Artificial Neural Network to prescient the severity of Parkinson's Disease

Mrs. Kirti Wanjale
Computer Engineering Department
Vishwakarma Institute Of Information Technology
Pune, India
kirti.wanjale@viit.ac.in

Miss. Madhavi Nagapurkar
Computer Engineering Department
Vishwakarma Institute Of Information Technology
Pune, India
madhavi.nagapurkar@viit.ac.in

Mr. Parag Kaldate
Computer Engineering Department
Vishwakarma Institute Of Information Technology
Pune, India
parag.kaldate@viit.ac.in

Mr. Onkar Kumbhar
Computer Engineering Department
Vishwakarma Institute Of Information Technology
Pune, India
onkar.kumbhar@viit.ac.in

Mr. Subhranil Bala
Computer Engineering Department
Vishwakarma Institute Of Information Technology
Pune, India
subhranil.bala@viit.ac.in

***Abstract*—Parkinson's disease is a central nervous system disorder in which dopamine generating cells present in the substantia nigra (a part of our brain) gets damaged or perished. Due to a deficiency of dopamine, which acts as a neurotransmitter in our body, symptoms of Parkinson's disease are caused. The change in vocal speech is one of the major symptoms shown and can be used to detect this disease. In this paper, there is an implementation and proprietary comparison of various machine learning models. The proposed model comparison shows that ANN is the comparatively better ML model to detect and predict Parkinson's disease which has accuracy for total UPDRS is 83.56 % and for motor UPDRS is 85.135 % which are the scales for measuring the severity elaborated further. We have utilized "Sklearn", "TensorFlow" and "Keras" python AI libraries to actualize all the ML models. It sums up that the accuracy generated by the proposed model is 84.45 % which is substantially higher than the research work done so far.**

***Keywords*—Random Forest, Decision Tree, Linear Regression, SVR, ANN, Parkinson disease, Machine Learning, Keras, Tensorflow, Adam, UPDRS**

I. INTRODUCTION

Parkinson's disease is the second most common neurological disarray. It affects dopamine generating nerve cells in the midbrain. The cause of which is still unknown, but various factors involved are certain genetic mutations and environmental changes. Dopamine is a chemical that acts as both hormone and neurotransmitter and due to deficiency of this chemical, there are certain repercussions like tremors, bradykinesia, rigid muscles, impaired posture, and speech changes. Research shows that vocal changes can be observed approximately five years before the actual diagnosis of the disease. Therefore, it is one of the major symptoms of this disease and hence we have used this symptom to detect and predict the disease. There is no proprietary methodology for the prognosis of PD which led the researchers to apply machine learning for the same. Implementation of various machine learning models like multiple linear regression, polynomial regression, random forest, decision trees, SVR and ANN is done. A comparative study of the results of the implementation of these models is done and we have come up with various conclusions.

II. RELATED WORK

Several researchers have done the research to apply Machine Learning for the prediction and detection of Parkinson's Disease. In a paper by Srishti Grover, Saloni Bhartia, Akshama, Abhilasha Yadav, Seeja K [1] DNN model was implemented on the static dataset and patients were classified in two categories "severe" or "non-severe" in the view of the estimation of motor and total UPDRS values. A.BOUROUHOU, A.JILBAB, C.NACIR, A.HAMMOUCH [2] implemented 3 classifiers, namely KNN, SVM and Naive Bayes classifier on a dataset having 26 extracted features and concluded that SVM was the best classifier. Satyabrata Aich et al. [3] implemented a non-linear classifier along a decision tree for the classification of PD and non-PD patients. Meilin Su, Keh-Shih Chuang[4] performed a dynamic selection of features based on fuzzy entropy which used similarity measures to discard non-relevant features. R. Viswanathan et al. [5] used an approach in which the features were extracted based on sustained consonant /m/ for the classification of PD Patients. Elmehdi BENMALEK, Jamal ELMHAMDI, Abdelilah JILBAB [6] have mapped the extracted features to UPDRS using neural network and least square regressions. Meysam Asgari, Izhak Shafran [7] have done a computational approach for computing various features by performing 3 tasks and then calculated UPDRS value for severity prediction of PD. Mehrbakhsh Nilashi, Othman Ibrahim & Ali Ahani [8] have applied Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR) for the prognosis of progression of PD. Aarushi Agarwal, Spriha Chandrayan and Sitanshu S Sahu [9] have presented a predictive model using Extreme Machine Learning techniques. Athanasios Tsanas, Max A. Little, Patrick E. McSharry, Jennifer Spielman, and Lorraine O. Ramig [10] have tested the accuracy of usage novel algorithms for the discrimination of PD and healthy patients. In a paper written by Evaldas Vaiciukynas et al [11] have done work in detecting Parkinson's Disease by processing speech signals from sustained phonation. Phonation relates to the vowel/a/voicing errand and discourse to the way to express a short sentence in the Lithuanian language. Timothy J. Wroge et al [12] have done work on diagnosis and prediction of Parkinson's disease using non-intrusive voice bio-markers as input features for the machine learning algorithms. They have provided an analysis of correlation of the viability of different machine learning classifiers in disease diagnosis with noisy and high dimensional information.

