

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 11, November 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

0

Impact Factor: 8.379

9940 572 462

6381 907 438

🛛 🖂 ijircce@gmail.com

im 🛛 🙋 www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

XGB-Classifier for Detecting Parkinson's Disease

¹Mr. Uday Kumar G S, ²Ms. Lydia A, ³Mrs. K. Shandhini, ⁴Mrs. S. Saranya Devi

Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College, Kuppam, India¹

Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College, Kuppam, India²

Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College, Kuppam, India³

Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College, Kuppam, India⁴

ABSTRACT: Parkinson disease is a neurodegenerative disorder that affects nervous system and the root cause of it is falling rates of dopamine levels in the forebrain. The machine learning model is implemented to significantly improve diagnosis method of Parkinson disease. In this work it indicates that the ensemble techniques XGBoost classification (Extreme gradient boosting) algorithm achieved the high test accuracy rate (94.8%) compared to different classification algorithm. The performance of the methods has been assessed with a reliable dataset from UCI Machine learning repository

KEYWORDS: Parkinson disease, Random forest, XGBoost Classifier, Data analysis, ANN etc.

I. INTRODUCTION

Parkinson's disease is a progressive disorder of the central nervous system affecting movement and inducing tremors and stiffness. It has 5 stages to it and affects more than 1 million individuals every year in India. This is chronic and has no cure yet. It is a neurodegenerative disorder affecting dopamine-producing neurons in the brain. Biomarkers derived from human voice can offer insight into neurological disorders, such as Parkinson's disease (PD), because of their underlying cognitive and neuromuscular function. PD is a progressive neurodegenerative disorder that affects about one million people in the United States, with approximately sixty thousand new clinical diagnoses made each year. Historically, PD has been difficult to quantify and doctors have tended to focus on some symptoms while ignoring others, relying primarily on subjective rating scales. Due to the decrease in motor control that is the hallmark of the disease, voice can be used as a means to detect and diagnose PD. With advancements in technology and the prevalence of audio collecting devices in daily lives, reliable models that can translate this audio data into a diagnostic tool for healthcare professionals would potentially provide diagnoses that are cheaper and more accurate. We provide evidence to validate this concept here using a voice dataset collected from people with and without PD. This paper explores the effectiveness of using supervised classification algorithms, such as deep neural networks, to accurately diagnose individuals with the disease. Our peak accuracy of 94% provided by the machine learning models. The organizati-onal framework of this study divides the research work in the different sections. The Related work is presented in section 2. Further, in section 3 presented Parkinson disease, in section 4 and 5voice impairment dataand data Analysis discussed. in section 6, the description of machine learning and Implemen-tation are mentioned. In section 7, experimental work is shown with Machine learning models. Conclusion and future work are presented by last sections 8.

II. RELATED WORK

Many studies have been conducted on the detection of Parkinson disease using different machine learning algorithms on different datasets containing different features statistical measures. Shivangi[19], has used Deep dense ANNs (Artificial Neural Networks) on the voice recordings and CNNs (Convolutional Neural Networks) on VGFR dataset is used to predict the Parkinson disease. A three layer neural networks is used on the vertical ground force reaction data and a four layer neural network is used on voice data which is in a comma separated values (CSV) format for processing the data and making prediction.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

XGBoost and feature selection used by IqraNissar[10], and applied various machine learning models on the voice dataset and found that XGBoost with mRMR feature selection gave best results compared to other models. In the study it is observed that feature selection using minimum-Redundancy and maximum-Relevance has produced better accuracy results in making the prediction using XGBoost model.

Wu Wang[25], have used several indicators like rapid eye movement and olfactory loss, dopaminergic imaging markers and cerebrospinal fluid data for making the predictions using several machine learning algorithms KNN, SVM and ensemble techniques for getting better accuracy.

Hadoop and data mining used for data processing by RichaMathur[15], and presented a data mining methodology using Weka tool and machine learning algorithms ANN and KNN for making predictions on the processed data. In this study a layered big data framework is used for processing the un-structured data and Attributeselection filter provided by Weka tool for feature selection. A 10-fold cross validation technique is used for mining the data.

