Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Comprehensive and Comparative Analysis of Different Types ML Algorithm for Detecting Parkinson's Disease

Muppala Sunny Chowdhary[1], Jacob Jasper Roy Paga[2], Deepa D[3], GowriManohari V[4], Nithya S[5]

[1,2]Assistant Professor,
Sathyabama Institute of Science and Technology, India
[3,4,5]Assistant Professor,
Sathyabama Institute of Science and Technology, India

Article Info
Page Number: 588- 594
Publication Issue:
Vol 71 No. 3s2 (2022)

Article History
Article Received*: 28 April 2022*
Revised*: 15 May 2022*
Accepted*: 20 June 2022*
Publication*: 21 July 2022*

Abstract
Biomarkers derived from the human voice can offer knowledge into neurological problems, like Parkinson's infection (PD), on account of their basic psychological and neuromuscular capacity. PD is a reformist neurodegenerative issue that affects around 1,000,000 individuals in India, with roughly 60,000 new clinical determinations made every year. Because in engine control that is the sign of the infection, voice can be utilized to identify and analyze PD. This paper has been created for directed characterization calculations, like Xtreme Gradient Boosting (xgboost), which is another calculation for Machine Learning.

Keywords: - Parkinson disease; Classification; Machine learning classifier; Deep neural network; Diagnosis

I. INTRODUCTION

Parkinson's illness is a neuropathological disorder that can affect the engine elements of the human organism. It has been estimated that over a million individuals are experiencing this condition in North America. Parkinson's illness is portrayed by the degeneration of certain synapse groups that are liable for creating synapses that incorporate dopamine, serotonin and acetylcholine. The deficiency of dopamine's outcome in the indications like uneasiness, gloom, weight reduction and visual issues. Different indications that can be found in individuals with Parkinson's illness are helpless equilibrium, voice weakness and quake. Different examination contemplates have shown that 90% of individuals who experience the ill effects of PD have discourse and vocal issues which incorporate dysphonia, droning and hypophonia. Along these lines, the corruption of voice is viewed as the underlying manifestation of Parkinson's infection.

The reason and fix of PD are yet obscure yet the accessibility of different medication treatments offers the critical alleviation of indications particularly at its previous stages, in this manner improving the existing nature of patients and lessens the assessed cost of the Pathology. Theinvestigation of voice estimation is straightforward and non-intrusive. In this manner, to follow the movement of PD the estimation of voice can be utilized.

For evaluating the movement of PD, different vocal tests have been contrived which incorporate supported phonations and running discourse messages. The telemonitoring and telediagnosis frameworks have been generally utilized as these frameworks depend on discourse signals which

are affordable and simple to utilize. Consequently, in this paper, there is an endeavor to investigate a superior AI-based model for the early discovery of PD from the voice tests of the subject.

II. Purpose Of The Project

Early finding of a Parkinson's infection could be helpful for the recognizable proof of individuals who can partake in preliminaries of its representatives, or at last to attempt to end sickness movement once-powerful illness changing medications have been distinguished.

III. Problem Statements

The goal of this project is to create a model which accurately reveals the presence of the disease Parkinson’s disease in an individual.

IV. Dataset Information

4.1. Dataset

This dataset contains biological voice assessments performed on 23 Parkinson's disease patients (PD). The central place of the data is focused on identifying strong individuals from those with the condition.

From this data set, we can get some important information about the disease. For instance, we can see that the number of instances is 197.

4.2. Attribute Information

There are 147 measurements with Parkinson’s and 48 without Parkinson’s. ***This dataset had been downloaded from the UCI ML repository.***

V. Proposed System

Various machine learning methods were used to create the proposed system. The researchers put it to the test at UCI's machine learning repository. The data acquired throughout the study was then examined and the system was trained.

