# An LSTM based Deep learning model for voice-based detection of Parkinson's disease



5 authors, including:

Danish Raza Rizvi
Jamia Millia Islamia


Iqra Nissar
Jamia Millia Islamia


Sarfaraz Masood
Jamia Millia Islamia


Faiyaz Ahmad
Jamia Millia Islamia

International Journal of Advanced Science and Technology
Vol. 29, No. 5s, (2020), pp. 337-343

# An LSTM based Deep learning model for voice-based detection of Parkinson's disease

Danish Raza Rizvi[1], Iqra Nissar[2], Sarfaraz Masood*[3], Mumtaz Ahmed[4], Faiyaz Ahmad[5]

[1,2,3,5]*Department of Computer Engineering, Jamia Millia Islamia, New Delhi-110025*
[4]*Department of Civil Engineering, Jamia Millia Islamia, New Delhi-110025*

***Abstract***

*Parkinson's disease (PD), a chronic neuropathological disorder, occurs when certain brain cell clusters are not apt to engender dopamine. As a result, people feel difficulty in writing, speaking, walking and performing various other activities. Numerous research investigations have shown that voice impairment is the most underlying symptoms found in the number of Parkinson's disease patients. In this work, we attempt to explore the possibilities of a deep neural network (DNN) and long short-term memory (LSTM) network-based model for predicting Parkinson's disease using a subject's voice samples. The various simulations were performed on the dataset to exhibit the efficacy of the models along with their comparison to the conventional machine learning techniques. The results obtained show high values of various metrics including an accuracy of 97.12% and 99.03% for DNN and LSTM respectively which strongly suggest their efficiency for the detection of PD.*

**Keywords**: *Deep Neural Networks, Long Short-Term Memory Network, Parkinson's disease detection, Voice signals*

## I. INTRODUCTION

PD is a neurological disorder which deteriorates the motor functions [16] as well as non-motor functions of the body which includes speech disorders [17]. Dr. James Parkinson describes it as a shaking palsy [18]. Out of the world's total population, over 10 million people are diagnosed with PD as per the survey done by American Parkinson Disease Association (APDA) [28]. It is commonly seen in elder aged people; whose age is over 60 [28]. The cause, as well as cure of Parkinson's disease, is yet unknown [20, 23], but it can be treated through medication during its earlier stages which offers a significant mollification of symptoms [28]. The primary symptoms of this disease are difficulty in movement, stiffness in body parts, poor balance [4] [6] [24] tremor, voice impairment, and bradykinesia.

Researchers have shown that 90% of people with Parkinson have speech problems and vocal impairment [12] [22] which are the earliest indicators for the disease. The vocal tremor, monotone, hoarseness, reduced loudness, breathiness and imprecise articulation [11] are few of the vocal impairment symptoms. Using the continuous vowel phonations or running speech, the level of local impairment can be evaluated [1] [7]. The symptoms of this disease are instigated by the death of certain brain cells that produce neurotransmitters such as dopamine, serotonin, and acetylcholine [26].

The diagnosis of PD from voice impairments is very popular because the telemonitoring and telediagnosis systems that are established on voice signals are very economical and can be used easily, and thus the physical visits of PD subjects to clinics are lowered [20] [21] which in turn reduces the workload on medical personnel. The PD diagnosis generally consists of three steps that include data pre-processing, extraction of features and classification [24] [27]. In the data pre-processing step, the segmentation of speech signals with time windows is performed. In the feature extraction step, the extraction of several features is done from each pre-processing segment. This is concluded by a classification process as the last step.

## II. LITERATURE REVIEW

Several notable attempts by various researchers to develop techniques for predicting Parkinson's disease in subjects. The following is a brief review of the work done so far.

ISSN: 2005-4238 IJAST

Benba et al [2] collected the voice recordings from 50 subjects to discriminate the subjects into two groups (PD patients and neurological diseases patients). Using the Perceptual Linear Prediction (PLP), MFCCs, and ReAlitiveSpecTrAl PLP (RASTA-PLP) cepstral techniques, feature extraction was performed. Five supervised machine learning classifiers were used and an accuracy of 90% was reported with the linear SVM on first 11 PLP coefficients.