ShailRaval[18], has used Random Forest(RF) on static spiral for detecting tremor and used Ensemble techniques like RF (Bagging method), AB (Boosting method) and HV (Voting method) for increasing the proposed accuracy.

Securing data using cloud done by Sumathiet.al[22,23], and has published a paper on securing the medical information processing using scheduling strategy in cloud.

III. PARKINSON DISEASE

Parkinson Disease is a brain neurological disorder. It leads to shaking of the body, hands and provides stiffness to the body. No proper cure or treatment is available yet at the advanced stage. Treatment is possible only when done at the early or onset of the disease. These will not only reduce the cost of the disease but will also possibly save a life. Most methods available can detect Parkinson in an advanced stage; which means loss of approx... 60% dopamine in basal ganglia and is responsible for controlling the movement of the body with a small amount of dopamine. More than 145,000 people have been found alone suffering in the U.K and in India, almost one million populations suffers from this disease and it's spreading fast in the entire world.

Parkinson's disease symptoms can be different for everyone. Early signs are mild that goes unnoticed. Symptoms usually begin on one side of your body and gets worsen on that side, afterwards it affects both the sides.

Parkinson's symptoms may include

- Tremor
- Slowed movement
- Rigid muscles.
- Impaired posture and balance.
- Loss of automatic movements
- Speech changes
- Writing changes
- Depression
- Anxiety
- Sleeping, and memory-related issues
- Loss of sense of smell along with balance problems.
- What causes Parkinson's disease is still unclear, but researchers found that several factors are responsible for triggering the disease.

The Parkinson's disease is due to a loss of neurons that produce a chemical messenger in the brain called dopamine. When there is a decrease in level of the amino acid named dopamine it leads to the abnormal brain activity, which leads to Parkinson's disease.

The cause of Parkinson's disease is still a question mark, but several factors appear to play a role, including:

- 1. *Genes* Certain mutation genes have been found by research that are very rare. The gene variants often increase the risk of Parkinson's disease but have a lesser effect on each genetic marker.
- 2. *Environment* Due to certain harmful toxins or chemical substances found in the environment can trigger the disease but have a lesser effect.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

3. Triggers

As a result people suffer from this disease for many years before diagnosis. The estimated results have shown that there are 7-10 million people are affected by Parkinson's disease worldwide. People with age above 50 are the one's who has the higher possibility of getting Parkinson's disease but still an estimated 4 percentage of people who are under the age 50 are diagnosed with Parkinson's disease. There is no cure or prevention for PD. However, the disease can be controlled in early stage. The data mining techniques is used as a effective way for early detection and diagnosis of the disease. Data mining techniques in medicine is a research area that combines sophisticated representational and computing techniques with the insights of expert physicians to produce tools for improving healthcare.Data mining is a statistical method for finding hidden patterns in datasets by constructing predictive or classification models that can be learned from past experience and applied in future cases, so there is a need for a more accurate, objective means of early detection, ideally one which can be used by individuals in their home setting.

IV. VOICE IMPAIRMENT DATA

• Data Set Information:

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data

Is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. Table4.1 explains the list of features description in dataset.

• Dataset Description

Attribute	Description		
MDVP:Fo(Hz)	Average vocal		
	fundamental frequency		
MDVP:Fhi(Hz)	Maximum vocal		
	fundamental frequency		
MDVP:Flo(Hz)	Minimum vocal		
	fundamental frequency		
MDVP:Jitter(%),MDVP	Several measures of		
Jitter(Abs),MDVP:RAP,MDVP:P	variation in fundamental		
PQ,Jitter:DDP	frequency		
MDVP:Shimmer,MDVP:Shimme	Several measures of		
r(dB),Shimmer:APQ3,Shimmer:	variation in amplitude		
APQ5,MDVP:APQ,Shimmer:DD			
Α			
NHR,HNR	Two measures of ratio of		
	noise to tonal		
	components in the voice		
Status	Health status of the		
	subject (one) -		
	Parkinson's, (zero) –		
	healthy		
RPDE,D2	Two nonlinear		
	dynamical complexity		
	measures		
DFA	Signal fractal scaling		
	exponent		
	spread1,spread2		
PPE	Three nonlinear		