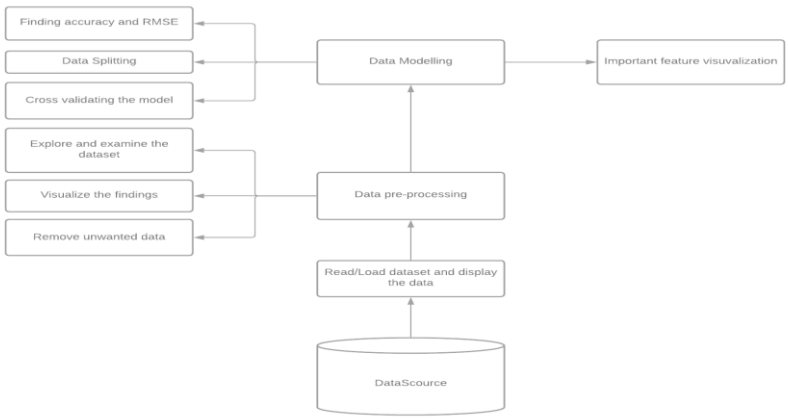


Fig 1. Workflow Diagram

1. Acquiring the dataset We have to download the voice sample dataset from UCI Learning Repository and store it on our PC.

2. Read/Load the dataset Now we will load the dataset from our PC and display it in our code using Pandas data-frame.

3. Data Pre-processing Here we will explore and examine our dataset (using .info, .corr, .describe methods), remove unwanted values and then visualize our findings using heat-map and bar-chart.

4. Data Modelling We will split our data into training and testing parts using the sklearn library, and we will predict using the XGBoost algorithm and find out the Accuracy and Root Mean Square Error of the model's prediction. Finally, we will use k-fold cross-validation to make our model more robust, and we will find out the mean Accuracy and mean Root Mean Square Error of the validation result.

5. Important feature visualization using XGboost Finally, we will visualize and find out the feature that has the highest importance among all the features.

1.1 Classification Algorithms

The algorithm used- ***Logistic regression, SVM, K-NN, DNN, XGBoost***

XGBoost is a machine learning algorithm that takes advantage of the speed and performance of gradient boosting.

How does the XGBoost Algorithm work?

- XGBoost fabricates truly short and straightforward choice trees iteratively.
- XGBoost begins by making the main straightforward tree which has a lackluster showing without help from anyone else.
- It at that point fabricates another tree that is prepared to anticipate what the primary tree couldn't and is itself a powerless student as well.
- The calculation goes on by consecutively assembling more frail students, every one amending and lessening the mistakes of the past tree until a halting condition is reached.

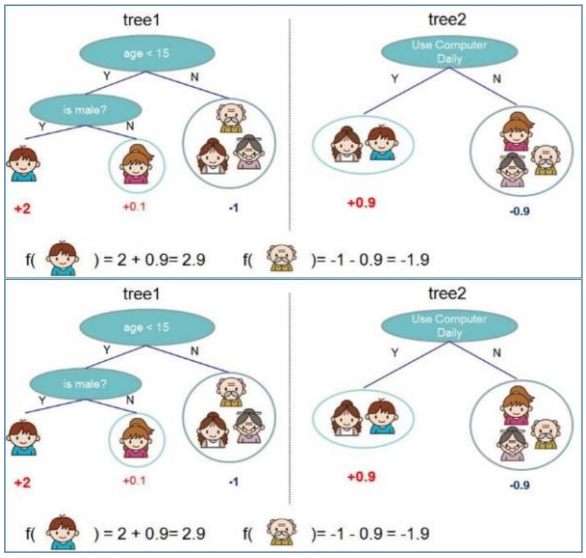


Fig 2.XGBoost example Advantages of XGBoost

- Better Speed and execution: XGBoost is similarly quicker and it has shown better execution over different calculations on an assortment of AI benchmark datasets.
- Regularization: Standard GBM execution has no regularization like XGBoost, in this manner, it additionally assists with diminishing overfitting.
- Equal Processing: XGBoost uses the force of equal handling and that is the reason it is a lot quicker than GBM.
- It utilizes various CPU centers to execute the model.
- Taking care of Missing Values: XGBoost has an in-constructed ability to deal with missing qualities.

1.2 Performances Evaluation Metrics

To evaluate the performance of the classifiers we need Evaluation/ Confusion matrices which are been shown in table 2.

