**RESEARCH ARTICLE** 

**OPEN ACCESS** 

## A Data Mining Tool In Neurlogical Disorder Prediction Using Feature Selection Techniques

P. Sravanthi<sup>[1]</sup>, B.Sai Sudha<sup>[2]</sup>, P.Sai Ganesh<sup>[3]</sup>, M.Dinesh<sup>[4]</sup>,

T.Ravi Kumar<sup>[5]</sup>, G.Rajasekharam<sup>[6]</sup>

Department of Computer Science and Engineering, Nadimpalli Satyanarayana Raju Institute of Technology, Andhra Pradesh,India

## ABSTRACT

Data mining techniques are used for a variety of applications. In healthcare Industry, Data mining plays an important role in predicting diseases. For detecting a disease number of tests should be required from the patient. By using Data mining techniques the number of tests can be reduced. This reduced test plays an important role in time and performance. Neurological disorders are diseases of the brain, spine and the nerves that connect them. The increasing capabilities of technologies are generating massive volumes of complex data at a rapid pace. Evaluating and diagnosing disorders of the nervous system is a complicated and complex task. Many of the same or similar symptoms happen in different combinations among the different disorders. Here, we provide a developed selected data mining methods in the area of neurological diseases diagnosis. Hence, it will help experts to gain an understanding of how data mining techniques can assist them in neurological diseases diagnosis and patients treatment. Gradient boosting tree algorithm is used to predict the diseases as it produces the accurate results. Here, our aim is to find the performance of different classification methods of large database.

Keywords:- Data Mining, Classification Rules, Voice dataset, Bar Graph.

## I. INTRODUCTION

## PROBLEM STATEMENT

Neurodegenerative disorders are the results of the progressive tearing and neurons loss in different areas of the nervous system. Neurons are the functional unit of brain .They are contiguous rather than continuous. A good healthy looking neuron as shown in fig 1 has extensions called dendrites or axons, a cell body and a nucleus that contains our DNA. DNA is our genome and hundred billion neurons contains our entire genome which is packaged into it .When a neuron get sick, it loses its extension and hence its ability to communicate which is not good for it and its metabolism become low so it starts to accumulate junk and it tries to contain the junk in the little packages in little pockets .When things become worse and if the neuron is a cell culture it completely loses its extension, becomes round and full of the vacuoles.

This work deals with the prediction of Parkinson's disorder which is now a days is a tremendously increasing incurable disease. Parkinson's disease is most spreading disease [19] which get its name from James Parkinson who earlier described it as a paralysis

agitans and later gave his surname was known as a PD. It generally affects the neurons which is responsible for overall body movements. Main chemicals are dopamine and acetylcholine which affects human brain.

There are various environmental factor which have been implicated in PD [20].below are the listed factor which caused Parkinson's disease in an individual.

## **II. OBJECTIVES**

The main objective is to predict the prediction efficiency that would be beneficial for the patients who are suffering from Parkinson and the percentage ratio will be reduced. Generally in the first stage Parkinson can be cured by the proper treatment. So it's important to identify the PD at the early stage for the betterment of the patients. The main purpose of this research work is to find the best prediction model i.e. the best machine learning technique which will distinguishes the Parkinson's patient from the healthy person. The techniques SVM, Adaboost, navie bayes, gradient descent, gradient boosting tree We have found that Neural network ,SVM, Linear Regression have been reported in various researches, whereas it has been found that only few researchers have explored Adaboost and gradient boosting tree. The experimental study is performed on the biomedical voice measurement from 31 people, 23 with Parkinson's disease. The prediction is evaluated using error rates. Further the Feature selection technique has been implemented with the aim to get the important features that can detect the Parkinson's disease.

#### **III. SCOPE**

**SCOPE**By using this feature selection technique we can predict the disease In initial stage and we can save the life of effected person Prediction of Parkinson disorder is one of the most important problem that has to be detected in



Fig 1:Structure of neuron present in human brain

the early phases of the commencement of the disease so as to reduce the disease progression rate among the individuals .Various researches have been made to find the basic cause and some have reached to the heights by proposing a system which differentiates the healthy people from those with any parkinson's data set using various machine learning techniques. Lots of preprocessing, feature selection and classification techniques have been implemented and developed in the past decades. Following is the given work done in the prediction of Parkinson's disorders.