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

measures of fundamental
frequency variation

Table 4.1: showing attribute information of dataset

Featur	М	Std	Min	25%	50%	75%	Max	
e	ea							
	n							
MDVP	19	154.	41.3	88.3	117.	148.	182.	260.
:Fo(Hz	5.0	2286	9006	3300	5720	7900	7690	1050
)		4			0	0	0	0
MDVP	19	197.	91.4	102.	134.	175.	224.	592.
:Fhi(H	5.0	1049	9154	1540	8625	8290	2055	0300
z)	19	1 116.	12.5	0	0 84.2	0 104.	0 140.	0 239.
MDVP	5.0	3246	43.5 2141	65.4 7600	84.2 9100	104. 0185	140. 0185	239. 1700
:Flo(H z)	5.0	3240	2141	7000	9100	0185	0185	0
MDVP	19	0.00	0.00	0.00	0.00	0.00	0.00	0.03
:Jitter(5.0	622	484	168	346	494	736	316
%)								
MDVP	19	0.00	0.00	0.00	0.00	0.00	0.00	0.03
:Jitter(5.0	622	484	168	346	494	736	316
%)								
MDVP	19	0.00	0.00	0.00	0.00	0.00	0.00	0.02
:RAP	5.0	330	296	068	166	250	383	144
MDVP	19	0.00	0.00	0.00	0.00	0.00	0.00	0.02
:RAP	5.0	330	296	068	166	250	383	144
MDVP :RAP	19 5.0	0.00 330	0.00 296	0.00 068	0.00 166	0.00 250	0.00 383	0.02 144
MDVP	19	0.02	0.01	0.00	0.01	0.02	0.03	0.11
:Shim	5.0	0.02 970	885	0.00 954	650	297	788	908
mer	5.0	210	005	221	050	277	/00	200
MDVP	19	0.28	0.19	0.08	0.14	0.22	0.35	1.30
:Shim	5.0	225	487	500	850	100	000	200
mer(d								
B)								
Shim	19	0.04	0.03	0.01	0.02	0.03	0.06	0.16
mer:D	5.0	699	045	364	473	836	079	942
DA	10	0.02	0.04	0.00	0.00	0.01	0.02	0.21
NHR	19 5.0	0.02 484	0.04 041	0.00 065	0.00 592	0.01 166	0.02 564	0.31 482
HNR	19	21.8	4.42	8.44	19.1	22.0	25.0	33.0
	5.0	8597	4.42 576	100	9800	8500	25.0 7550	4700
RPDE	19	0.49	0.10	0.25	0.42	0.49	0.58	0.68
	5.0	853	394	657	130	595	756	515
DFA	19	0.71	0.05	0.57	0.67	0.72	0.76	0.82
	5.0	809	533	428	475	225	188	528
Sprea	19	-	1.09	-	-	-	-	-
d1	5.0	5.68	020	7.96	6.45	5.72	5.04	2.43
		439		498	009	086	619	403
Sprea	19	0.22	0.08	0.00	0.17	0.21	0.27	0.45
d2	5.0	651	340	627	435	888	923	049

 Table 4.2: Showing statistical measures of dataset

The dataset which is used for analysis is created by Max Little of the University of oxford, in collaboration with the National Centre for Voice and speech, Denver, Colorado, who recorded the speech signals. This data set consists of a range of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

biomedical voice measurement with 195 samples of features from 31 people, 23 with Parkinson's disease (PD) and 8 of them are the control group. The data is in ASCII CSV format. A series of features was extracted which can be categorized as : name of the patients, three types of fundamental frequency(high, low and average), several measures of variation in fundamental frequency(jitter and its type) several measures of variation of amplitude(shimmer and its type), two measures of ratio of noise to tonal components in the voice(NHR and HNR), Two nonlinear dynamically complexity measure-es(RPDE,D2), DFA is a signal fractal scaling exponent, and Three nonlinear measures of fundamental frequency variation(spread1, spread2, PPE). The statistical measures of features in dataset are shown in table4.2. The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. There are around 6 recording per patients. The data set has about 75% of cases suffering from Parkinson disease and 25% of cases which are healthy as shown in the fig.4.1 with the help of bar plot.

