# Application of Neural Networks in Early Detection and Diagnosis of Parkinson's Disease

[1]Rashidah. Funke Olanrewaju, [1]Nur Syarafina Sahari, [2]Aibinu A. Musa and [3]Nashrul Hakiem
[1]Department of Electrical and Computer Engineering, Faculty of Engineering,
International Islamic University Malaysia, Kuala Lumpur, Malaysia.
[2]Federal University of Technology Minna, Niger State Nigeria.
[3]Department of Informatics Engineering Faculty of Science & Technology,
UIN Syarif Hidayatullah Jakarta.

***Abstract*— Parkinson's disease (PD) is a chronic neurological progressive disorder caused by lack of the chemical dopamine in the brain. Up to today, there is still no cure or prevention for PD, and usually the disease worsens gradually over time. However, this disease can be controlled with some treatment, especially in the early stage. Hence, this study proposes a method in early detection and diagnosis of PD by using the Multilayer Feedforward Neural Network (MLFNN) with Back-propagation (BP) algorithm. This MLFNN with BP algorithm is simulated using MATLAB software. The dataset information used in this study was taken from the Oxford Parkinson's Disease Detection Dataset. The output of the network is classified into healthy or PD by using K-Means Clustering algorithm. The performance of this classifier was evaluated based on the three parameters; sensitivity, specificity and accuracy. The result shows that network can be used in diagnosis and detection of PD due to the good performance, which is 83.3% for sensitivity, 63.6% for specificity, and 80% for accuracy.**

## I. INTRODUCTION

Parkinson's disease (PD) is the second most regular neurodegenerative disorder after Alzheimer's disease [1]. PD occurs due to neurological disorder in the brain [2]. In the brain's thalamic region, there is an area called substantia nigra, which has the neurons that produce a vital brain chemical known as dopamine. Dopamine is a neurotransmitter that helps in the transmission of signals in the brain and other vital areas. Fig. 1 shows the degeneration of the substantia nigra of the brain.

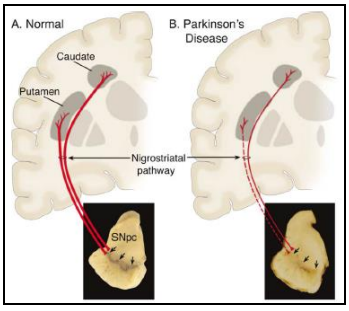


Fig. 1: Degeneration of the substantia nigra of the brain [3]

Fig. 1(A) shows the schematic representation of the normal nigrostrial pathway (red pathway)[3]. This nigrostrial pathway is made of dopaminergic neurons. In the Fig. 1(B), it shows the schematic representation of the nigrostriatal pathway (in red) of Parkinson's patient has degenerated.

The symptoms of PD usually related to the movement, such as "shakes" (tremor), slowness of movements, muscle stiffness and the symptoms may advance to imbalance (postural instability). Besides the motor symptoms, other symptoms are fatigue, anxiety, depression, slowness of thinking, difficulty concentrating, visual hallucinations, and sleep disturbances [4].

PD usually worsens gradually over time and currently there is no cure or prevention for it. However, it can be controlled with some treatment, especially at the early stage. Hence early detection is important. The common treatment given to the patient is based on restoring dopamine in the brain and deep brain stimulation (DBS) surgery [5].

The diagnosis of PD usually based on the medical history and neurological investigations. However, diagnosis is less accurate because of the various indicators and their similarities with other neurodegenerative diseases, such as progressive supranuclear palsy and multiple system atrophy [6]. The diagnosis also fails to identify this disease before the patient had a significant loss of dopamine neurons. Therefore, main objective of this research is to improve the diagnostic accuracy, and to find the best method to detect the disease in the early stage. It is very important to diagnose PD accurately since if the patient is misdiagnosed as healthy, his condition will worsens over time. In an effort to improve the diagnostic accuracy, the application of Artificial Neural Network (ANN), which is a computational structure paradigm modeled on the biological process that is inspired by the way biological nervous systems, such as the brain, processes information is used. Neural networks are appropriate tools for classification of nonlinear data compared to linear statistical models of classification [7].

The organization of the paper is as follows; Section. II gives a detail analysis of the proposed system while discussion of the result obtained on application of the system on public available PD is discussed in section III and section IV

concludes the paper.

## II. MATERIALS AND METHOD

The block diagram of the proposed method for Parkinson detection is shown in Fig. 2. It consists of a three step cascade system, namely Data Formatting section, MLFFNN with BP algorithm section and K-Means Clustering section.