## Machine learningalgorithms

- 1. Navie baye's
- 2. Svm (support vector machine)
- 3. Sgd (stochastic gradient descent)

4. Gbt (gradient boosting tree)

#### 1. Naive Baye's Classifiers

This article discusses the theory behind the Naive Bayes classifiers and their implementation.

Naive Bayes classifiers are a collection of classification algorithms based on **Baye's Theorem**.It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other

1.00										petra	en - Micro	off Doel									. 0
	mark 11	also relay	Particular	100	area Va	-															-
A (M	.0	-	(B) + (j	60010E		1	ne Teir	General		1.1	192	Normal	30	et i	Good		7 3	(1)	X accord	8 3	
-4 (00)	15		100				THE R. LANS.	1 100.00		Centha	-	Revetad			1000	(C	mark from	a long	A 18-	Gards Fred R.	
				-						April 1	g * Table *					_				Hiller 1 Select 1	
				-	_															-	-
- 40		//	1944														-				-
	100000	C.	and the second	100.0	100.0.00			1	100.000.00				N.				A	DIT A		1 1	- W.
			and a second		ALLEY LASS		ALL PROPERTY.											1000		the spread	
01,000	123.996	234.342	14.774	100704	0.00047	0.0097	0.00394	440.247	9.04574	0.420		0.9315	0.02971	0.08242	0.02201	21.053	1.0.45	100 0.00	10080 44	Tate 0.58049	1 2.900
19,901	122.4	248.83	114.879	110,000	0.00003	0.00405	110690	0.01394	0.060.04	0.625		0.04518	0.04,000	1.09403	0.02229	18.085			00/1 -4.0		3 2,4858
- 001	110.000	100.000			6.00000	0.00003	0.00000		0.00000	0.462	0.000701				0.0000	20.444			10.00	ALL OTHER	
10,000	222.004	100.075		LOUGHT .	0.00000	8.000004	0.00000		a covera	OLULY.	0.0000	0.04003		Contract of the local division of the local	0.003112		1 0.45		shada it.	LLTP GLIPHIC	
100	138.034	249.781	139.633	1002284	100011	1.00855	0.00908	8.412999	946425	0.984	0.0947	0.04425		0.2741	0.01791	10.000			1940a (9.0	A119 0.23400	
1,000	120.552	101.164	111.797		0.00001			0.001398	0.04705	0.455		0.03326			0.05222	21.378	1 0.42		10000 -4.7		1 1.04
- 503	103 111		124 224	0.0000	0.000033			A 00473	A DOMEST	0.124	4.06430	0.000.00			0.000044	24,403	1 04		1000		1.1.1.1
	100.000	100.00		2.0001	0.00000	0.00000	a contract	0.00000	0.00000	CL B. De	1.00020	0.000900	0.012.00		0.000044	and the second		total date	them to	PROFE OLIVIA	
	33.73	100.000	10.774	0.00000	1.00009	10000	944932	0.0088	0.00055	0.191	4.07775			0.09216		13.412	1 0.00	1903 Q.74	3380	9000 0.12770	2 6 4 6
1,100	93.098	100.000	11.129	100534	100009	0.00094	0.000342	0.00003	9.008.00	0.235	0.0140	9.017.0	0.00444	0.04374	0.00022	21.063	1.0.54	No. 10, 75	19462 0.0	1188 0.12500	5 2.43
1,740		110.74	85.077		0.00099	0.00054	0.0003	0.00703	0.02543	0.1997		0.01342			0.011000	25.110		1110 0.55	006 24		
101,101	12,004	113,6/1	86.192	0.0054	0.00000	0.00081	0.00030	0.00044	0.00752	0.249	0.00424	0.00041	0.02714	0.04272	0.01141	23.414	1 0.5	1320 0.1	4.02 -4.9	3003 0.35306	1 2.54
0,910	138.928	254.854	101.279	100294	1.00002	1.00554	010054	0.00000	0.05299	0.515	0.00956	0.00717	0.0114	0.01998	0.00530	25.703	1 0.	2005-0.54	10028 - 11.5	1715 0.15261	3 2.94