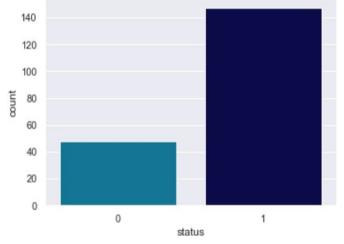


Fig.4.1: Show value counts for two categorical 0 for healthy and 1 for Parkinson's

• Feature Analysis

Feature importance refers to a class of techniques for assigning scores to input features to a predictive model that indicates the relative importance of each features when making a prediction. Feature important analysis provide insight into the dataset. The relative scores can highlight which features can be more relevant to the target. Feature importance analysis can provide insight into the model. Most important scores are calculated by a predictive model that has been fit on the dataset. In fig.4.2.it describes how different features in the Parkinson's disease dataset correlated to each other.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

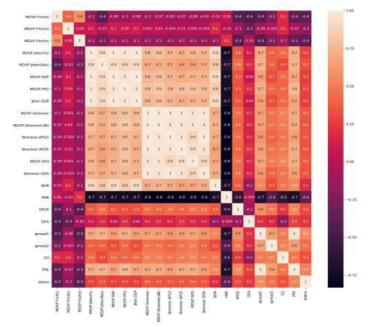


Fig.4.2: Correlation between features of parkinson's disease

V. ANALYSIS

Data Analysis is the methodical approach of applying the statistical measures to describe, analyse, and evaluate data. The researchers analyse patterns and relationships among variables. Univariate, Bivariate, and Multivariate are the major statistical techniques of data analysis.

A. UNIVARIATE ANALYSIS

Univariate analysis is the easiest methods of quantitative data analysis. As the name suggests, "Uni," meaning "one," in univariate analysis, there is only one dependable variable. It is used to test the hypothesis and draw inferences. The objective is to derive data, describe and summarize it, and analyse the pattern in it.

MDVP:Fo(Hz) Mean - Median = 5.4386

MDVP:Fo(Hz) 3 Standard Deviation below mean 30.0584

MDVP:Fo(Hz) 3 Standard Deviation above mean 278.3988

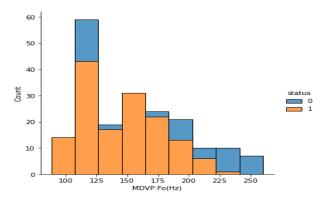
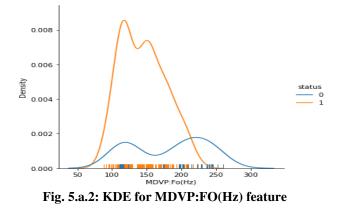


Fig. 5.a.1: Histogram Plot for MDVP:FO(Hz) feature

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |



Measures of variation of fundamental frequency ispositively skewed. For some people, average vocal fundamental frequency is higher than the average. Mainly because it is higher for healthy persons as seen in fig.5.a.2. Frequencies ranging from 100Hz to 200Hz are the ones of persons with PD. Especially 145Hz to around 160Hz frequency ranges are for PD persons only. There is good separation between two kde graphs indicating that this is a good indicator. As we can observe in Kde rug, values are densed near 150Hz on both sides suggesting need to look more here. Also two peaks suggests that this data might have come from two different sources. There are no outliers here.

- PPE Mean Median = 0.0125
- PPE 3 Standard Deviation below mean -0.0638
- PPE 3 Standard Deviation above mean 0.477

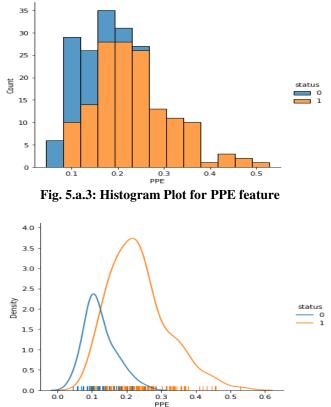


Fig. 5.a.4: KDE for PPE feature



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

It is a nonlinear measure of fundamental frequency variation. From 0.1 to 0.5 is the range which correlates to persons with PD. This is a good indicator as there is separation. As we can observe in fig.5.a.4, values are densed between 0.2 & 0.3 suggesting need to look more here. It might have outliers.