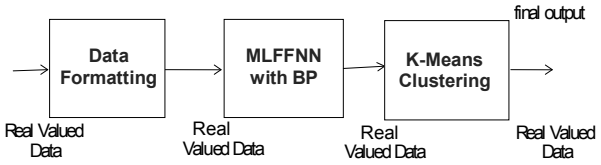


Fig. 2: Automatic Parkinson Detection Block Diagram

Fig. 3 shows the flowchart used in this research to detect and classification of PD. MATLAB software has been used to simulate the neural networks, K-Means Clustering, and performance evaluation.

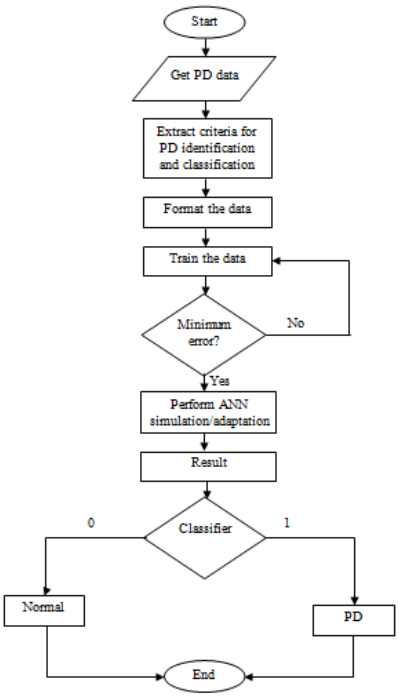


Fig. 3: Flowchart of PD detection and classification

### A. Data Formatting

The PD database used in this project is from the Oxford Parkinson's Disease Detection Dataset [8].

This dataset consists of biomedical voice measurements that were recorded from 31 people, with 23 of them diagnosed with PD. Each column represents a particular voice measure, and each row corresponds one of the 195 voice recording from the 31 people. There are about six recordings for each person and for each recordings, 22 attributes of the voice are measured. There is a "status" column which shows whether the individual is healthy (0) or PD (1).

In this research, only eight attributes have been chosen as the input of the ANN. The attributes used in this research are related to the frequencies of the voice. Since tremor is one of the symptoms of PD, the various measures of the frequencies are can be used as the input of ANN. The attributes used are MDVP:Fo (Hz), MDVP:Fhi (Hz), MDVP:Flo (Hz), MDVP:Jitter (%), MDVP:Jitter (Abs), MDVP:RAP, MDVP:PPQ, and Jitter:DDP.

### B. MLFNN with BP Algorithm

For the next section, the architecture used is the Multilayer Feed-forward Neural Networks (MLFFNN) that is trained with Back-propagation (BP) algorithm. Assuming the data inputs are represented by $X_i$ and weights by $W$, therefore the detail explanation can be described based on the Fig. 4.

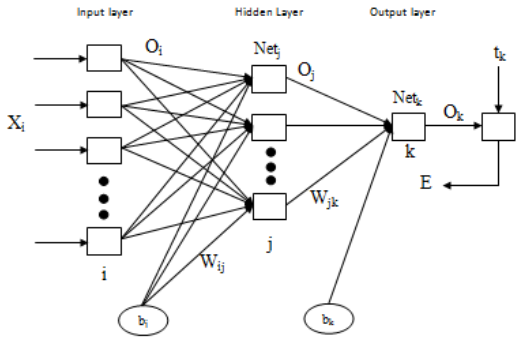


Fig. 4: Multilayer feed-forward with back-propagation algorithm

The neurons in each layer are fully connected to the neurons in the next layer, from layer $i$ to $j$ to $k$. Supposed that the network is designed with only one hidden layer neurons and generate only one output. $W_{ij}$ is the weight that connects the $i^{th}$ neuron from input layer to the $j^{th}$ neuron in the output layer, whereas $W_{jk}$ is the weight that connects the $j^{th}$ neuron from input layer to the $k^{th}$ neuron in the output layer. In the BP algorithm, the generalization of delta rule involves 2 phases, which are the forward phase and the backward phase [9].