III. Dataset

We are using the same dataset that has been used in the referred research paper(i.e. Predicting Severity Of Parkinson's Disease Using Deep Learning by *Shrishti Grover et al*). They have taken this dataset from the UCI-ML repository. The dataset used to be created by Athanasios Tsanas (tsanasthanasis '@' gmail.com) and Max Little (littlem '@' physics.ox.ac.uk) of the University of Oxford, in collaboration with 10 clinical centers in the US and Intel Corporation who developed the telemonitoring device to report the speech signals. The original study used a range of linear and nonlinear regression techniques to predict the clinician's Parkinson's ailment symptom rating on the UPDRS scale. This dataset consists of various biomedical voice measurements from forty-two people with early-stage Parkinson's ailment recruited to a six-month trial of a telemonitoring device for remote symptom development monitoring. The recordings had been automatically captured in the patient's homes. Minimum 200 recordings of each patient were considered in a dataset. Dataset is in ASCII-CSV format which is composed of 5,876 instances as rows and 22 columns including 16 voice parameters.

1.Jitter(%)	2.Jitter(Abs)	3.Jitter:RAP	4.Jitter:PPQ5
4.Jitter:DDP	5.Shimmer	6.Shimmer(dB)	7.Shimmer:APQ3
9.Shimmer:APQ5	10.Shimmer:APQ11	11.Shimmer:DDA	12.NHR
13.HNR	14.RPDE	15.DFA	16.PPE

Table 1: Input Attribute Information

Metric	Severe	Non-Severe	Scaling Range
Total-UPDRS	Above 25	0-25	0-176
Motor-UPDRS	Above 20	0-20	0-108

Table 2: Severity and Scale Range of Output Classes

The above table describes 16 parameters that are used for the prediction of the severity of the disease. Among all mentioned parameters, 1-13 are Linear parameters whereas14-16 belongs to non-linear parameters. The most important purpose of the statistics is to predict the motor and whole UPDRS rankings('MOTOR_UPDRS' and 'TOTAL_UPDRS') from the 16 voice measures.

IV. Model Selection

A) Multiple Linear Regression

Multiple Linear Regression also called multiple regression is a type of linear regression in which a linear relationship exists between multiple independent variables to predict the value of a single dependent variable. The formula of this type of linear regression is:-

$$Y=m_1x_1+m_2x_2+m_3x_3+m_4x_4+.........+m_nx_n+c.$$

mi = slope coefficients

Xi = independent variable

Y = dependent variable

C = y-intercept

B) Polynomial Regression

Polynomial regression is a sort of linear regression in which the connection between the independent and dependent instance is n^{th} degree polynomial. The advantage of this regression is that polynomial fits a wide range of curvature.

$$Y = c + m_1x + m_2x^2 +....+ m_nx^n$$

mi = slope coefficients

x = independent variables

y = dependent variable

c = y-intercept

C) Support Vector Regression

Support Vector Regression works on principles similar to that of SVM. SVR can be seen as the adapted form of SVM in which the value of the dependent variable is not categorical instead it is numerical. One of the benefits of SVR is that it allows the construction of non-linear models without changing the independent variables and hence allowing a better understanding of the model. Various kernel functions are used in the model generation of SVR like (a) Linear, (b)Polynomial, (c)Sigmoid and (d) Radial Basis.

D) Decision Tree

Decision Tree comes under a supervised Machine Learning algorithm in which target variables' value is predicted using various decision rules. Decision Tree is usually used for classification, and entropy as a measure of impurity is used as a criterion. To implement a decision tree for regression we need an impurity metric that is suitable for continuous variables. So, the impurity metric using weighted mean squared error is defined and calculated for this purpose.

