection between the units of the same layer [29]. The visible layer (**v**) consists number of visible units (v_1, v_2 ,. . . . . . v_i) to process the pattern's features which are entering the network as unlabeled data while the hidden layer (**h**) contains number of hidden units (h_1, h_2 ,. . .. . . h_j) with binary value which receive their data from the visible units and can reconstruct them.

All visible units communicate with the hidden units as a bidirectional matrix of weight (**W_{ij}**) associated symmetrically, in addition to the visible bias (a_i) and hidden bias (b_i) [30], [27]. In some classification problems, such as patches of the image and the speech signals, data of the patterns are continuous nonlinear values (not binary). Although DBN is designed to manipulate binary data, it can also deal with the real input data in the visible layer with a particular way. According to Hinton [29], in order to manipulate these real values, independent Gaussian noise can be used instead of the binary visible units. Therefore, the energy function can be represented as shown in equation (1). All input data should be normalized to have zero mean and unit variance.

$$E(v,h) = \sum_{i\in vis} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j\in hid} b_j\, h_j - \sum_{ij} \frac{v_i}{\sigma_i} h_j w_{ij} \qquad (1)$$

where σ_i is the standard deviation of the Gaussian noise for visible unit i.

The learning can be more complicated when both of the visible and the hidden units are Gaussians. In this regard, "the individual activities are held close to their means by quadratic "containment" terms with coefficients determined by the standard deviations of the assumed noise levels" [29]. Then, the energy function is become as shown in equation (2).

$$E(v,h) = \sum_{i\in vis} \frac{(v_i - a_i)^2}{2\sigma_i^2} + \sum_{j\in hid} \frac{(h_j - b_j)^2}{2\sigma_j^2} - \sum_{ij} \frac{v_i}{\sigma_i} \frac{h_j}{\sigma_j} w_{ij} \qquad (2)$$

In this paper, the contrastive divergence (CD) algorithm is used to optimize the network's parameter (connection weights). The following procedure is describing the steps that are followed to implement this algorithm with RBM. In general, the learning in the RBM focuses on calculating the three primary factor: an unbiased sample of hidden unit (negative phase), an unbiased sample of state of a visible unit (positive phase) and final matrix of weight as the follow:

1. Calculating the probabilities of hidden units and sample their vectors from training sample according to (3). Then, computing the outer product of **v** and **h** (positive phase).

$$E(v,h) = \sum_{i\in vis} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j\in hid} b_j\, h_j - \sum_{ij} \frac{v_i}{\sigma_i} h_j w_{ij} \qquad (3)$$

2. Sampling the reconstruction vector **v**' of the visible unit from the vector **h**. Then resampling the hidden activations **h**'. (Gibbs sampling step) as indicated in (4).

$$E(v,h) = \sum_{i\in vis} \frac{(v_i - a_i)^2}{2\sigma_i^2} + \sum_{j\in hid} \frac{(h_j - b_j)^2}{2\sigma_j^2} - \sum_{ij} \frac{v_i}{\sigma_i} \frac{h_j}{\sigma_j} w_{ij} \qquad (4)$$

3. Calculating the outer product of **v**' and **h**' (negative phase).

4. Updating the matrix of weight rule according to (5) :

$$\Delta w_{ij} = \varphi \langle v_i h_j \rangle_{data} - \langle v_i^{'} h_j^{'} \rangle_{model} \qquad (5)$$

where φ is a learning rate.

5. Update the visible bias (a_i) and hidden bias (h_i) as written in (6) and (7).

$$a = a + f(v - v') \qquad (6)$$

$$b = b + f(h - h') \qquad (7)$$

As described so far and shown in Figure 2, DBN is divided into multiple RBM networks. The first RBM is a construction of the visible and the first hidden layer. Then the second RBM is built from communicating the first hidden layer and the second hidden layer and so on.