1,100	100.000	179.104	2.22	0.0077	0.00003	0.00993	0.00000	1.00000	9.09942	0.154	9.04726	0.000332	0.01707	0.02104	0.00041	24,009	1 0.45	1000 0.00	10003 -0.4	3002 0.23408	1 2.30
1. 16.0	102.045	285.265	12.675	5100794	0.00007	0.00121	0.00049	0.10000	91138.78	0.558	0.077000	0.00977	0.01240	0.00000	0.000479	26.522	1 0.67	1241 (1.6)	PROF. (B.	THE OWNER	3,836
10,100	147.105	227.433	81.129	couses	0.00003	0.00057	0.00003	0.004/1	0.01501	0.125	0.00772	0.00000	0.01309		0.00839	10.110	1.0.56	NUM OF	15/40 0.3		5 2.30
0,997	101.188	101.174	82.764	0.0544	1.00004	010031	0.00242	0.006.02	010047	0.547	0.064%	0.013	0.02076	0.02968	11069	22.513	1.0.9	PUB GP	1002 -0.0	NERE OF TAXAB	1 2 26
0,901	188.778	232.341	75.603	0.00758	0.00004	0.00284	0.00587	0.00855	0.03327	0.348	0.03441	0.03893	0.0345	0.04322	0.02910	20.376	1 0.65	1000 0.80	19417 -2.9	3117 0.43432	3 3.00
1,000	100.044	171.0.0		1000712	E18963	P.A.CORT	010032	a passa	0.05517	11.940	0.00471		0.01/01		0.0110	10.00	1 0.88	1008 0.71	1982 .7.9	THE CLEAR	
0,000	128,465	188.394	342.022	100753	000005	8,005/2	0.00000	610132	0.03995	0.546	0.01721	0.02376	0.0431	0.03104	0.033000	17.153	1.0.54	KO04 0.0	10008 -4.5	3447 0.34017	1 2 8 9
1,311	121.80	285.748	85.782	0.0084	0.00005	010028	0.1642	0.00385	O Dates	0.126	0.03967	0.00.661	0.04055	0.05	0.000/1	17.546	1 0.84	8.0 0.0	HERO -4.0	2044 0.76704	1 60
10,000	153.84	172,84	28.128	1.0044	0.00003	0.00032	0.00207	0.00095	0.04137	0.37	0.00021	0.02391	0.04325	0.06003	0.05849	11.493	1 0.12	8117 0.81	10001 -0.1	3006 0.13762.	2 2.55
10,000	107.83	101.221		100042	*******	4.0022	8.16247	******	wareful .	0.377		1.0256			will/8	44.46	+ 614	rme Gal	man d.A	men to Jh2180	
10,100	173,917	106.735	97.24	000439	100003	0.00221	0.10258	9.10063	9.04292	0.956	9.102587	0.2247	-1.03772	0.06562	=1084	20.422	1 0.55	0.84	100x 5.2	ware 0.1102.0	3 234
1,000	184.659	200.041	14.774	000047	1.00025	0.00.00	0.0039	0.1034	a Daras	0.594	0.007.84	0.00946	0.05447	0.02714	0.01779	21.811	1 0.38	Petr 0.71	1969 -5.5	3765 0.7708	5 2.64
1,000	234.4	208.007	77,968	0.00633	0.00000	0.00000	0.00075	0.00948	0.03707	0.581	0.01732	0.02265	0.0376	0.05101	0.02567	22,000	1 0.52	2748-0.72	17948 0.5	7184 0.23985	3 2.84
1,940	DIDH	108.111	75.541	0.00422	110003	0.1025	110034	0.0075	0.01996	0.585	atteasy	0.001009	0.05872	0.03966	0.10392	2.46	1 0.45	aug. 0.75	ADDS -A.D	NU8 012020	3 2.50
9,900	148,845	208.701	81,737	0.00499	0.00003	0.0025	0.00275	0.00049	0.02929	0.298	1.00883	0.01144	0.05826	0.0205	0.01328	25.119	1 0.35	1275 0.72	59652 6.2	7149 0.29610	2 2.35
0,940	155.454	117.041	80.055	0.0045	0.00007	110100	8145.75	1104.76	9.05758	0.565	0.00709	0.00013	0.09961	0.02.407	0.005/7	25.90	3.0.45	perm dias	ana -0.8	Anna in Frank	5 2.243
10,000	182.568	298.340	77.63	0.00502	0.00003	0.0628	0.00033	0.00041	0.01791	0.298	0.00795	0.03331	0.01799	0.9256	out tr	25.676	1.0.42	1/05 0.71	(3797 - 0.4	3533 0.77986	3 1.331
# part	Genera /	1 1 1 1 1							Allina						CONTRACT.	a.03		028.0.0			
																			(D)2	ALF CONTRACT OF	
	0	0	-					100					-								

Fig 2: represent data set

To start with, let us consider a dataset.