E) Random Forest

Random Forest is also a supervised Machine Learning model that is based on ensemble learning. Ensemble Learning is a method that integrates predictions from various Machine Learning Algorithms. A model that incorporates many models is called an ensemble model. Random Forest comes under bagging techniques. In Random Forest, multiple decision trees are executed in parallel and no interaction is done among these trees when they are built. Results of multiple decision trees are aggregated using averaging or model votes.

F) Feed forward neural Network

Feed-forward neural networks are particularly used for supervised learning in cases where the records to be realized is neither sequential nor time-dependent. These are the artificial neural networks where the connections do not form a cycle. It is the first type of ANN invented and is simpler than a recurrent neural network which is its counterpart. The information there follows solely forward direction besides forming any loops, consequently termed as a feed-forward neural network. It is one of the biologically stimulated classification algorithms.

V. Algorithm and work flow

The Parkinson Telemonitoring dataset that we have gathered from the UCI Repository contains 5876 instances, 16 input biomedical parameters as input and 2 parameters as output. As from the above comparisons among all the models, we have selected ANN as the best model to work with as it comes under the neural network which works under the principle of adapt and learn. Adapting various process situations and based on changing environments, learning those circumstances is an imp task in which neural networks perform very well. Initially, the dataset is divided into an 80:20 ratio to create train and test datasets with no randomization. Then the values are scaled with the help of min-max scaler (all values are converted between 0 – 1). After scaling, we created our Neural Network. The neural network that we have used is the Feed-Forward neural network. It has the following configuration:-

1. Input layer with 16 neurons.
2. Output layer with 2 neurons.
3. Three hidden layers each with 100 – 200 – 100 neurons.
4. Kernel Initializer used for assigning weights is normal (not zero but approx near to zero).
5. Activation function used for hidden layers and input layers is a rectified linear unit and for the output layer is linear in nature as output format is numerical.
6. An optimizer that we have used finally after testing seven varieties is the ADAM optimizer as it works well with noisy or sparse gradients.

The different types of optimizers that we have tested the model upon are ADAM, ADAGRAD, NADAM, SGD, ADADELTA, RMSPROP, ADAMAX. For accuracy we have used 3 types of loss functions – Mean square error, mean square logarithmic error and mean absolute error. The learning rate that we have used is the default for Adam optimizer that is 0.00146 but we have also tested the model on variable learning rates to see how the model performs and what we observed is default LR is better for our model. For testing the model, we have batch size as 20 and total epoch size as 1000. To be able to use the best model out of 1000 epochs, we have created checkpoints so that our model gets saved every time accuracy in the form of loss function reduces by a small amount. After all model training is done, we manually choose the last model which has the lowest loss value among all. After selecting the best model, for testing, now we first inverse transformed all our output and then converted it into a vector of 0s and 1s using the range provided for both the outputs. For accuracy, we have created a confusion matrix.

Function Of Adam Optimizer

Adam optimizer is an algorithm that we have used for the iterative setting of weights of the neurons in the network as per the epochs. We have chosen this particular optimizer over other optimizers as it requires less tuning and also it can handle high variability data. the main reason for its choice over others was that it hardly requires any resources to adjust itself with the neural network which in turn gives a huge benefit in terms of accuracy and sensitivity achieved. Also, Adam optimizer takes into account various benefits of other optimizers and work accordingly.

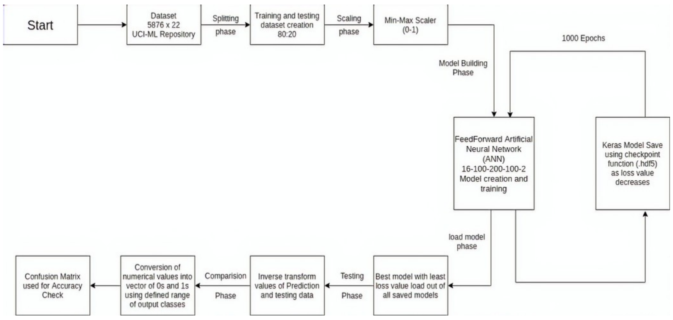

Figure 1: ANN Model Work Flow

	ADAM	NADAM	ADADELTA	ADAGRAD	SGD	RMSPROP	ADAMAX
TEST MOTOR UPDRS	73.00	74.38	75.65	75.65	72.42	76.40	73.95
TEST TOTAL UPDRS	72.40	74.55	74.55	75.74	73.78	76.51	74.29
TRAIN MOTOR UPDRS	97.27	99.91	99.89	81.93	77.53	99.87	100
TRAIN TOTAL UPDRS	94.72	99.85	99.89	81.36	79.14	99.74	99.97