In another work, for the PD detection, Abdullah Caliskan et al [5] used a DNN Classifier that has the stacked autoencoder network which is combined with the softmax classifier. For their work, they used two datasets, the Parkinson Speech dataset (PSD) and Oxford Parkinson's Disease dataset (OPD). An accuracy of 93.79% and 68.05% was achieved on OPD and PSD datasets respectively.

A method was proposed by Little et al [6] to detect the PD, by analyzing the voice samples. 195 sustained vowels were collected from 31 people; in which healthy subjects were 8 and 23 were PD subjects. The support vector machine (SVM) was used to classify the subjects into PD and healthy subjects. Their proposed model achieves an accuracy of 91.4%.

David Gil et al. [8] proposed an SVM and artificial neural network (ANN) based method. The dataset [06] was used for the detection of PD. They used SVM with linear and pukkernel and achieved an accuracy of 91.79% and 93.33% respectively. The multilayer perceptron (MLP) an ANN-based method achieves an accuracy of 92.31%.

S. Grover et al [9] suggested a DNN to predict the Parkinson's disease severity. Voice recordings of 42 patients were collected with 200 recordings per subject, leading to a total of 5,875 valid recordings. Data normalization was performed by using min-max normalization. The total UPDRS was used as the target variable to analyze the PD severity with output classes being 'severe' and 'non-severe'. Their model achieved an accuracy of 94.22%.

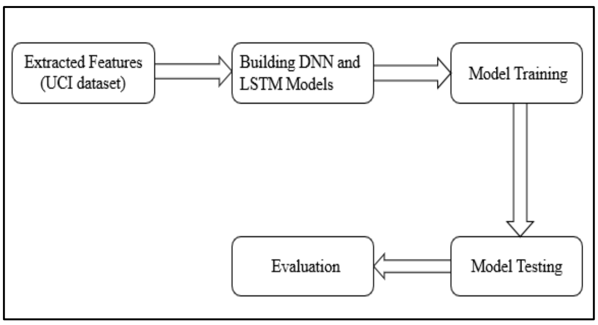
Sakar et al [24] suggested a model for diagnosis of voice-based PD using SVM and k nearest neighbor (k-NN). For validation, summarized Leave-One-Out (s-LOO) and Leave-One-Subject-Out (LOSO) were used. Various metrics like Matthews's correlation coefficient (MCC), sensitivity, accuracy, and specificity were used to assess the model efficiency. Their model achieved an accuracy of 82.50% and 85% for k-NN and SVM respectively.

All the works discussed above deal with construction of conventional machine learning based model for detection of Parkinson's disease based on subject's voice samples. In this paper, there is an attempt to explore the possibilities of building state of the art deep learning-based for the early detection of PD. The success of these models is evaluated by using various performance metrics. The results obtained from the proposed deep learning-based models were later also compared against the other PD detection methods.

## III. TECHNIQUES AND MODEL USED

In this work, we attempt to investigate the potential of developing a better deep learning model especially the long short-term memory, for prediction of Parkinson's disease based on subjects' voice samples. The proposed framework of our models is shown in Fig. 1.

Figure 1: Proposed Framework for Deep Learning




### *A. Deep Neural Networks*

A DNN is a special kind of artificial neural network (ANN) having many hidden layers presented between the input and the output layer [3] [19]. The DNN transforms the input into the output by finding the correct mathematical manipulation. In this work, a deep neural network with one input layer, one output layer, and three hidden layers is designed. The input layer had 28 units while the first hidden layer had 256 units. The second and third hidden layer had 128 units each. The output layer consists of only two units with the value 1for the PD subject and the value of 0 for a healthy subject. During the training for around 100 epochs, the network was tuned on hyperparameters to obtain better results. Fig. 2 shows the architecture of our DNN.

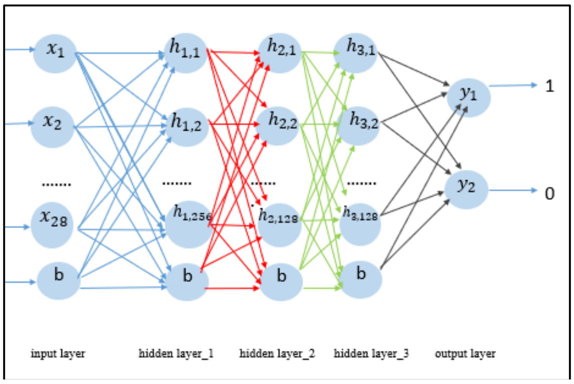



Figure 2.The selected DNN Architecture for the proposed work.