Consider a fictional dataset that describes the features for prediction of Parkinson's disease.

The dataset is divided into two parts, namely, feature matrix and the response vector.

- Feature matrix contains all the vectors(rows) of dataset in which each vector consists of the value of dependent features.
- Response vector contains the value of class variable(prediction or output) for each row of feature matrix.

## 2. Support Vector Machine (svm)

I guess by now you would've accustomed yourself with linearregression and logistic regression algorithms. If not, I suggest you have a look at them before moving on to support vector machine. Support vector machine is another simple algorithm that every machine learning expert should have in his/her arsenal. Support vector machine is highly preferred by many as it produces significant accuracy with less computation power. Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks. But, it is widely used in classification objectives.

The objective of the support vector machine algorithm is to find a hyper plane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.



Fig 3 : Possible hyper planes



# Fig 4 : Hyper planes in 2D and 3D feature space



## Fig 5 : represents support vector

#### **3.** Gradient Descent technique

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.



Fig 6 :shows the SGD technique



## Step-by-step

Now let's run gradient descent using our new cost function. There are two parameters in our cost function we can control: mm (weight) and bb (bias). Since we need to consider the impact each one has on the final prediction, we need to use partial derivatives. We calculate the partial derivatives of the cost function with respect to each parameter and store the results in a gradient.

#### 4. Gradient boosting tree

Although most of the Kaggle competition winners use stack/ensemble of various models, one particular model that is part of most of the ensembles is some variant of Gradient Boosting (GBM) algorithm. Take for an example the winner of latest Kaggle competition: Michael solution Jahrer's with representation learning in Safe Driver Prediction. His solution was a blend of 6 models. 1 LightGBM (a variant of GBM) and 5 Neural Nets. Although his

success is attributed to the semi-supervised learning that he used for the structured data, but gradient boosting model has done the useful part too.Even though GBM is being used widely, many practitioners still treat it as complex black-box algorithm and just run the models using pre-built libraries. The purpose of this post is to simplify a supposedly complex algorithm and to help the reader to understand the algorithm intuitively. I am going to explain the pure vanilla version of the gradient boosting algorithm and will share links for its different variants at the end. I have taken base DecisionTree code from **fast.ai** library (fastai/courses/ml1/lesson3rf foundations.ipynb) and on top of that, I have built my own simple version of basic gradient boosting model.



Fig 7 : represents the iteration ,parallel, sequential



Fig8: Samplerandomnormallydistributed residuals with mean around 0





## **IV. PROPOSED SYSTEM**

We have proposed new method to collect symptoms of diseases we are using data set of Parkinson disease by using some machine learning algorithms for more accuracy In prediction we used algorithms like

- a. NAVIE BAYES
- b. SVM
- c. SGD
- d. GBT

While compared with previous research's we have got better results those are shown below





## V. ADVANTAGES

- We have taken voice data set of Parkinson disease
- We have taken 195 patients samples 23 feature from voice data set
- ['MDVP:Fo(Hz)', 'MDVP:Fhi(Hz)', 'MDVP:Flo(Hz)', 'MDVP:Jitter(%)', 'MDVP:Jitter(Abs)', 'MDVP:RAP', 'MDVP:PPQ', 'Jitter:DDP', 'MDVP:Shimmer', 'MDVP:Shimmer(dB)', 'Shimmer:APQ3', 'Shimmer:APQ5', 'MDVP:APQ', 'Shimmer:DDA', 'NHR', 'RPDE', 'DFA', 'spread1', 'spread2', 'D2', 'PPE']
- From these features we are comparing the patients data and predicting the disease
- We are divided the prediction into 3 sub parts like 50,100,150
- In the above bar graph we can see the results
- By dividing the samples into 3 parts we can know the which algorithm is giving better output and accurate results
- In our observation (GBT) is giving best accurate results
- Time complexity is very less
- Reliability
- Portability
- By using this algorithm's we can increase life span of the patient if it is detected in initial stages

# SRS ( SOFTWARE REQUIREMENT SPECIFICATION )

software requirements specification (SRS) is a detailed description of a software system to be developed with its functional and non-functional requirements. The SRS is developed based the agreement between customer and contractors. It may include the use cases of how user is going to interact with software system. The software requirement specification document consistent of all necessary requirements required for project development. To develop the software system we should have clear understanding of Software system. To achieve this we need to continuous communication with customers to gather all requirements.