Figure 4: Accuracy Comparision for All 7 Optimisers For Training And Testing Data

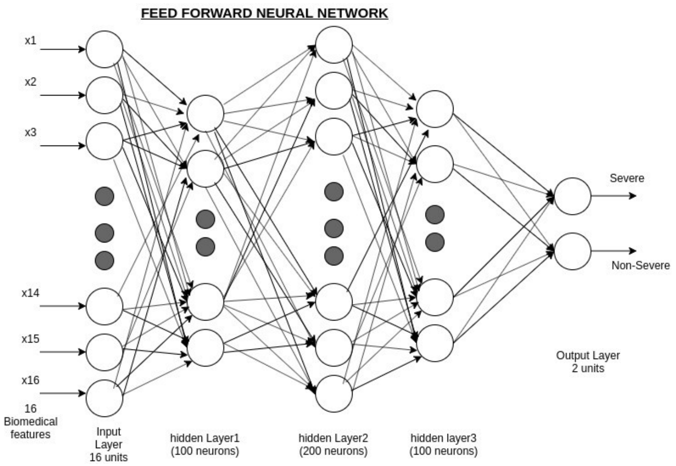

Figure 2: Proposed Neural Network

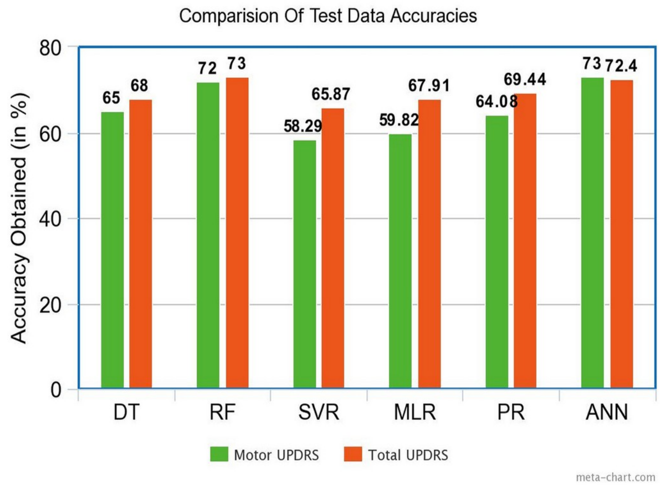

Figure 5: Accuracy Comparision for Regressions Methods for training and testing data:- 1)Decision Tree 2) Random Forest 3) SVR 4) Multiple Linear Regression 5)Polynomial Regression 6)ANN

RESULTS AND PLOTS

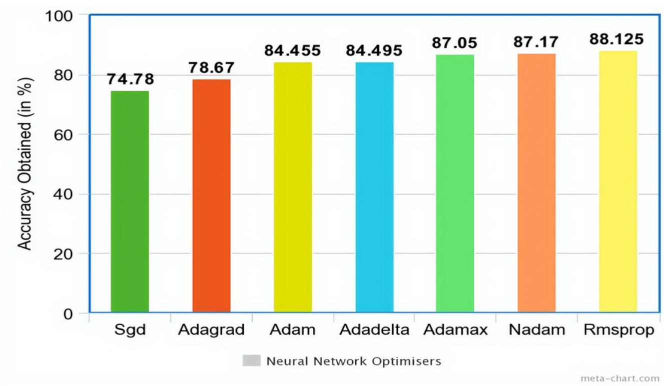

Figure 3: Comparision Of Accuracies For Train And Test Data For 7 Different Types Of Optimizers.

