

DETECTING PARKINSON'S DISEASE USING XGBOOST CLASSIFIER MACHINE LEARNING

Nikisha Jadhav^{*1}, Sangita Phad^{*2}, Snehal Kamble^{*3}

^{*1}*Bachelor of Engineering, Department of Information Technology, Vasantdada Patil Pratishthan's College of Engineering & Visual Arts, Mumbai, Maharashtra, India.

^{*2}*Bachelor of Engineering, Department of Information Technology, Vasantdada Patil Pratishthan's College of Engineering & Visual Arts, Mumbai, Maharashtra, India.

^{*3}*Bachelor of Engineering, Department of Information Technology, Vasantdada Patil Pratishthan's College of Engineering & Visual Arts, Mumbai, Maharashtra, India.

ABSTRACT

Parkinson's Disease (PD) is a progressive disorder central nervous system that affecting the various movements. Symptoms are different Tremors are common, but the disorder is also commonly causes stiffness and slowing of the movement. Using machine learning algorithm it's easy to use analyze the voice pattern variations to predicting the existence PD patients. This paper proposes the predictive model that effectively diagnoses PD with maximum accuracy using the dataset to extrapolated data from voice recordings of Parkinson's patients. xgboost is a new machine learning algorithm to design with speed and performance in mind. xgboost stands for extreme gradient boosting and is based on decision tree. This algorithm is used to predict the disease.

Keywords: Machine Learning Algorithm xgboost, Convolution Neural Network, Support Vector Machine.

I. INTRODUCTION

Parkinson's Disease is a chronic, progressive disease which affects the movement throughout the body. Parkinson's Disease Symptoms can be different of everyone. There are many symptoms of PD, including tremors, muscle rigidity, changes in facial expressions, handwritten drawings. Current diagnostic tests for PD are limited. There is no standard diagnostic tests, such as blood test so a clinician's opinion is required to neurological test make a diagnose. Parkinson's Disease has 5 stage to it and affects the more than 1 million individuals every year in India . This is a chronic and has no cure yet. PD is a neurodegenerative affecting dopamine producing neurons in the brain. Parkinson's disease is the second most common neurodegenerative disease affecting 1% of the population over 55 years of age

Machine Learning

Machine Learning is the data science process that allows computer to gain information insight to the pattern and existing data to predict the outcomes and trends of the data to programmed identified them. This can make tasks like diagnosing Parkinson's disease more automated efficient, and accurate with a robust machine learning model and also identify patterns and characteristics in the data in ways that humans may not noticed. Machine learning algorithm to train data input and use statistical analysis in order to create specific output. Machine learning has various algorithm to used predict the disease.

II. DESIGN AND IMPEMENTION

Methodology:

This section explained the different steps to achieve the prediction of Parkinson's disease using various machine learning. Data Gathering, Data preprocessing, Model selection, Training, Evaluation, prediction.



Figure 1: Workflow Of Model

1.1 Data Gathering:

The first step is data gathering. This step is very important because the quality and quantity of the data you gather will directly affects the level of your prediction model.

1.2 Data Preparation:

In this step data is visualized well to spot the relationship between the parameters present in the data so as to take the advantage of as well as to get the data.

1.3 Model Selection:

In this model selection there are various model is used to till date by research and scientist. Some are meant by image processing, some for sequences like text, numbers or patterns.

1.4 Training:

Training the dataset is the main task of machine learning. We will apply the data to progressively improve the selected model to predict the better actual result should be approx. to predict one.

1.5 Evaluation:

The metrics we have calculated are ROC, Accuracy , Specificity , Precision etc. which will highlights the best algorithm among all.

1.6 Prediction:

In this phase we finally get the model ready to detect the prediction of Parkinson's disease based on the given dataset.

Algorithm:

XGBoost(Extream Gradient Boosting):

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It is used in prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. It implements machine learning algorithms under the Gradient Boosting framework.

Flowchart:

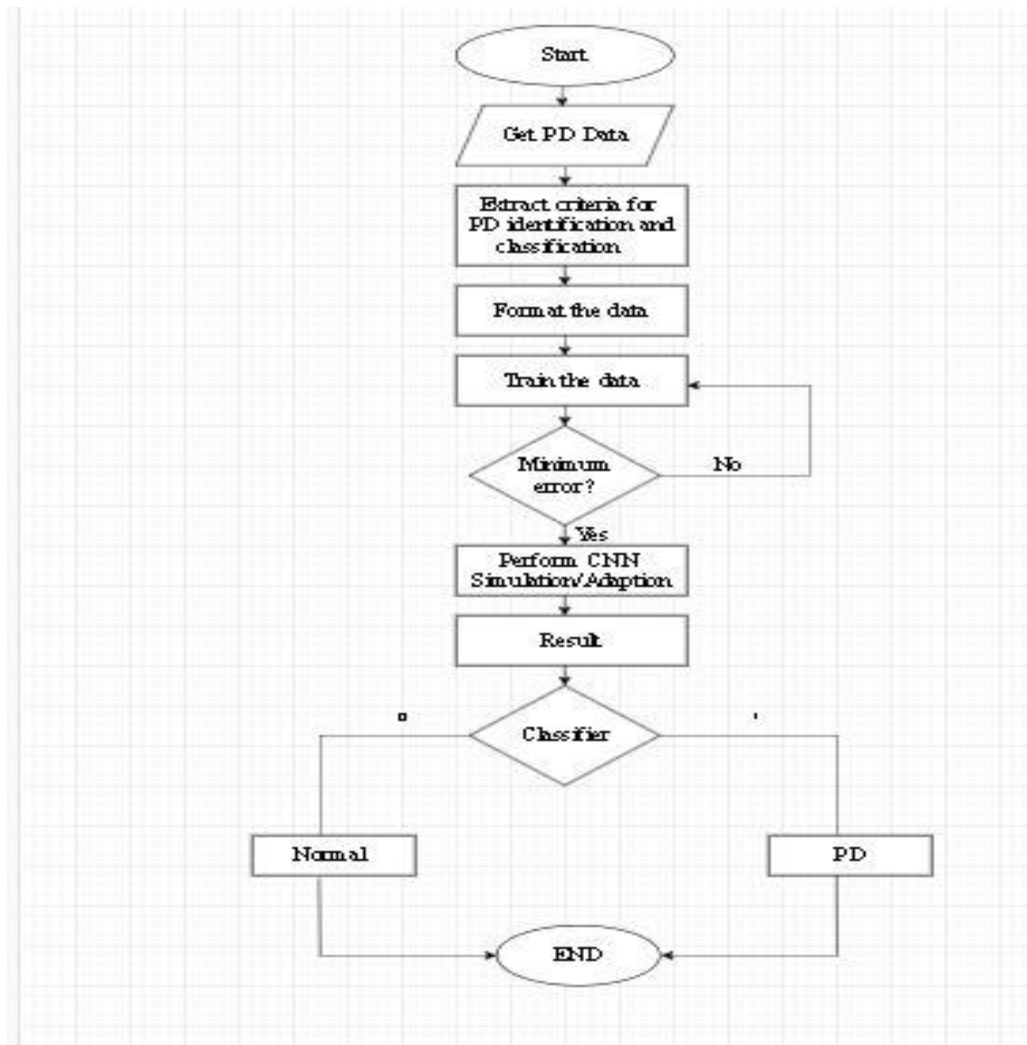


Figure 2: Flowchart

III. RESULTS AND DISCUSSION

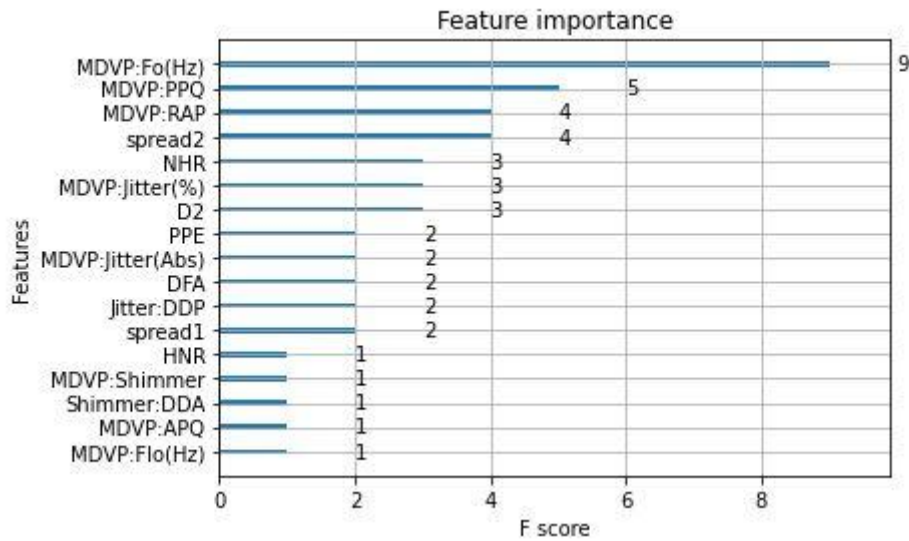
Audio Recording Dataset:

	name	MDVP:F0(Hz)	MDVP:F1(Hz)	MDVP:F2(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	...
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...