## Fig 11: shows the software requirement specification

#### SPIRAL MODEL

Spiral model is one of the most important Software Development Life Cycle models, which provides support for Risk Handling.

In its diagrammatic representation, it looks like a spiral with many loops.

The exact number of loops of the spiral is unknown and can vary from project to project.

Each loop of the spiral is called a Phase of the software development process. The exact number of phases needed to develop the product can be varied by the project manager depending upon the project risks. As the project manager dynamically determines the number of phases, so the project manager has an important role to develop a product using spiral model.

The Radius of the spiral at any point represents the expenses(cost) of the project so far, and the angular dimension represents the progress made so far in the current phase.

Below diagram shows the different phases of the Spiral Model:



Fig 12: Shows the working of spiral model

## **Advantages of Spiral Model**

- Below are some of the advantages of the Spiral Model.
- **Risk Handling**: The projects with many unknown risks that occur as the development proceeds, in that case, Spiral Model is the best development model to follow due to the risk analysis and risk handling at every phase.
- Good for large projects: It is recommended to use the Spiral Model in large and complex projects.
- Flexibility in Requirements: Change requests in the Requirements at later phase can be incorporated accurately by using this model.

• **Customer Satisfaction**: Customer can see the development of the product at the early phase of the software development and thus, they habituated with the system by using it before completion of the total product.

## VI. FUCTIONAL REQUIREMENTS

- A Functional Requirement (FR) is a description of the service that the software must offer.
- Functional Requirements are also called Functional Specification.
- Capturing the patient medical data as a Input .
- Handling the large-scale datasets.
- observational and multi-center study that includes early untreated Parkinson's Disease patients along with age and gender-matched healthy normal subjects, to identify progression biomarkers in Parkinson's Disease.
- Output will be the different machine learning comparative analysis graphs

## NON-FUNCTIONAL REQUIREMENTS

- A non-functional requirement defines the quality attribute of a software system.
- Non-functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs.
- login for authorized users.
- data backup using SQL queries.
- It should run in offline also

## DATA FLOW DIAGRAM



Fig 13 : data flow diagram

## ARCHITECTURAL DIAGRAM



Fig 14: It shows the architecture diagram

## SOFT WARETOOLS

Tools that are used for implementation of the problem solution are as follows:

python 3.7.4( 64 bit)

Microsoft Excel 2007.

Parkinson's data set.

Machine learning algorithms.