	Predicted PD Patient	Predicted Healthy Person
Actual PD Patient	TP	FN
Actual Healthy Person	FP	TN

Table 1. Evaluation/ confusion matrices

From table 2 yields the following metrics, which are written as equations. Accuracy, sensitivity, and specificity are all terms used to describe how accurate, sensitive, and particular something is.

$$\textbf{Accuracy} = ((TP+TN)/(TP+TN+FP+FN))*100$$

$$\textbf{Sensitivity} = (TP/(TP+FN))*100$$

$$\textbf{Specificity} = (TN/(TN+FP)*100$$

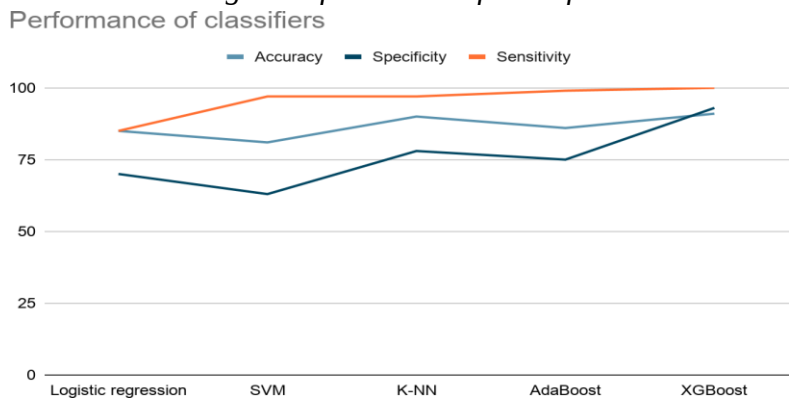
In this study, the task was to investigate the correctness of a classification test using AI-classified systems. The main components of the test were presented as follows: Data layer 1, layer 2, and layer 3. The outputs of the test were then summarized as follows: Evaluations of the correctness of the classifications were then conducted.The show assessment evaluations like diagram precision, character, affectability, and MCC were thusly organized to survey the presence of these classifiers and made into table 1. In like manner, Table 2 shows the presentations of these classifiers.

Classifier	Performance evaluation metrics			
	Parameters	Accuracy (%)	Specificity (%)	Sensitivity (%)
Logistic regression	C=10	85	70	85
SVM	C=10,g =0.025	81	63	97
K-NN	K=5	90	78	97
AdaBoost	-	86	75	99
XGBoost	-	91	93	100

Table 2.The performance evaluation of classifiers.

As indicated by the above table, we can see that the strategic relapse grouping exactness is 85%, affectability 85% and particularity 70 %. SVM acquired 81% exactness, 63% particularity and 97% affectability. K-NN got 90% precision, 78% particularity and 97% affectability. AdaBoost acquired 86% exactness, 75% explicitness and 99% affectability. XGBoost exhibitions were brilliant when contrasted with other arrangementcalculations. The XGBoost acquired 91% precision, 93% particularity and 100% affectability. Thus for PD conclusion, the XGBoost is the best AI classifier.

Fig 3.Performances of classifiers.



VI. Conclusion

The Parkinson's disease research field is currently very important, and early discovery of the condition can improve the patient's quality of life. Speech analysis has yielded considerable results as a result of recent improvements in methodology. The major goal of this project is to demonstrate how speech signals may be used to diagnose Parkinson's disease. Because voice measures are noninvasive, speech processing has had tremendous promise in the identification of Parkinson's disease for many years. We employed the XGBoost machine learning method in our project. The machine learning model generated an accuracy of 91.167 percent with an RMSE of0.277350, which is pretty impressive. The model also has a Mean Test RMSE of 0.281145 when using k-fold cross-validation to test it.which is very close to the RMSE given by our model initially. Finally using the **XGBoost'splot_importance()** function we have found out that the feature **MDVP.Fo(Hz) (i.e. Average vocal fundamental frequency)** has the highest importance score among all the features. Thus the proposed model is a reliable model to detect Parkinson's disease due to its efficient accuracy rates.