**The forward phase:** For hidden layer output, consider

$$O_j = \Phi\left(Net_j\right) \tag{1}$$

where

$$Net_j = \sum_j W_{ij} O_i + b_i \tag{2}$$

where $b_j$ is the bias of the hidden node and can be set to zero, $\Phi$ is the sigmoid activation function. For output layer $k$, the network output is given as;

$$O_k = \Phi(Net_k) \tag{3}$$

where

$$Net_k = \sum_k W_{jk} O_j + b_j \tag{4}$$

**The backward phase between output and hidden layer:** The backward phase includes the calculation of the signal error and the weight update of the network. The network error '*E*' is developed as follow:

$$E = t_k - O_k \tag{5}$$

where $t_k$ is the target output and $O_k$ is the output of network which the output of the output layer. The objective function is to minimize the sum squared of error function, where the average sum squared error of the network is as given;

$$E = \frac{1}{N}\sum_{k=1}^{N}(t_k - O_k)^2 = \frac{1}{N}\sum_{k=1}^{N}E^2 \tag{6}$$

where $N$ is the total number of training pattern, $E$ is the error function to be minimized.
The network weight update between the hidden layer $j$ and output layer $k$ is given by;

$$W_{jk}^{new} = W_{jk}^{old} + \Delta W_{jk} \tag{7}$$

where

$$\Delta W_{jk} = \eta \nabla E|_{W_{jk}} \tag{8}$$

$\eta$ is learning rate, $\nabla$ is the gradient.
Therefore, (8) can be re-written as a partial derivative given by:

**The backward phase between hidden layer and input layer:** Adjusting between hidden layer and the input layer by:

$$W_{ij}^{new} = W_{ij}^{old} + \Delta W_{ij} \tag{9}$$

where $\Delta W_{ij}$ is the rate of change in weight.
Therefore the new weight update is;

$$W_{ij}^{new} = W_{ij}^{old} + \eta \delta_j O_i \tag{10}$$

### C. K-Means Clustering

The output of ANN is classified into two groups by using K-means Clustering. K-means algorithm classifies items into k groups [10]. The grouping is done by minimizing the sum of squared distances (Euclidean distances) between items and the corresponding centroid. K-Means Clustering is as follows:

1. Set the value of *k*
2. Choose the initial centroid randomly
3. Find the Euclidean distance between each point and the centroids, and assign the point into the group that has closest distance
4. Compute the new centroid for each cluster
5. Repeat steps 3 – 4 until the value of the centroids did not change

### D. Evaluation Criteria

The performance of the classifier is evaluated based on three aspects; sensitivity, specificity, and accuracy. Sensitivity measures the predicted output with respect to the change in input. In other words, sensitivity shows the ratio of the true positives that are correctly identified. This is contrast with specificity, which measures the ratio of true negatives that are correctly identified. The relationship between the predicted value and the actual value is called accuracy. Accuracy measures how close the predicted value to the actual value.
Besides that, the Receiver Operating Characteristic (ROC) graph is also used to measure the performance of this classifier. ROC graph is often used to assess the performance of a classifier. It is a useful and clear possibility for organizing classifiers and measuring their performance. ROC graph plots the relationship between True Positive Rate (TPR) and False Positive Rate (FPR). TPR is equal to the sensitivity of the classifier whereas FPR is equal to the (1-specificity) of the classifier. The area under the ROC curve (AUC) is a measure of the classifier accuracy [11].

## III. RESULTS AND DISCUSION

In this research, Multi-layer Feed-forward Neural Network (MLFNN) with Back-Propagation (BP) is applied. The back-propagation learning algorithm with single hidden layer of MLFNN has been used. The input layer consists of 8 nodes and the hidden layer consists of 10 nodes. There is only a single output of this network which gives the real-valued output between 0 and 1. The activation function used in this research is the log sigmoid for both hidden and output layer. The initial weights were chosen randomly. 65 datasets have been fed into the network for the training phase. For the testing phase, another 65 datasets have been fed into the network. Fig. 5 shows the error for each epoch, or iteration in the training phase. The graph shows that the mean squared error is decreasing. This shows that the network is learning. After about 50 epochs, the mean squared error slightly decreasing until 500 epochs. This shows that the network has been optimized.

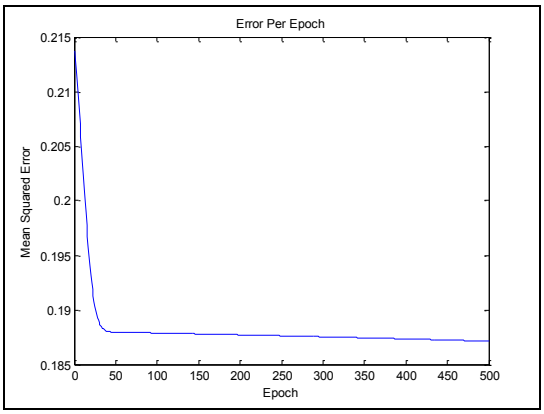


Fig. 5: Error per epoch

Then, the network was tested using another 65 set of data. The output of the network was classified into two groups by using K-means Clustering. The output of the network has the minimum value of 0.7394 and maximum value of 0.7721. This set of output has been classified into Cluster One and Cluster Two based on the distance of each point to the mean of these clusters. Table 1 shows the mean, maximum and minimum values of each cluster, and Fig. 6 shows the scatter plot of each point. This figure shows that the output has been classified into two groups; Cluster One and Cluster Two. Cluster One is for people with PD and Cluster Two is for people who is healthy. The red dot is the centroid of Cluster One and the red asterisk is the centroid for Cluster Two.