Html web page for user convenience

## SCREEN SHOTS OUTPUT

Re & Se & De Deg Opter Hann Hep Franz 3.1. (appel 3.4.4005H19. appel 1.1 2015, 10:2020 [NE v.1014 6 Mt 10054]) o win12 [We have, 'uppel's, 'waining or timeses} in terms information.
<pre>Pyber 31.4 (appk31.4.468598112, Al 1 202, 20240) SEX v.1856 64 int (BR668] on vibit Type "Bar", Younging", Younging", Younging Young Young</pre>
No.         HORDER: Clypticson/perimer.py           Horder data and scenarizing         Horder data and scenarizing           That make a dynamics 16         Horder data and scenarizing           Number of primers without Perimers 0         Horder data and scenarizing           With Perimer of Perimers without Perimers 0         Horder data and scenarizing           With Perimer of Perimers without Perimers 0         Horder data and scenarizing           With Perimer of Perimers without Perimers 0         Horder data and scenarizing           With Perimer of Perimers without Perimers 0         Horder data and scenarizing           With Perimer of Perimers without Perimers 0         Horder data and scenarizing           With Perimer of Perimers without Perimers 0         Horder data and scenarizing           With Perimers 0         Horder data and scenarizing         Horder data and scenarizing           With Perimers 10         Horder data and scenarizing         Horder data and scenarizing           With Perimers 2010         Horder data and scenarizing         Horder data and scenarizing           With Perimers 2010         Horder data and scenarizing         Horder data and scenarizing           With Perimers 2010         Horder data and scenarizing         Horder data and scenarizing           With Perimers 2010         Horder data and scenarizing         Horder data and scenanizing
BUBLE Classifier in an anti-anti-anti-anti-anti-anti-anti-anti-
State data set sourcefully         In           Intel state of primes: 105         Matter of primes: 106           Matter of primes: 106         Matter of primes: 106           Matter of primes: 106         Matter of primes: 106           Matter of primes: 106         Matter of primes: 106           Ventra of primes: 106         Matter of primes: 106           Ventra of primes: 106         Ventra of primes: 106           Ventra of primes: 106         Ventra of prime: 106           Ventra of prime: 106 </td
han lambe of painters 18 Marker of painters with Archiners 10 Marker of painters of the Archiner of Marker o
hanke of actions 21 hanke of actions with Actionses 10 hanke of actions with Actionses 10 hanke of actions with Actionses 10 hanke of actions without Actionses 10 hanke without
hante of printers with Antinance 10 Matter of printers without Antinance 0 Hanter of printers without Antinance 0 Hanter of printers without Antinance 0 Hanter of printers without Antinance 0 Northeader, North-Share, Nort
The solute of the solution of
[WEF-Shinty, WEF-Shinty, WEF-Shinty, W., Seet-Shinty, WEF-Shinty, W
S. YURANEY, Manaserika, War, War, War, Yura, Yuran, W. YHYI Happe colum: Status Marine wilker Miller wilker Miller wilker 1000, 11,500 (10,500), and 10,500 (10,500) 1000 (10,500) (10,500), and 10,500 (10,500) 1000 (10,500) (10,500) (10,500) (10,500) 1000 (10,500) (10,500) (10,500) (10,500) (10,500) 1000 (10,500) (10,500) (10,500) (10,500) (10,500) 1000 (10,500) (10,5
Partice volume:         Number volume:           VUET-Particle VUET-Pa
Names         Names <td< td=""></td<>
Bit Source Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11380         Different Statute         Bit Source Statute         Bit Source Statute           1 11390         Different Statute         Bit Source Statute         Bit Source Statute           1 11390         Ditttttt         Different Statute <t< td=""></t<>
1 12.40 14.60 14.13 14.50 14.885 1.4885 1.4885 1 2 14.60 14.11 14.15 14.50 14.885 1.4885 1.4885 1 3 14.60 14.11 14.15 14.50 14.50 14.885 1 4 14.14 14.70 14.54 14.55 14.585 1 5 000 x 11 claums] Suffling of their inter at raining sets completed training set: 55 sugles Main Age: Training set test of sugles Training Set test of su
III.0.1
III.G.W     III.S.W     III.S.W     III.S.W       III.G.W     III.S.W     III.S.W     III.S.W       III.G.W     III.S.W     III.S.W     III.S.W       Statilized of that ison tests and training sets complexed       Training sets: Stamples       Hold sets: G samples
Inflate 1 Inflate 1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.
(5 row a 11 columns) Sufficient of that is non-set and training sets complexed Training set: 55 supples Training set: 55 supples Training a subscription statistic set size of 50 Training a subscription size of subscription set size of 50 Training a subscription size of subscription set size of 50 Training a subscription size of subscription set size of 50 Training a subscription size of subscription set size of 50 Training a subscription size of subscription set size of 50 Training a subscription set size of 50 Training a set size of statistic set of 50 Training a set size of statistic set of 50
Shariling of stat and next and transports completed Text set: 6 samples Text set: 6 samples Text set: 6 samples Text set status in 1.000 seconds Ref predictions in 1.000 seconds. Ref predictions in 1.000 seconds. Fination g at waits of status o
Transming set: Do Mangades Transming set: Do Mangades Transming a Descriptional studies of the set size of 50 Transming a Descriptional studies of the set size of 50 Theorem for transming set of the seconds. For words for transming set of the seconds. Fit sources for the second sec
Text Set D: Sample:         Training a Guardiad W string at training set size of 50         Training at the formation of the Seconds.         If source for training set 1.708.         The source for training set 1.708.
Manne Anger: Tranzie a Nanne Maria e Sanzanie a training est size of 50 Tranzie a Nanne Sanzanie. Rake predictions in 1.000 secondo. El nome for training estTAN. Main Predictions in 1.000. Secondo. El nome for test est5000. Quert Merche Machine: Training est est staining est size of 51
Transming = Searchaff Stating = Transming = Searchaff Stating = Se
Timene Yoon, in L. Mully Seconds Repredictions in 1.00% Seconds. 21 Source for training set: 1.708. Rear for training in L. ANNA. Support Peters Rechimes: Theraping Sec Stating set size of SL Training Sec Stating Seconds
Make predictions 14, 10,008, Seconds. Have for maining and 1,003. Have for particular to 1,1030. Seconds for tests set: 0,1000. Seconds Have Mark (and 1,000). Finding 8 are within particular of 51 Training 8 are of 51
11 stores for transmis set: 0.000. 21 stores for test set: 0.0000. 25 stores for test set: 0.0000. Support Forest Schinger Training a Set similary set: similary of Si
Mar prediction in 1.100, Secondo. Farma for start as 1.000. Support Petros Rechimes: Theraday as at in a constant of 51 Theraday as at in a constant of 51
ri zure za zer sec 1.000. Supur Recime: Tening a ST using a training set size of SL
Support Rectin Retines: Training a Pri using a training set size of SU Training and D. M. M. Seconda
Treannag a SVC using a treannag set size of 50 Treined model in 0.0000 seconds
Trained model in U.UUU seconds
and the second
Nate predictions in 0.000 seconds.
FI Sorre for training set: U.9288.
Made predictions in U.1000 seconds.
FI some for test set: 0.860.
Studiastic Gradient Descent:
Training a SEOClassifier using a training set size of 50
Treined model in U.1849 seconds
Wade predictions in U.1159 seconds.
Fl score for training set: 1.0656.
Wade predictions in 0.1000 seconds.
(2) and for high rate ( ) (200
📲 🛤 k = k = k = k = k = k = k = k = k = k