### *B. The LSTM model*

In the conventional feed-forward neural networks, the data flow happens in a sequential manner, the output of the first hidden layer is passed as input to the next layer and so on. However, in problems like speech recognition, such network was observed not to perform well. Hence alternate solution as Recurrent neural network (RNN) was proposed by Williams and Zipser [29] so as to increase the performance on such problems. RNN memorizes the previous information to calculate the current output. However, the problems that are associated with the RNN is the vanishing gradient and limited long-term memory. Both of these problems are addressed by the use of LSTM. Hochreiter and Schmidhuber [13] were the first to give the LSTM architecture. Fig. 3 shows the architecture of a single LSTM cell.

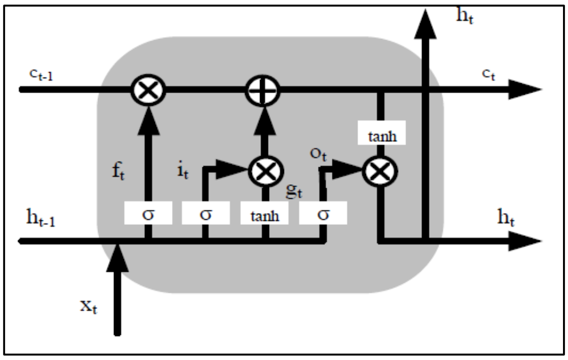



Figure 3. Single Cell LSTM Architecture. [15]

In this work, the deep learning LSTM model consists of an input layer having a number of units as the total attributes in input data. The model is having 1 LSTM layer with the varying LSTM units that pass input from the input layer and produces the target value. The training process was done with the rmsprop optimizer and also involved tuning of the network parameters like a number of hidden units, batch size, and a number of epochs in order to achieve better results. The dropout layer was also used to prevent the model to overfit.

## IV. EXPERIMENT DESIGN & RESULTS

This section presents the experimental design, the dataset along with the evaluation of the result obtained. Two separate sets of experiments were performed in this work. The first experiment was carried on DNN based model while the second one was based on the LSTM model. In the DNN based model, a dropout of 0.5 was applied to avoid overfitting of the network. The Adam optimizer and categorical cross entropy loss were used. A nonlinear activation function Relu was used in the dense hidden layers, and the softmax activation was applied to the output layer that will perform classification. In the other LSTM based model, the Rmsprop optimizer, categorical cross entropy loss and a sigmoid nonlinear activation function were used. During the training, 0.5% dropout is used in the hidden layer which will drop 50% of its activations. This ensures the model not to rely on any single node.

### *A. Study Subjects in the Dataset*

To evaluate the model, we have used PSD dataset [24] having voice samples of 40 subjects, out of which 20 were healthy subjects (10 female, 10 male) and 20 were PD subjects (6 female, 14 male). The age of an individual varies between 45 and 83 years. 26 speech samples -were taken from the people (control and PD subjects), that includes sustained vowels, short sentences, numbers (1-10) and words. 26 features that are based on time-frequency were extracted. Table 1 presents the description of the selected PSD dataset [24].

Table 1. Time & Frequency-Based Attribute Description.

| Parameters | Extracted Features |
|---|---|
| **Frequency** | Local Jitter ($C_2$), Absolute and local Jitter ($C_3$), Rap Jitter ($C_4$), ppq5Jitter ($C_5$), ddpJitter ($C_6$) |
| **Amplitude** | Local Shimmer ($C_7$), Local and dBShimmer ($C_8$), apq3Shimmer ($C_9$), apq5Shimmer ($C_{10}$), apq11Shimmer ($C_{11}$), ddaShimmer ($C_{12}$), |
| **Harmonicity** | Auto-correlation ($C_{13}$), NoisetoHarmonic ($C_{14}$), HarmonictoNoise ($C_{15}$) |
| **Pitch** | Median pitch ($C_{16}$), Mean pitch ($C_{17}$), Standard deviation (std) ($C_{18}$), Min. pitch ($C_{19}$), Max. pitch ($C_{20}$) |
| **Pulse** | Pulse count ($C_{21}$), Period count ($C_{22}$), Mean period ($C_{23}$), Std of period ($C_{24}$) |
| **Voicing** | The fraction of locally unvoiced frames ($C_{25}$), voice break count ($C_{26}$), voice break degree ($C_{27}$), |