Audio Model Accuracy Score:

XGBoost model accuracy score: {0:0.4f} 0.9230769230769231

Audio Result:



Wave Drawing Result:

Random Forrest vs XGBoost Classifier

Accuracy:

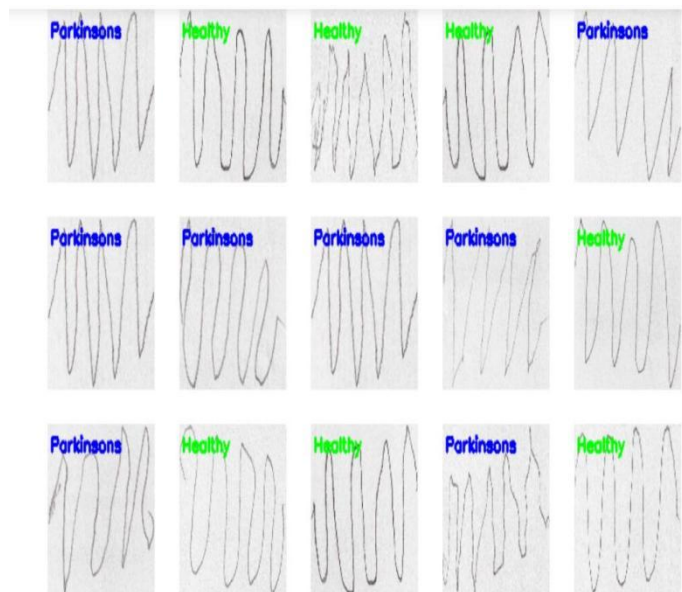
Random Forrest=76.67%, XGBoost=73.33%

Sensitivity:

Random Forrest=73.33%, XGBoost=73.33%

Specificity:

Random Forrest=80.00%, XGBoost=73.33%



Spiral Drawing Result:

Random Forrest vs XGBoost Classifier

Accuracy:

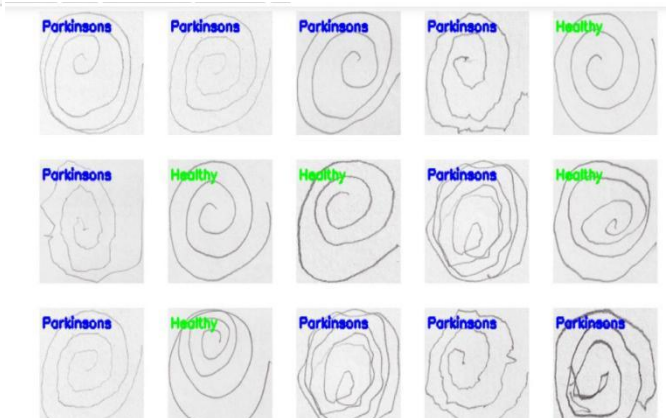
Random Forrest=86.67%, XGBoost=73.33%

Sensitivity:

Random Forrest=80.00%, XGBoost=73.33%

Specificity:

Random Forrest=93.33%, XGBoost=73.33%



IV. CONCLUSION

We have proposed an effective approach to generate and accurate predictive model for Parkinson's disease using xgboost classifier. This method is able to identify the PD subjects with an accuracy of 80-90%. From our extensive study, It is evident that the sustained vowels carry sufficient information to predict Parkinson's disease. In future studies, different feature selection or reduction methods can be examined to improve the classification accuracy

V. REFERENCES

- [1] Shivangi, Anubhav Johri and Ashish Tripathi Department of Computer Science Jaypee Institute of Information Technology 978-1-7281-3591-5/19/\$31.00 ©2019 IEEE "Parkinson Disease Detection Using Deep Neural Networks".
- [2] Aarushi Agarwal, Spriha Chandrayan and Sitanshu S Sahu Department of Electronics and Communication Engineering Birla Institute of Technology, Mesra, Ranchi, India 978-1-5386-5163-6/18/\$31.00 ©2018 IEEE "Prediction of Parkinson's Disease using Speech Signal with Extreme Learning Machine".
- [3] Enes Celik Department of Computer Science Kırklareli University Kırklareli, Turkey 978-1-7281-1013-4/19/\$31.00 ©2019 IEEE "Improving Parkinson's Disease Diagnosis with Machine Learning Methods".
- [4] Akshaya Dinesh North Brunswick Township High School Rutgers School of Engineering 978-1-5386-2534-7/17/\$31.00 ©2017 IEEE "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recordings".
- [5] Mosarrat Rumman Abu Nayeem Tasneem Sadia Farzana¹, Monirul Islam Pavel¹ and Dr. Md. Ashraful Alam¹ Department of Computer Science & Engineering BRAC University 66 Mohakhali, Dhaka, Bangladesh 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) "Early detection of Parkinson's disease using image processing and artificial neural network". © 2021, IRJET |