#### International Journal of Computer Science Trends and Technology (IJCST) - Volume 8 Issue 2, Mar-Apr 2020



## International Journal of Computer Science Trends and Technology (IJCST) - Volume 8 Issue 2, Mar-Apr 2020



Fig 16 : UPLOAD TRAINING SET

The first two options generate accuracy graphs for two different sets of data. The third option generates a scatter graph with relations between parameters.

Upload test file

Generate graph

Go to Intro Screen Test totalcases for Parkinson's

Get Scatter Graph

Diagnose Individual for Parkinson's

Upload training set



#### Fig 18: Generated graph

## Fig 19 : CLICK ON THE UPLOAD TEST FILE

The first two options	generate accuracy grap	hs for two different sets of data
The third option gener	rates a scatter graph wit	h relations between parameters
Upload training se	t Upload test	file Get Scatter Gra
	Generate grap	h.
	Generate grap	h.

Fig 20 : TEST FILE OUTPUT

## Fig 17 : TRAINING SET OUT PUT

Shuffling of data into test and training sets complete!Training af SamplesNaive Bayes:Training a GaussianNB using a training set size of 8. . .Support Vector Machines:Training a Stoclassifier using a training set size of 8. . .Stochastic Gradient Descent:Training a GaussianNB using a training set size of 8. . .Set soure details for 100 setMaive Bayes:Training a Stoclassifier using a training set size of 8. . .Support Vector Machines:Training a GaussianNB using a training set size of 8. . .Set soure details for 100 setMaive Bayes:Training a Stoclassifier using a training set size of 8. . .Stochastic Gradient Descent:Training a GaussianNB using a training set size of 8. . .Support Vector Machines:Training a GaussianNB using a training set size of 8. . .Support Vector Machines:Training a GaussianNB using a training set size of 8. . .Stochastic Gradient Descent:Training a GradientDoostingClassifier using a training set size of 8. . .Stochastic Gradient Descent:Training a GradientDoostingClassifier using a training set size of 8. . .Support Vector Machines:Training a Stoclassifier using a training set size of 8. . .Support Vector Machines:Training a GradientDoostingClassifier using a training set size of 8. . .Support Vector Machines:Training a GradientDoostingClassifier using a training set size of 8. . .Support Vector Machines:Training a GradientDoostingClassifier using a training set size of 8. . .<td



## Fig 21: Generated graph

## Fig 22 :CLICK ON GET SCATTER GRAPH

-	Go to intro screen	Test totalcases for Parkinson's
The first has entions	onorato accuracy granhe	for hup different note of data
The third option genera	tes a scatter graph with re	elations between parameters.
Upload training set	Upload test file	e Get Scatter Graph
	Generate graph.	