### *B. Performance Evaluation Metrics*

The results of the suggested models were assessed by using precision, sensitivity, specificity, and accuracy as evaluation metrics. False Positives ($F_p$), True Positives ($T_p$), True Negatives ($T_n$), and False Negatives ($F_n$) rates are used to express each of these selected metrics. Sensitivity, also known as the True Positive rate, is termed as the system's ability to classify the PD subjects correctly. It is calculated as:

$$Sensitivity = \frac{T_p}{T_p+F_n} \qquad (1)$$

Specificity, a True Negative rate and is defined as the system's ability to classify the subjects into healthy subjects correctly. It is given by:

$$Specificity = \frac{T_n}{T_n+F_p} \qquad (2)$$

Precision is the ratio of True Positive relevant occurrences to the total number of retrieved instances. It is given by:

$$Precision = \frac{T_p}{T_p+F_p} \qquad (3)$$

Accuracy is the systems overall ability to classify the subjects into control and PD subjects correctly. It is calculated as:

$$Accuracy = \frac{T_n+T_p}{T_p+F_p+T_n+F_n} \qquad (4)$$

### *C. Experimental Results*

Neural network parameters may be tuned to get efficient results. As there is no particular strategy for choosing the values of parameters, hence are usually selected heuristically. With an exploratory experiment, the optimal size of 1st, 2$^{nd}$ & 3rd hidden layer of the DNN was observed to be 256, 128 & 64 respectively when trained for 100 epochs. Similarly for the LSTM network, certain parameters like a number of epochs, batch size, and hidden layer units, were also tuned before commencing for the actual training of the model. It was be observed that the LSTM model

ISSN: 2005-4238 IJAST

achieved the best results of with the batch size as 16 and the size of hidden layer as 32 with 80 epochs.

For statistical significance, 8 separate DNN and LSTM models were constructed and analyzed with the optimal value of the parameters obtained in the previous step. Statistical quantifiers like mean, median, std. deviation etc. were evaluated from the performance metrics obtained for each of these models for both the DNN and LSTM experiments.

Table 4. Description of performance values for each model of DNN and LSTM.

| Model# | DNN | | | | LSTM | | | |
|---|---|---|---|---|---|---|---|---|
| | **Prec.** | **Sens.** | **Spec.** | **Acc. (%)** | **Prec.** | **Sens.** | **Spec.** | **Acc. (%)** |
| Model 1 | 0.96 | 0.93 | 0.98 | 96.15 | 0.96 | 0.93 | 0.98 | 96.15 |
| Model 2 | 0.95 | 0.95 | 0.94 | 95.19 | 0.99 | 0.98 | 0.98 | 98.55 |
| Model 3 | 0.91 | 0.91 | 0.90 | 90.87 | 0.97 | 0.94 | 0.99 | 97.11 |
| Model 4 | 0.94 | 0.91 | 0.96 | 93.75 | 0.99 | 0.99 | 0.98 | 99.03 |
| Model 5 | 0.97 | 0.95 | 0.98 | 97.12 | 0.98 | 0.99 | 0.96 | 98.07 |
| Model 6 | 0.92 | 0.86 | 0.96 | 91.83 | 0.98 | 0.96 | 0.98 | 97.59 |
| Model 7 | 0.95 | 0.88 | 0.99 | 94.23 | 0.98 | 0.97 | 0.99 | 98.7 |
| Model 8 | 0.97 | 0.97 | 0.91 | 94.71 | 0.98 | 0.97 | 0.98 | 98.45 |
| ***Mean*** | *0.95* | *0.92* | *0.95* | ***94.23*** | *0.98* | *0.97* | *0.98* | ***97.96*** |
| ***Std*** | *0.02* | *0.03* | *0.03* | ***2.08*** | *0.01* | *0.02* | *0.01* | ***0.96*** |
| ***Median*** | *0.94* | *0.92* | *0.96* | ***94.47*** | *0.98* | *0.97* | *0.98* | ***98.26*** |
| ***Min*** | *0.91* | *0.86* | *0.90* | ***90.87*** | *0.96* | *0.93* | *0.96* | ***96.15*** |
| ***Max*** | *0.97* | *0.97* | *0.99* | ***97.12*** | *0.99* | *0.99* | *0.99* | ***99.03*** |