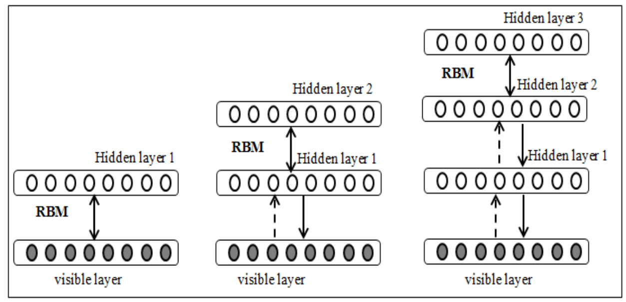

Figure 2: General architecture of DBN

After applying the above algorithm on each RBM in the network and optimizing its parameter, the pre-trained network needs to be fine-tuned by using supervised methods. The back-propagation algorithm is used for the fine tuning, and it can be considered as a second stage of learning [27]. As it is supervised learning mode, it needs to label the data that will pass through the whole network from the visible layer till to output layer. With this algorithm, pairs of training data (input and target vectors) should be provided to the classifier through labeling the data with its corresponding pattern[31],[32].

III. RESULTS AND DISCUSSION

A. Experimental setup

This paper uses data retrieved from a Parkinson's Data Set [27], [30] that are collected by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who developed a tele-monitoring device to record the speech signals. [23] extracted features of the 195 samples of voices that have been recorded from 31 people (23 with Parkinson's disease and eight healthy people). This paper considers 16 attributes (features) for each sample, which is explained in Table I, that have real values. Therefore, we normalized them to have zero mean and unit variance in order to fit the proposed work. Table I illustrates a brief description for the considered attributes. Finally, the status of each pattern is either 1 (for the PD) or 0 for the healthy person.

TABLE I. BRIEF DESCRIPTION OF THE DATASET'S ATTRIBUTES

The attribute	**Description**
MDVP: Fo (Hz)	Average vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
MDVP: Flo(Hz)	Minimum vocal fundamental frequency
MDVP (%)	MDVP jitter as a percentage
MDVP (Abs)	MDVP Jitter absolute jitter in microseconds
Jitter: RAP	Phonation Relative average perturbation, the average absolute difference between a period and the average of it and its two neighbors, divided by the average period
Jitter: PPQ	Phonation Five-point period perturbation quotient, the average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period.
Jitter DDP	Phonation Average absolute difference between consecutive differences between consecutive periods, divided by the average period
Shimmer MDVP	Phonation Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.
MDVP: Shimmer(dB)	This is the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20. MDVP calls this parameter ShdB, and gives 0.350 dB as a threshold for pathology.
Shimmer : APQ3	Phonation Three-point amplitude perturbation quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbors, divided by the average amplitude.
Shimmer : APQ5	Phonation Five-point amplitude perturbation quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbors, divided by the average amplitude.
MDVP : APQ	Phonation Average absolute difference between consecutive differences between the amplitudes of consecutive period.
MDVP: Fo (Hz)	Average vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency

As explained in the preceding sections, this system has been designed by using a DBN and the work concentrates on optimizing that DBN to reach the best results. Therefore, the methodology of the paper is divided into two tasks. The first task is to determine the best structure that can lead to the highest applicable accuracy. Then, the second task is optimizing the connection weights to attained the best applicable classification and lowest error.

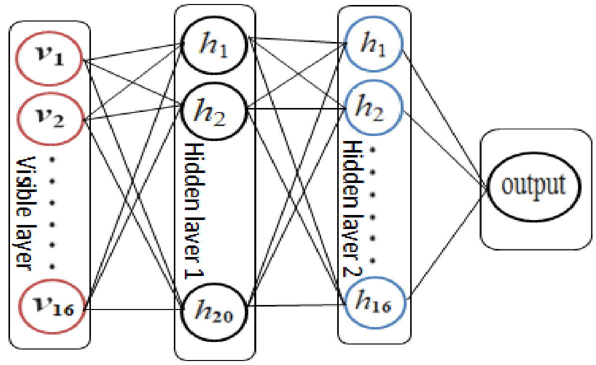

Figure 3: The overall structure of the proposed system

The overall design of the proposed system is illustrated in Figure 3. It can be seen from this figure that network has been conceived as a stack of two Restricted Boltzmann Machines. First, the visible layer is the layer that will receive the input (features of the patterns) to be manipulated in multiple processing layers. By considering 16 of the features that extracted by [23], the number of the units in the visible layer become 16 neurons. As the system will classify the pattern into one of two probabilities, either healthy (0) or patient (1), only one unit is required in the output layer. The layers between the visible and the output are called hidden layers, and they have the major role in processing the data and getting the best classification results.