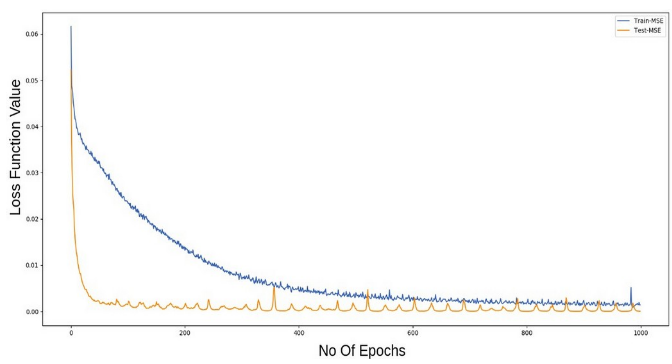

Figure 6: Variation In The Mean Square Error Value For 1000 Epochs For Training And Testing Data.

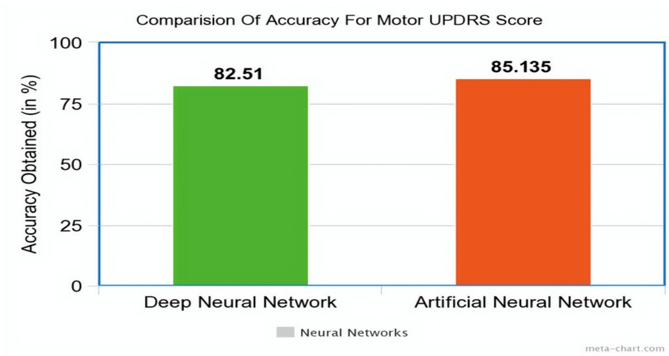

Figure 7: Comparison of Motor UPDRS Score Accuracies Of ANN and DNN

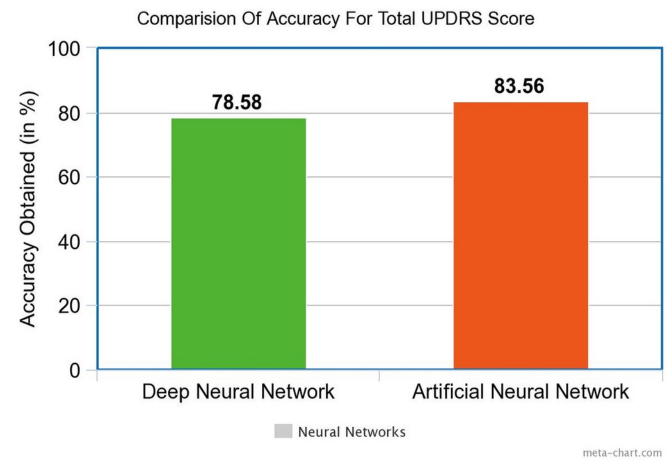

Figure 8: Comparison of Total UPDRS Score Accuracies Of ANN and DNN

The dataset taken from the UCI-ML repository has 16 biomedical speech parameters which are given as input to an ANN and the output generated is the total UPDRS score and the Motor UPDRS Score. In the case of total UPDRS, The average accuracy acquired for the test and train data is 83.56% (test data - 72.40% and train data - 94.72%) and for Motor UPDRS, Average accuracy obtained was 84.63% (test data - 73.0% and train data - 97.27%). We correlate our results with that of the research done by Grover, S., Bhartia, S., Akshama, Yadav, A., & K.R., S., because they came up with their model for the same UCI-ML Parkinson's Telemonitoring Voice Dataset. They used DNN to predict the severity of PD and their model produced an average accuracy for the total UPDRS and Motor UPDRS (test data and train data) was 78.58% and 82.51% respectively. A comparison of the accuracy is shown in the figure7 and figure8. Also, we have tested 7 different optimizers concerning the epochs and accuracy obtained by the model from them and the results are plotted in figure3. Individually for both the outputs we have created a table for all types of optimizer and numbers are represented in a 4x7 matrix shown in figure4. We have also seen the rate at which the loss value decreases both in case of Training and Testing from Mean Square Error in figure6. For Comparision of ANN with classical Machine Learning Models, we have used 6 conventional models like Decision Trees, Random Forest, Support Vector Machines, Multiple Linear Regression, etc whose individual results for train and test data is plotted in figure5.

Platform used for Performance analysis

A) Bar Chart

We have utilized single bar charts as well as double bar charts for the analysis of overall performances as it clarifies better trends as in contrast to different methods. It additionally lets in a visible take a look at the consequences estimating key values at a glance.