As per Table 4, the LSTM model yielded an accuracy of 99.03% with a standard deviation of less than 1% meaning that most of the values of accuracy were around the mean of 97.96 % which in itself is considerably better than the results of most of the recent works in this area. For the DNN models, the range of min and max accuracies was within 90% to 97% which is better than the earlier works but lesser than those of the LSTM models.

## V. DISCUSSION

The performance evaluation of our proposed models was analyzed and compared with all other conventional detection systems for PD. Table 5 shows the comparative analysis of accuracies of the proposed models with the existing models on Sakar et al dataset [24].

Table 5: Comparison of various models for PD detection on the Sakar et al. dataset [24].

| Author | Technique applied | Max Accuracy observed (%) |
|---|---|---|
| **Sakar et al (2013)** | SVM | 82.50 |
| **Sakar et al (2013)** | k-NN | 85.0 |
| **Abdullah Caliskan et al (2017)** | DNN classifier | 68.05 |
| **Proposed model** | DNN | **97.12** |
| **Proposed model** | LSTM | **99.03** |

Sakar et al [24] proposed SVM and k-NN based classifiers and used LOSO and s-LOO cross-validation schemes. Abdullah Caliskan et al [05] proposed a DNN classifier a stacked autoencoder with the softmax classifier for this problem. Their proposed model achieves an accuracy of 68.05% on PSD dataset [24]. However, both the proposed models, i.e. the DNN and the LSTM based prediction models outperform all the models that were used on the same dataset achieving the best accuracy of 97.12% and 99.03% respectively as shown in Fig. 4.

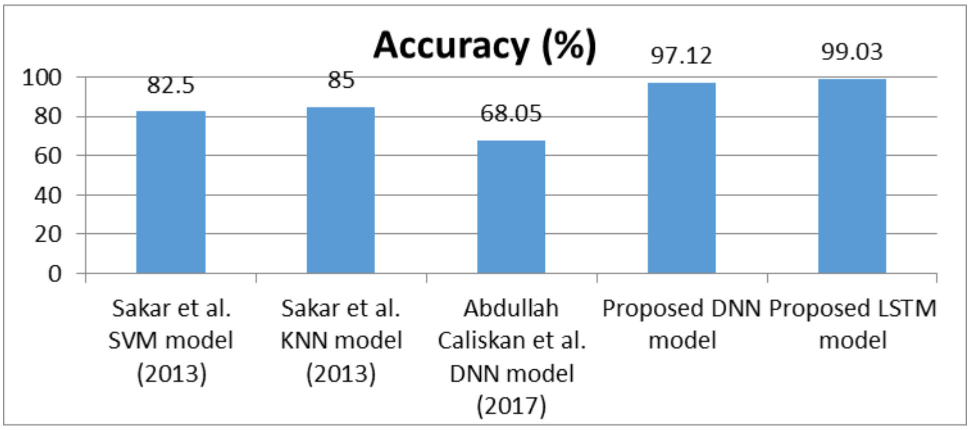



Figure 4: Comparison of Accuracies of the proposed LSTM model with previous works on the dataset.

## VI. CONCLUSION

In this work, the problem of Parkinson's disease detection is attempted through state of the art deep learning approaches which require noninvasive voice samples of the subject. Such a model can be used for the PD diagnosis remotely and also will be helpful to monitor the progress of the patient. This will reduce the patient's physical visits to the clinics. The idea was to use the extracted features of voice samples and then perform the PD classification. The fine-tuning of hyperparameters of the network was also performed so as to achieve the best possible results on the selected PD dataset. The proposed models achieve an accuracy of 97.12% and 99.03% for DNN and LSTM models, respectively, suggesting that these are reliable models to detect Parkinson's disease. The high value of mean accuracies with low standard deviation values of our proposed models suggests that they are better than the existing Parkinson's disease detection models based on voice samples.