Deciding the number of the hidden layers and number of units in each of these layers is not an easy task because it is changing based on the problem and type of the data. Although many studies have presented different solutions to determine that, there is no definitive method to decide these values. Hinton [29] points out that the number of the training cases and their dimensionality and redundancy can contribute in determining the number of these units. In this direction, training cases with high redundancy need big sets, and fewer parameters are required because using more parameter can lead to overfitting [29]. In fact, there is no optimal value and the selection often depends on the try and error technique within a particular range. To calculate the number of the hidden layers in the network, many studies suggest that the process may start with one layer then add another and stop before reaching the overfitting while others argue that network with three [33] or two hidden layers are enough to handle most types of the problems.

In the proposed design, structure of two hidden layers is the best choice to present high accuracy with no overfitting. For the first hidden layer, the selected numbers have been chosen from 32 hidden units (double of the number of the visible units) to 8 units (half of the hidden units). Then, the experiments results demonstrates that 20 hidden units in the first layer and 17 units in the second hidden layer are considered to be the best case to attain the higher accuracy.

As it indicated previously, the deep belief network is a stacked Restricted Boltzmann Machines (RBMs). Consequently, the structure of each of the two RBMs is described in Table II.

TABLE II. NUMBER OF UNITS IN EACH RBM

	Number of the visible units	Number of the hidden units
RBM 1	16	20
RBM 2	20	17

Finally, the type activation function that has been used in all of the processing units is a logistical activation function and its function can be written as shown in (8)

$$f(\mathrm{x})=\frac{1}{1+e^{-x}} \quad (8)$$

B. Results

To illustrate the performance of the proposed method by using the data set. We divided the sample into two groups of data. The first group is the training sample that are accounted for 74% of the total samples. The remaining samples will be used to test and validate the system and examine it accuracy.

In the first training step, unsupervised learning operation takes its place in optimizing the two Restricted Boltzmann Machines (RBM) that are stacked on top of each other. After 25000 iterations of the unsupervised training with unlabled training set, the Restricted Boltzmann Machines provide trained initial weights for the connection lines between the visible layer and the first hidden layer (W1) and also between the first hidden layer and the second hidden layer (W2). During the training the process and after learning the first RBM, its output enters as an input to the second RBM. By means, the two network has been trained asynchronously. Therefore, they need to be tuned to handle the input data in efficient mode by using supervised learning methods. The backpropagation has been used to tune the two RBMs with each other and also to tune them with output layer. Mean Square Error (MSE) that is explained in (9) is used as a measure for the performance of the network during the training procedure to reach the minimum error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (9)$$

Where $\hat{y}$ is a vector of actual data, and $\hat{y}_i$ is the vector of desired data.

After 8700 iterations of a supervised learning, the MSE reached its steady state rate. The number of the neurons in each hidden layers one of the uncertain issues and has a significant impact on the efficiency of the classification operation. Different numbers of hidden units have been applied and tested till to reach the optimal values as shown in Table III. Learning rate (0.001) in general is small enough to ensure convergence.

During the testing phase, the system has been examined by using the another set of patterns (testing set). It succeeded in classifying 47 patterns out of 50 in the testing set and failed with three patterns only. Therefore, the overall accuracy of the system is 94%. This percent is good enough to generate reliable system able to diagnosis the patients.

TABLE III. THE ACCURACY OF THE SYSTEM WITH DIFFERENT UNITS NUMBER

	Number of the neurons			The accuracy
	Visible layer	***Hidden layer 1***	***Hidden layer 2***	***Hidden layer 1***
1	16	10	8	88%
2	16	16	15	92%
3	16	20	15	90%
4	16	20	17	94%
5	16	25	19	92%
6	16	30	20	89%

C. Discussions

In This paper aims to produce an automatic system to diagnosis the PD and specify their health status. We use a trusted dataset that have been published on UCI website. Many researchers worked on this dataset and applied it in classification and regressing processes. By comparing the results of this work with other papers that utilized the same data, this paper achieves the best performance and accuracy. In [21], three probabilistic neural networks (PNN) have been used to classify that data. Its models are hybrid search (HS), Monte Carlo search (MCS) and incremental search (IS). The highest result for that paper doesn't exceed 81.28%. Moreover, eight algorithms have been applied to classify the data by [25], but their accuracies were below 90.26%, while the accuracy of [25] reached to 92.9% by using the neural networks with the same data. On the other hand, the proposed system surpasses all of these works by achieving 94% accuracy when it tested with the validation and test set. Table IV compares the accuracies that resulted from of the preceding papers with our apporach.