Table 1: Mean, Maximum and Minimum Values of each cluster

| | **Mean Value** | **Minimum Value** | **Maximum Value** |
|---|---|---|---|
| **Cluster One** | 0.7456 | 0.7394 | 0.7532 |
| **Cluster Two** | 0.7614 | 0.7563 | 0.7721 |

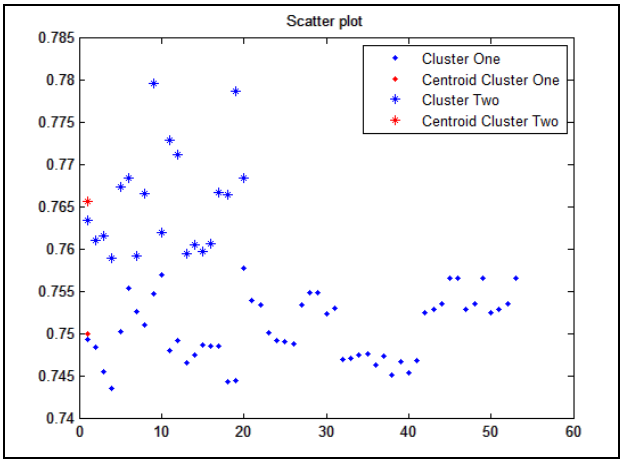


Fig. 6: Scatter plot of each point

The performance of this classifier is evaluated based on the sensitivity, specificity, and accuracy. These three aspects are measured using the following formulae [12]:

$$Sensitivity\ (\%) = \frac{True\ \ Positives}{True\ \ Positives + False\ \ Negatives} \times 100 \quad (11)$$

$$Specificity\ (\%) = \frac{True\ \ Negatives}{True\ \ Negatives + False\ \ Positives} \times 100 \quad (12)$$

$$Accuracy\ (\%) = \frac{True\ \ Positives + True\ \ Negatives}{Total\ \ Number\ \ of\ \ Samples} \times 100 \quad (13)$$

where the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are explained in Table 2.

Table 2: Abbreviations for performance parameters

| **Abbreviation** | **Explanation** |
|---|---|
| True Positive (TP) | The number of people who actually have PD and diagnosed with PD<br>➢ Target Output = 1<br>➢ Network Output = 1 |
| True Negative (TN) | The number of people who actually healthy and diagnosed as healthy<br>➢ Target Output = 0<br>➢ Network Output = 0 |
| False Positive (FP) | The number of people who actually healthy but diagnosed with PD<br>➢ Target Output = 0<br>➢ Network Output = 1 |
| False Negative (FN) | The number of people who actually have PD but diagnosed as healthy<br>➢ Target Output = 1<br>➢ Network Output = 0 |

The performance of the classifier can be analyzed based on Table 4 and Fig. 8. Table 3 shows the percentages of sensitivity, specificity, and accuracy of this classifier whereas Fig. 8 shows the ROC graph of this classifier.

Table 3: Performance of the classifier

| **Performance Parameters** | **Percentage (%)** |
|---|---|
| Sensitivity | 83.3 |
| Specificity | 63.6 |
| Accuracy | 80.0 |