## References

[1] Baken, R. J., & Orlikoff, R. F. (2000). Clinical measurement of speech and voice. Cengage Learning.
[2] Benba, A., Jilbab, A., & Hammouch, A. (2016). Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis. IEEE transactions on neural systems and rehabilitation engineering, 24(10), 1100-1108.
[3] Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural networks: Tricks of the trade (pp. 437-478). Springer, Berlin, Heidelberg.
[4] Cunningham, L., Mason, S., Nugent, C., Moore, G., Finlay, D., & Craig, D. (2010). Home-based monitoring and assessment of Parkinson's disease. IEEE Transactions on Information Technology in Biomedicine, 15(1), 47-53.
[5] Caliskan, A., Badem, H., Basturk, A., & Yuksel, M. E. (2017). Diagnosis of the parkinson disease by using deep neural network classifier. Istanbul University-Journal of Electrical & Electronics Engineering, 17(2), 3311-3318.
[6] Dastgheib, Z. A., Lithgow, B., & Moussavi, Z. (2012). Diagnosis of Parkinson's disease using electrovestibulography. Medical & biological engineering & computing, 50(5), 483-491.
[7] Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., & Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. European Archives of Oto-rhino-laryngology, 258(2), 77-82.
[8] Gil, D., & Manuel, D. J. (2009). Diagnosing Parkinson by using artificial neural networks and support vector machines. Global Journal of Computer Science and Technology, 9(4).

ISSN: 2005-4238 IJAST

[9] Grover, S., Bhartia, S., Yadav, A., & Seeja, K. R. (2018). Predicting severity of Parkinson's disease using deep learning. Procedia computer science, 132, 1788-1794.
[10] Gulli, A., & Pal, S. (2017). Deep learning with Keras. Packt Publishing Ltd.
[11] Harel, B., Cannizzaro, M., & Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. Brain and cognition, 56(1), 24-29.
[12] Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. Behavioural neurology, 11(3), 131-137.
[13] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[14] Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry, 79(4), 368-376.
[15] Jithesh, V., Sagayaraj, M. J., & Srinivasa, K. G. (2017, February). LSTM recurrent neural networks for high resolution range profile based radar target classification. In 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 1-6). IEEE.
[16] Olanow, C. W., Stern, M. B., & Sethi, K. (2009). The scientific and clinical basis for the treatment of Parkinson disease (2009). Neurology, 72(21 Supplement 4), S1-S136.
[17] O'Sullivan, S. B., Schmitz, T. J., & Fulk, G. (2019). Physical rehabilitation. FA Davis.
[18] Langston, J. W. (2002). Parkinson's disease: current and future challenges. Neurotoxicology, 23(4-5), 443-450.
[19] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature 521.
[20] Little, M., McSharry, P., Hunter, E., Spielman, J., & Ramig, L. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nature Precedings, 1-1.
[21] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomedical engineering online, 6(1), 23.
[22] Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and hearing Disorders, 43(1), 47-57.
[23] Rigas, G., Tzallas, A. T., Tsipouras, M. G., Bougia, P., Tripoliti, E. E., Baga, D., & Konitsiotis, S. (2012). Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. IEEE Transactions on Information Technology in Biomedicine, 16(3), 478-487.
[24] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.
[25] Singh, N., Pillay, V., & Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. Progress in neurobiology, 81(1), 29-44.
[26] Standaert D. G., Saint-Hilaire M. H., Thomas C. A. Parkinson's disease Handbook. American Parkinson Disease Association, New York, USA, 2015.
[27] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2009). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE transactions on Biomedical Engineering, 57(4), 884-893.
[28] Van Den Eeden, S. K., Tanner, C. M., Bernstein, A. L., Fross, R. D., Leimpeter, A., Bloch, D. A., & Nelson, L. M. (2003). Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. American journal of epidemiology, 157(11), 1015-1022.
[29] Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. Neural computation, 1(2), 270-280.

ISSN: 2005-4238 IJAST