TABLE IV. COMPARISON OF THE ACCURACY BETWEEN THE PROPSED WORK AND PREVIOUS WORKS BY APPLYING THE SAME DATASET

Article	Method	Accuracy
[25]	IS	79.78%
	MCS	80.92%
	HS	81.28%
[22]	DMNeural	84.3%
	Decision tree	84.3%
	Regression	88.6%
	Neural network	92.9%
[25]	Bayes Net	80.00%
	Naïve Bayes	69.23%
	Logistic	83.66%
	Simple Logistic	84.61%
	KStar	89.74%
	ADTree	86.15%
	J48	80.51%
	LMT	86.15%
	Random Forest	90.26%
The Propsed work	**DBN**	**94.00%**

IV. CONCLUSION

In this paper, an efficient diagnosis system for Parkinson's disease using deep belief neural network (DBN) is presented. Through to recognize the voice of patients, the onset of Parkinson's disease can be diagnosed. The result indicates that the deep belief network gives an improvement for diagnosis with 94% accuracy. This result shows that DBN has succeeded in reach the highest accuracy with that dataset.

As a future work, this can be developed to give a score for Parkinson's patients that can measure the degree of the harming instead of only classifying them into healthy and patient.

REFERENCES

[1] Van Den Eeden, S.K., Tanner, C.M., Bernstein, A.L., Fross, R.D., Leimpeter, A., Bloch, D.A., and Nelson, L.M.: 'Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity', American Journal of Epidemiology, 2003, 157, (11), pp. 1015-1022

[2] Lesage, S., and Brice, A.: 'Parkinson's disease: from monogenic forms to genetic susceptibility factors', Human Molecular Genetics, 2009, 18, (R1), pp. R48-R59

[3] Lang, A.E., and Lozano, A.M.: 'Parkinson's Disease', New England Journal of Medicine, 1998, 339, (15), pp. 1044-1053

[4] Cunningham, L., Mason, S., Nugent, C., Moore, G., Finlay, D., and Craig, D.: 'Home-Based Monitoring and Assessment of Parkinson's Disease', Information Technology in Biomedicine, IEEE Transactions on, 2011, 15, (1), pp. 47-53

[5] Dastgheib, Z.A., Lithgow, B., and Moussavi, Z.: 'Diagnosis of Parkinson's disease using electrovestibulography', Medical & Biological Engineering & Computing, 2012, 50, (5), pp. 483-491

[6] Marino, S., Ciurleo, R., Di Lorenzo, G., Barresi, M., De Salvo, S., Giacoppo, S., Bramanti, A., Lanzafame, P., and Bramanti, P.: 'Magnetic resonance imaging markers for early diagnosis of Parkinson's disease', Neural Regeneration Research, 2012, 7, (8), pp. 611-619

[7] Rigas, G., Tzallas, A.T., Tsipouras, M.G., Bougia, P., Tripoliti, E.E., Baga, D., Fotiadis, D.I., Tsouli, S.G., and Konitsiotis, S.: 'Assessment of Tremor Activity in the Parkinson’s Disease Using a Set of Wearable Sensors', Information Technology in Biomedicine, IEEE Transactions on, 2012, 16, (3), pp. 478-487

[8] Jankovic, J., and Tolosa, E.: 'Parkinson's Disease and Movement Disorders' (Lippincott Williams & Wilkins, 2007. 2007)

[9] Duffy, J.: 'Motor speech disorders: Substrates, differential diagnosis, and management. St. Louis, MO: Mosby-Year Book', in Editor (Ed.)^(Eds.): 'Book Motor speech disorders: Substrates, differential diagnosis, and management. St. Louis, MO: Mosby-Year Book' (Inc, 2005, edn.), pp.