Though the model works efficiently, this is limited by the richness of the dataset with which it is being trained. The selected dataset has only 197 instances, hence in the future if we use a dataset with more no samples it would help the model generalize even better.

References

[1] Naranjo, L., Perez, C. J., Campos-Roca, Y., & Martin, J. (2016).Addressing voice recording replications forParkinson's disease detection. Expert Systems with Applications, 46, 286-292.

[2] Gil-Martín, M., Montero, J. M., & San-Segundo,R. (2019). Parkinson's disease detection from drawing movements using convolutional neural networks.Electronics, 8(8), 907.

[3] Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K. M., Dorsey, E. R., & Little, M. A. (2015). Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. Parkinsonism & related disorders, 21(6), 650-653.

[4] El-Agnaf, O. M., Salem, S. A., Paleologou, K. E., Curran, M. D., Gibson, M. J., Court, J. A., ... &Allsop, D. (2006). Detection of oligomeric forms of α-synuclein protein in human plasma as a potential biomarker for Parkinson's disease.The FASEB journal, 20(3), 419-425.

[5] Memedi, M., Nyholm, D., Johansson, A., Pålhagen, S., Willows, T., Widner, H., ...& Westin,J. (2015). Validity and responsiveness of at-home touch screen assessments in advanced Parkinson's disease. IEEE Journal of biomedical and health informatics, 19(6), 1829-1834.

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

[7] Aishwarya, R., Chowdhary, M. S., Abinay, A., Phanindhra, K., Yamini, N., &Toppo, R. R. (2020, July). Cognizance the Action of Human by Applying Enhanced Techniques. In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-8). IEEE.

[8] Chowdhary, M. S., Aishwarya, R., Abinay, A., &Harikrishna, P. (2020, October). Comparing Machine-Learning Algorithms for Anticipating the Severity and Non-Severity of a Surveyed Bug.In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 504-509).IEEE.

[9] Davidson, C. M., de Paor, A. M., Cagnan, H., & Lowery, M. M. (2015). Analysis of oscillatory neural activity in series network models of Parkinson's disease during deep brain stimulation. IEEE Transactions on Biomedical Engineering, 63(1), 86-96.

[10] Hariharan, M., Polat, K., &Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease.Computer methods and programs in biomedicine, 113(3), 904-913.

[11] Chakraborty, S., Aich, S., Han, E., Park, J., & Kim, H. C. (2020, February). Parkinson's Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach. In 2020 22nd International Conference on Advanced Communication Technology (ICACT) (pp. 298-303).IEEE.

[12] Iakovakis, D., Hadjidimitriou, S., Charisis, V., Bostantjopoulou, S., Katsarou, Z., Klingelhoefer, L., ...&Hadjileontiadis, L. J. (2018). Motor impairment estimates via touchscreen typing dynamics toward Parkinson's disease detection from data harvested in-the-wild. Frontiers in ICT, 5, 28.

[13] Deepa, D., R. Vignesh, A. Sivasangari, SujaCherukullapurathMana, B. KeerthiSamhitha, and Jithina Jose. "Visualizing road damage by monitoring system in cloud." International Journal of Electrical Engineering and Technology 11, no. 4 (2020): 191-203.

[14] Deepa, D., Sachitananda Jena, Yadavalli Ganesh, M. S. Roobini, and AnithaPonraj. "Threat Level Detection in Android Platform Using Machine Learning Algorithms."In Advances in Electronics, Communication and Computing, pp. 543-551.Springer, Singapore, 2021.

[15] Sivasangari, A., D. Deepa, T. Anandhi, AnithaPonraj, and M. S. Roobini. "Eyeball based Cursor Movement Control." In 2020 International Conference on Communication and Signal Processing (ICCSP), pp. 1116-1119.IEEE, 2020.

[16] Ishwarya, M. V., D. Deepa, S. Hemalatha, A. VenkataSaiNynesh, and A. PrudhviTej. "Gridlock surveillance and management system." Journal of Computational and Theoretical Nanoscience 16, no. 8 (2019): 3281-3284.