So out of all independent features only PPE, spread2, spread1, MDVP:Fo(Hz) are the features which looks like good indicators as of now

B. BIVARIATE ANALYSIS

In Bivariate Analysis, there are two variables wherein the analysis is related to cause and the relationship between the two variables.

Analysis Of Bar Plots:

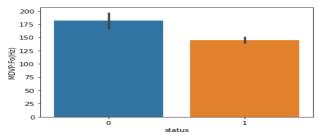


Fig. 5.b.1: Bar Plot for MDVP:Fo(Hz) feature

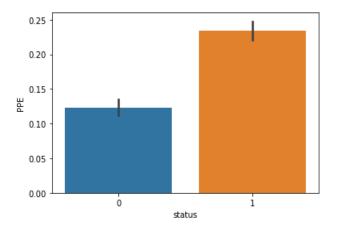


Fig. 5.b.2: Bar Plot for PPE feature

Here it is observed in fig.5.b.1 and fig.5.b.2, that features PPE, spread2, NHR, PPE, spread2, spread1, MDVP:Fo(Hz), Shimmer:APQ3, MDVP: Shimmer(dB), MDVP:Shimmer, Jitter: DDP, MDVP:RAP, MDVP:PPQ, MDVP:Jitter(%), MDVP:Jitter(Abs) looks encouraging because of the visible significant difference between the two classes.



Analysis Of Pairplot

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

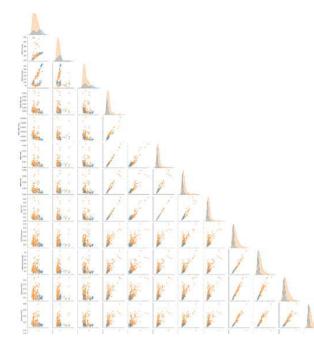


Fig 5.c: Pairplot for the features in dataset

Looking at diagonals in fig.5.c: MDVP:Fo(Hz), MDVP:Fl-o(Hz), spread1,spread2,PPE looks to have some separation. These are the features which looks to be good indicators of our target class Some independent features have linear relation with other independent features:MDVP:Jitter(%), MDVP:Jitter(A-bs),MDVP:RAP,MDVP:PPQ,Jitter:DDPhave some linear relation among each other. All of these are measures of variation in fundamental frequency so it would be a better approach to keep one of them and drop the remaining features.Similarly, MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:A-PQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA also have linear relation among each other. All of these are measures of variation in fundamental amplitude so it would be a better approach to keep one of them and drop the remaining features.

VI. IMPLEMENTATION

In this Python machine learning project, using the Python libraries scikit-learn, numpy, pandas, and XGBoost, a model is built using an XGB Classifier. The data is loaded, got the features and labels, scaled the features, then the dataset is split, built an XGB Classifier, and then calculated the accuracy of our model. It is a disease which is a disorder in the nervous system. Parkinson's disease affects the movement of the human body. In today's world, around 1 million people are suffering from this disease. This is a disorder which produces neurodegenerative dopamine-producing neurons in the brain. The following system will detect Parkinson's symptoms in the human body.

The project will be made by a new machine learning algorithm called the XGBoost. XGBoost is a new Machine Learning algorithm designed with speed and performance in mind. XGBoost stands for Extreme Gradient Boosting and is based on decision trees. In this project, we will import the XGB Classifier from the XGBoost library; this is an implementation of the scikit-learn API for XGBoost classification. To build a model to accurately detect the presence of Parkinson's disease in an individual.

A. Methodology

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance that is dominative competitive machine learning. XGBoost provides a wrapper class to allow models to be treated like classifiers or regressors in the scikit-learn framework. This means we can use the full scikit-learn library with XGBoost models.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

XGB Classifier

The XGBoost model for classification is called XGB Classifier. Initialize an XGB Classifier and train the model. This classifies using eXtreme Gradient Boosting- using gradient boosting algorithms for modern data science problems. It falls under the category of Ensemble Learning in ML, where we train and predict using many models to produce one superior output. We can create and and fit it to our training dataset. Models are fit using the scikit-learn API and the model.fit() function. To detect the presence of Parkinson's Disease in individuals using various factors. We used an XGB Classifier for this and made use of the sklearn library to prepare the dataset.