Fig 23 : Generated scatter graph

## **VII. FUTURE SCOPE**

In this study we have used machine learning techniques, however very few researches have been done on machine learning algorithms. In future, the work can be extended by using auto encoders to reduce the number of feature and to extract the most important from them. Also the dataset used in this work is not so complex, so auto encoder did not learn well from that but with complex dataset it would definitely give better results.

## **VIII. CONCLUSION**

In this work, various prediction models for Parkinson's disease detection. For this purpose four machine learning techniques i.e. are naviebaye's, support vector machine(SVM),StochasticGradient

Descent(SGD),Gradient tree boosting(GBT). To obtain the desired results, are as well as four performance metrics are seen. These four metrics are accuracy, sensitivity, ROC, specificity.

From the results, GBT outstands from all the other ML techniques with the accuracy. After that , we tried to selected the most important and minimum number of features from the speech articulation data of 195 people where we have 23 features as explained in dataset description .For that we have used feature selection techniques whose working is shown below by changing the number of features selected as it is giving the overall accuracy 96.6%, which is better in comparison to all other machine learning techniques when compared with 50,100 and 150 sub sets feature's performance metrics.

## REFERENCES

- [1] Kamal Nayan Reddy, Challa, Venkata Sasank Pagolu and Ganapati Panda, "An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques", in Proceedings of the International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016, pp. 1446-145, 2016.
- [2] Geeta Yadav, Yugal Kumar and G. Sahoo,

"Predication of Parkinson's disease using Data Mining Methods: a comparative analysis of tree, statistical and support vector machine classifiers", in Proceedings of the National Conference on Computing and Communication Systems (NCCCS), pp. 1-4, 2012.

- [3] Paolo Bonato, Delsey M. Sherrill, David G. Standaert, Sara S. Salles and Metin Akay, "Data Mining Techniques to Detect Motor Fluctuations in Parkinson's Disease", in Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4766-4769, 2004.
- [4] Sonu S. R., Vivek Prakash and Ravi Ranjan, "Prediction of Parkinson's Disease using Data Mining", in Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1082-1085, 2017.
- [5] Aarushi Agarwal, Spriha Chandrayan and Sitanshu S Sahu, "Prediction of Parkinson's Disease using Speech Signal with Extreme Learning Machine", in Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 1-4, 2016.
- [6] Akshaya Dinesh and Jennifer He, "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recording", in Proceedings of the IEEE MIT Undergraduate Research Technology Conference (URTC), pp. 1-4, 2017.
- [7] Giulia Fiscon, Emanuel Weitschek, Giovanni Felici and Paola Bertolazzi, "Alzheimer's disease patients classification through EEG signals processing", in Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM). pp 1-4, 2014.
- [8] Pedro Miguel Rodrigues, Diamantino Freitas and Joao Paulo Teixeirab, "Alzheimer electroencephalogram temporal events detection by K-means", in Proceedings of the International Conference on Health and Social Care Information Systems and Technologies HCIST. pp. 859 – 864, 2012.

[9] Daniel Johnstone1, Elizabeth A. Milward1, Regina Berretta1 and Pablo Moscato1, "Multivariate Protein Signatures of Pre-Clinical Alzheimer's Disease in the Alzheimer's Disease Neuroimaging Initiative (ADNI) Plasma Proteome Dataset", in Proceedings of the Disease Neuroimaging Initiative, vol-7, pp. 1-17, 2017.

## **AUTHOR DETAILS**

P. SRAVANTHI is presently pursuing B.Tech (CSE) Department of Computer Science Engineering from Nadimpalli satyanarayana raju institute of technology.
B.SAI SUDHA is presently pursuing B.Tech(CSE) Department of Computer Science Engineering from Nadimpalli satyanarayana raju institute of technology
M.DINESH is presently pursuing B.Tech (CSE) Department of Computer Science Engineering from Nadimpalli satyanarayana raju institute of technology
P.SAI GANESH is presently pursuing B.Tech (CSE) Department of Computer Science Engineering from Nadimpalli satyanarayana raju institute of technology
MR.T.RAVI KUMAR (M.TECH),is working as an Assistant Professor in Department of computer science and engineering in Nadimpalli satyanarayana raju institute of technology, sontyam, Visakhapatnam.



MR.G.RAJASEKHARAM (M.TECH,PH.D), is working as an Associate Professor,(HOD) in the Department of Computer Science and Engineering in Nadimpalli satyanarayana raju institute of technology, sontyam,Visakhapatnam.