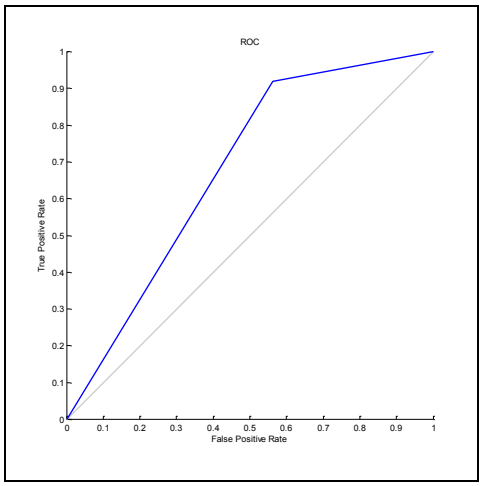


Fig. 7: ROC Graph

From on Table 4, the sensitivity, specificity, and accuracy of this classifier is 83.3%, 63.6%, and 80.0%, respectively. This shows that in terms of sensitivity and accuracy, this network performs very well, and in terms of specificity, this network can be considered good since the percentage is relatively high. Hence, this method is accurate to detect and classify Parkinson's disease. The accuracy of neural network can be improved by initializing the network and train the network again. Each time the training is executed, different parameters are generated and it might produce different solutions. The other approach is by increasing the number of hidden neurons. Larger number of neurons in the hidden layer makes the network more flexible because the network has more parameters it can optimize. Besides that, increasing the number of training data can also improve the network. The additional data makes the network generalizes better to new data.

In Fig. 8, the ROC curve shows the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR). TPR is equal to the sensitivity and FPR is equal to the (1-specificity). The accuracy can be measured from the area under the curve. If the area under the curve is larger, the network is more accurate. This ROC graph was plotted using MATLAB function *plotroc*. This *plotroc* function plot the ROC curve based on the network output and the target output. From this ROC curve, we can conclude that this network is accurate [11].

## IV. CONCLUSION

Neural networks for detection and classification of Parkinson's disease (PD) has been applied in this research. Multi-layer Feed-forward Neural Networks (MLFNN) with Back-Propagation algorithm has been applied on Patient dataset to classify PD and healthy patient. The output of this MLFNN is classified into two groups; Cluster One which is PD and Cluster Two which is healthy, by using K-Means Clustering. Lastly, the performance of the classifier is measured based on the three parameters; sensitivity, specificity, and accuracy. This study proves that the MLFNN with BP algorithm can be used in diagnosis and detection of PD due to the good performance, which is 83.3% for sensitivity, 63.6% for specificity, and 80% for accuracy. The ROC graph proves that this network is accurate because of the large area under the graph.

However, in order to apply this network in the real life, more detailed study should be done.

## REFERENCES

[1] L. Shen-Yang, S. D. Puvanarajah, and N. M. Ibrahim. "Parkinson's Disease: Information for People Living with Parkinson's." Novartis Corporation (Malaysia) Sdn. Bhd. and Orient Europharma (M) Sdn. Bhd., 2011.

[2] R. Arvind, B. Karthik, N. Sriraam, and J. K. Kannan. "Automated detection of PD resting tremor using PSD with recurrent neural network classifier," in *2010 International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom),* 2010, pp. 414-417.

[3] W. Dauer, and S. Przedborski. "Parkinson's disease: mechanisms and models." *Neuron*, vol. 39, no. 6, pp. 889-909, 2003.

[4] V. Giuliano. "A stem cell cure for Parkinson's Disease – so close and yet so far away." Internet: http://www.anti-agingfirewalls.com accessed June 30, 2012 [October 28, 2013].

[5] A. Dovzhenok, C. Park, R. M. Worth, and L. L. Rubchinsky. "Failure of delayed feedback deep brain stimulation for intermittent pathological synchronization in Parkinson's disease." *PloS one*, vol. 8, issue 3, pp. 1-10, 2013.

[6] S. S. Ahmed, W. Santosh, S. Kumar, and H. T. Christlet. "Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection." *J Biomed Sci*, vol. 16, no. 63, 2009.

[7] R. Das. "A comparison of multiple classification methods for diagnosis of Parkinson disease." *Expert Systems with Applications*, vol. 37, pp. 1568-1572, 2010.

[8] "UCI Machine Learning Repository: Parkinsons Data Set." Internet: http://archive.ics.uci.edu/ml/datasets/Parkinsons, June 26, 2008 [December 4, 2013].

[9] R. F. Olanrewaju, O. Khalifa, and K. N. A. Latif. "Computational Intelligence: It's Application in Digital Watermarking." *Middle-East J. Sci. Res.*, vol. 13, pp 25-30, 2013.

[10] A. K. Jain. "Data clustering: 50 years beyond K-means." *Pattern Recognition Letters* , vol. 31, no. 8, pp. 651-666, 2010.

[11] Z.-C. Qin. (2005). "ROC analysis for predictions made by probabilistic classifiers," in *Proceedings of 2005 International Conference on Machine Learning and Cybernetics. 5*, pp. 3119-3124, 2010. IEEE.

[12] N. Gupta, A. Rawal, V. L. Narasimhan, and S. Shiwani. "Accuracy, Sensitivity and Specificity Measurement of Various Classification Techniques on Healthcare Data." *Journal of Computer Engineering* , vol. 11, no. 5, pp. 70-73, 2013.