Training Features

- 1. When searching for best feature value for node split, XGBoost provides an option to search on the feature value's quantiles or histogram instead of try all the feature values to split node.
- 2. When building feature histogram, XGBoost may split feature data into multiple computers to calculate histogram, then merge back to generate a aggregate histogram, this like Hadoop Map-reduce operation, and the generated histogram will be cached for next split.
- 3. XGBoost can automatically handle missing values in feature. In tree node split step, XGBoost will either assign all missing value instances to left or right child, depend on which side has larger gain.
- 4. XGBoost provide lots hyper-parameters to deal with overfitting

B. XGBoost Analysis

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now. Please see the chart below for the evolution of tree-based algorithms over the years.

Step1: Finding similarity scores for each groups (left and right) using the equation Similarity score = $\frac{(\sum Residuali)^2}{\sum [Pi \times (1 - Pi) + \lambda]}$ Step2: Calculating Gain to decide the thresholds for better split Gain = Left similarity + Right similarity- Root similarity
Step3: Calculating Cover for each leaf node to determine minimum number of residuals in a node in the tree using the Cover equation $Cover = \Sigma [Pi \times (1 - Pi)]$
Step4: Tree Pruning is made based on Gain value, using the below equation Gain $-\gamma$ Pruning is made if difference is Negative Step5: Output values are calculated for every leaf node in the tree Output = Σ [Pi × (1 – Pi) + λ
Step6: Calculating new prediction for the constructed tree New prediction = Initial predicted value + learning rate (ε) ×

Fig 6.1: Algorithm for XGBoost Decision Tree

We could simply compare the new residuals and found that whether we have taken a small step in the right direction. We keep building other trees based on new residuals and make new prediction gives smaller residuals until residuals are small or reached maximum number.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

VII. EXPERIMENTAL RESULTS

This section shows the analysis of different learning models through the experimental work. The paper conducts the experiments on different well known learning models using Voice impairment data taken.

The Table7.1.Presents the comparison of various learning models used in the experiment. All the learning models are doing well with minimum accuracy of 82.05 % gained by SVC and 94.87 % gained by XGBoost classifier.

Models	Samples	ML Classifier	Test Accuracy
Existing Method	50	SVM XGBoost	81.16 77
Proposed Method	50 50 50	XGBoost Classifier Logistic Regression SVC	94.87 89.74 82.05

Table 7.1: The Comparison tables of different models

VIII. CONCLUSION

The analysis which algorithm provide the high accuracy of prediction for the Parkinson's disease dataset, here the classification accuracy was studied and compared, with good performance and fast implementation XGBoost with multiple fold data achieved a high accuracy with 94.87%. This system provides the comparison between machine learning classifiers of LR and XGBoost in PD disease diagnosis with high dimensional data.

Future work

In future work, we can focus on different techniques to predict the Parkinson disease using different datasets. In this research, we using binary attribute (1- diseased patients, 0-non-diseased patients) for patient's classification. In the future we will use different types of attributes for the classification of patients and also identify the different stages of Parkinson's disease

REFERENCES

- 1. Abdullah Caliskan, "Diagnosis of the Parkinson Disease by using Deep Neural Network Classifier.", (2017), IU-JEEE, Vol. 17(2), (2017), 3311-3318.
- 2. C.A. Antoniades, "The relationship between abnormalities of saccadic and manual response times in Parkinson's disease," (2013), Journal of Parkinson's disease, vol. 3, no. 4, pp. 557–563.
- 3. Das R. "A comparison of multiple classification methods for diagnosis of Parkinson disease", (2010), Expert Syst Appl. 2010; 37(2):1568–72.8.
- 4. Djuric-Jovicic, M., "Finger tapping analysis in patients with Parkinson's disease and atypical parkinsonism.", (2015), J. Clin. Neurosci. 30, 49–55. doi: 10.1016/j.jocn.2015.10.053.
- 5. Ene M., "Neural network-based approach to discriminate healthy people from those with Parkinson's disease.", (2008), Ann Univ Craiova Math ComputSci Ser. 35:112–6.
- 6. GeethaRamani R, "Parkinson Disease Classification using Data Mining Algorithms.", (2011), Int J ComputAppl (0975–8887).2011;32(9):17–22.
- 7. Gil D, "Diagnosing Parkinson by using artificial neural networks and support vector machines0", (2009), Glob J ComputSci Technol. 2009;9(4):63-71.
- 8. HongyoonChoia, "Refining diagnosis of Parkinson's disea-se with deep learning based interpretation of dopamine transporter imaging.", (2017), NeuroImage Clinical, 16 (2017) 586–594.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111054 |