Future Scope

In the paper, we have taken the data from the UCI – ML repository wherein all the data were already recorded and preprocessed. What we can do differently from this type of static data processing is that we can implement dynamic data processing in which we can create some kind of front end application that will record the voice of the user and do preprocess like extracting all the features on the back-end side. This will, in turn, provide dynamic training of model with some tuning changes in the existing model. Also, we can integrate different sources of data of predicting Parkinson's Disease like hand movements, deep brain simulation data, etc. Also, a complete analysis of the data capturing mechanism can be given to increase the trust of the application. we can take the help of the Internet Of Things as well to increase the accuracy we can use data from sensors placed at various positions during testing.

Conclusion

In this paper, we have implemented an ANN which is the feed-forward neural network. It is found to be providing exceptional results in predicting the severity of Parkinson's disease when compared with other regression models. The model seems to use its capabilities to adapt and learn to its full potential. It can analyze the nonlinear patterns within the dataset with great ease. With the use of Ann, we have also taken into consideration all the problems that might come in future modifications which other models do not provide very efficiently. It also allows greater flexibility in training and testing phase cases when there is an abundant change in the dataset due to which other models collapse but Ann seems to have a better handling mechanism for this as it provides humans like understanding behavior.

References

[1] Grover, S., Bhartia, S., Akshama, Yadav, A., & K.R., S. (2018). Predicting Severity Of Parkinson's Disease Using Deep Learning. *Procedia Computer Science*, *132*, 1788-1794. doi:10.1016/j.procs.2018.05.154

[2] Bourouhou, A., Jilbab, A., Nacir, C., & Hammouch, A. (2016). Comparison of classification methods to detect the Parkinson disease. *2016 International Conference on Electrical and Information Technologies (ICEIT)*. doi:10.1109/eitech.2016.7519634

[3] Aich, S., Younga, K., Hui, K. L., Al-Absi, A. A., & Sain, M. (2018). A nonlinear decision tree based classification approach to predict the Parkinson's disease using different feature sets of voice data. *2018 20th International Conference on Advanced Communication Technology (ICACT)*. doi:10.23919/icact.2018.8323864

[4] Su, M., & Chuang, K. (2015). Dynamic feature selection for detecting Parkinson's disease through voice signal. *2015 IEEE MTT-S 2015 International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO)*. doi:10.1109/imws-bio.2015.7303822

[5] Viswanathan, R., Khojasteh, P., Aliahmad, B., Arjunan, S., Ragnav, S., Kempster, P., … Kumar, D. K. (2018). Efficiency of Voice Features Based on Consonant for Detection of Parkinson's Disease. *2018 IEEE Life Sciences Conference (LSC)*. doi:10.1109/lsc.2018.8572266

[6] Benmalek, E., Elmhamdi, J., & Jilbab, A. (2017). Multiclass classification of Parkinson's disease using cepstral analysis. *International Journal of Speech Technology*, *21*(1), 39-49. doi:10.1007/s10772-017-9485-2

[7] Asgari, M., & Shafran, I. (2010). Extracting cues from speech for predicting severity of Parkinson'S disease. *2010 IEEE International Workshop on Machine Learning for Signal Processing*. doi:10.1109/mlsp.2010.5589118

[8] Nilashi, M., Ibrahim, O., & Ahani, A. (2016). Accuracy Improvement for Predicting Parkinson's Disease Progression. *Scientific Reports*, *6*(1). doi:10.1038/srep34181

[9] Agarwal, A., Chandrayan, S., & Sahu, S. S. (2016). Prediction of Parkinson's disease using speech signal with Extreme Learning Machine. *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. doi:10.1109/iceeot.2016.7755419

[10] Tsanas, A., Little, M., McSharry, P., & Ramig, L. (2009). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *Nature Precedings*. doi:10.1038/npre.2009.3920.1

[11] Vaiciukynas, E., Verikas, A., Gelzinis, A., & Bacauskiene, M. (2017). Detecting Parkinson's disease from sustained phonation and speech signals. *PLOSONE*, *12*(10),e0185613. https://doi.org/10.1371/journal.pone.0185613

[12] Wroge, T. J., Ozkanca, Y., Demiroglu, C., Si, D., Atkins, D. C., & Ghomi, R. H. (2018). Parkinson's Disease Diagnosis Using Machine Learning and Voice. 2018 IEEE Signal Processing in Medicine and Biology Symposium(SPMB). https://doi.org/10.1109/spmb.2018.861567