[10] Logemann, J.A., Fisher, H.B., Boshes, B., and Blonsky, E.R.: 'Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients', Journal of Speech and Hearing Disorders, 1978, 43, (1), pp. 47-57

[11] Rahn, D.A., Chou, M., Jiang, J.J., and Zhang, Y.: 'Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis', Journal of Voice, 2007, 21, (1), pp. 64-71

[12] Sapir, S., Spielman, J.L., Ramig, L.O., Story, B.H., and Fox, C.: 'Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings', Journal of Speech, Language, and Hearing Research, 2007, 50, (4), pp. 899-912

[13] Caballol, N., Martí, M.J., and Tolosa, E.: 'Cognitive dysfunction and dementia in Parkinson disease', Movement Disorders, 2007, 22, (S17), pp. S358-S366

[14] Froelich, W., Wróbel, K., and Porwik, P.: 'Diagnosing Parkinson's disease using the classification of speech signals', Journal of Medical Informatics & Technologies, 2014, 23, pp. 187--194

[15] Khemphila, A., and Boonjing, V.: 'Parkinsons disease classification using neural network and feature selection', World Academy of Science, Engineering and Technology, 2012, 64, pp. 15-18

[16] Bhattacharya, I., and Bhatia, M.P.S.: 'SVM classification to distinguish Parkinson disease patients'. Proc. Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, Coimbatore, India2010 pp. Pages

[17] Gil, D., and Johnson, M.: 'Diagnosing parkinson by using artificial neural networks and support vector machines', Global Journal of Computer Science and Technology, 2009, 9, (4), pp. 63-71

[18] Mandal, I., and Sairam, N.: 'Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system', International journal of medical informatics, 2013, 82, (5), pp. 359-377

[19] Hashmi, S.F.: 'A Machine Learning Approach to Diagnosis of Parkinson's Disease', 2013

[20] Shahbakhi, M., Far, D.T., and Tahami, E.: 'Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine', Journal of Biomedical Science and Engineering, 2014, 2014

[21] Ene, M.: 'Neural network-based approach to discriminate healthy people from those with Parkinson's disease', Annals of the University of Craiova-Mathematics and Computer Science Series, 2008, 35, pp. 112-116

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

[23] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., and Ramig, L.O.: 'Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease', Biomedical Engineering, IEEE Transactions on, 2009, 56, (4), pp. 1015-1022

[24] Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A., and Moroz, I.M.: 'Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection', BioMedical Engineering OnLine, 2007, 6, (1), pp. 23

[25] Sriram, T.V., Rao, M.V., Narayana, G.S., Kaladhar, D., and Vital, T.P.R.: 'Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms', International Journal of Engineering and Innovative Technology (IJEIT) Volume, 2013, 2, pp. 44-52

[26] Deng, L., and Yu, D.: 'Deep learning: methods and applications', Foundations and Trends in Signal Processing, 2014, 7, (3–4), pp. 197-387

[27] Hinton, G.E., and Salakhutdinov, R.R.: 'Reducing the Dimensionality of Data with Neural Networks', Science, 2006, 313, (5786), pp. 504-507

[28] Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H.: 'Greedy layer-wise training of deep networks', Advances in neural information processing systems, 2007, 19, pp. 153

[29] Hinton, G.: 'A practical guide to training restricted Boltzmann machines', Momentum, 2010, 9, (1), pp. 926

[30] Hinton, G.E.: 'Training Products of Experts by Minimizing Contrastive Divergence', Neural Computation, 2002, 14, (8), pp. 1771-1800

[31] Al-Fatlawi, A., Ling, S., and Lam, H.: 'A Comparison of Neural Classifiers for Graffiti Recognition', Journal of Intelligent Learning Systems and Applications, 2014, 2014

[32] Jabardi, M.H.N., and Kaur, H.: 'Artificial Neural Network Classification for Handwritten Digits Recognition', International Journal of Advanced Research in Computer Science, 2014, 5, (3)

[33] Hinton, G.E., Osindero, S., and Teh, Y.-W.: 'A fast learning algorithm for deep belief nets', Neural computation, 2006, 18, (7), pp. 1527-1554