- 9. Imanne El Maachi, "Deep 1D-Convnet for accurate Parkin-son disease detection and severity prediction from gait", (2019), Published by Elsevier Ltd., doi: https://doi.org/10.1016/j.eswa.2019.113075
- IqraNissar, "Voice-Based Detection of Parkinson's Disease through Ensemble Machine Learning Approach: A Performance Study", (2020), EAI Endorsed Transactions on Per-vasive Health and Technology, doi: 10.4108/eai.13-7-2018.162806
- 11. John Prince, "A Deep Learning Framework for the Remote Detection of Parkinson's Disease Using Smart-phone Sensor Data", (2018), IEEE, 978-1-5386-3646-6/18
- Kamal Nayan Reddy Challa, "An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Technique.", (2016), International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016.
- 13. Little M, "Suitability of dysphonia measurements for telem-onitoring of Parkinson's disease", (2008), Nat Prec. 2008. https://doi.org/10.1038/npre.2008.2298.1.
- 14. Luca Parisi, "Feature-driven Machine Learning to Improve Early Diagnosis of ParkinsonÕs Disease", (2018), EXPERT SYSTEMS WITH APPLICATIONS, ELSEVIER, doi: 10.1016/j.eswa.2018.06.003.
- 15. RichaMathur, "Parkinson Disease Prediction Using Machine Learning Algorithm", (2019), Springer Nature Singapore Pte Ltd., 10.1007/978-981-13-2285-3_42
- 16. Rustempasic I, "Diagnosis of Parkinson's disease using Fuzzy C-means clustering and pattern recognition.", (2013), SouthEastEur J Soft Comput. 2013. https:// doi. org/ 10. 21533/ scjournal. v2i1. 44.
- SatyabrataAich, "A Nonlinear Decision Tree based Class-ification Approach to Predict the Parkinson's disease using Different Feature Sets of Voice Data.", (2018), International Conference on Advanced Communications Technology (ICACT). ISBN 979-11-88428-01-4.
- ShailRaval, "A Comparative Study of Early Detection of Parkinson's Disease using Machine Learning Techniques", (2020), IEEE Xplore, Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0
- 19. Shivangi, "Parkinson Disease Detection Using Deep Neural Networks", (2019), IEEE, 978-1-7281-3591-5/19
- 20. Srishti Grover, "Predicting Severity Of Parkinson's Disease Using Deep Learning", (2018), International Conference on Computational Intelligence and Data Science (ICCIDS 2018). Procedia Computer Science 132 (2018) 1788–1794.
- 21. Srishti Grover, "Predicting Severity Of Parkinson's Disease Using Deep Learning, (2018), Published by Elsevier Ltd., 10.1016/j.procs.2018.05.154
- 22. Sumathi D, Poongodi P, "AN improved scheduling strategy in cloud using trust based mechanism", (2015), Int. J. Comput. Electr. Autom. Control Inf. Eng., 9(2), 637-641
- 23. Sumathi D, Poongodi P, "Secure medical information processing in cloud: Trust with swarm based scheduling", (2016), Journal of medical imaging and health informatics 6(7), 1636-1640
- 24. Sunil Yadav, "Hybrid Machine Learning Classifier and Ensemble Techniques to Detect Parkinson's Disease Patients", (2021), Springer Nature Singapore Pte Ltd, https://doi.org/10.1007/s42979-021-00587-8
- 25. Wu Wang, "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning", (2020), IEEE Access, DOI:10.1109/ACCESS.2020.3016062











INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

🚺 9940 572 462 应 6381 907 438 🖂 ijircce@gmail.com



www.ijircce.com