PARKINSON’S DISEASE SEVERITY PREDICTION

Submitted By:

Appidi, Harinadh

Kasturi, Aditya Karthikeya

1 | Introduction:

Parkinson Disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease. Population prevalence of PD increases from about 1% at age 60 to 4% by age 80. Parkinson's disease (PD) is a type of movement and non-movement disorder. It happens when nerve cells in the brain don't produce enough of a brain chemical called dopamine. There is no specific test for PD, so it can be difficult to diagnose. Doctors use a medical history and a neurological examination to diagnose it. Early symptoms of PD include tremor, rigidity, and difficulty walking; cognitive decline is common at later stages. It is more common in men than in women. Vocal impairment is also common with studies reporting 70% - 90% prevalence after the disease onset. Early and accurate diagnosis of PD is critical for effective treatment, but unfortunately PD diagnosis is not efficient. Tracking Parkinson's disease (PD) symptom progression often uses the unified Parkinson's disease rating scale (UPDRS) that requires the patient's presence in clinic, and time-consuming physical examinations by trained medical staff. UPDRS consists of 3 sections. 1) Mentation, Behavior and Mood; 2) Activities of Daily Living; 3) Motor. Motor UPDRS score ranges from 0-108 and Total UPDRS score range from 0-176.

Telemonitoring of Parkinson Disease is non-invasive and potentially reliable, cost-effective screening of persons with PD while removing the burden of frequent visits to clinic for both the patients and health care provider. Telemonitoring of PD heavily depends on recording of speech signals and can be easily integrated into telemedicine applications. In this project, a genuine effort is made to predict Total UPDRS and Motor UPDRS scores from speech signals using different statistical techniques and training machine learning models. Root Mean Squared Error, Absolute Mean Error and R-squared metrics are used to measure model ability in mapping speech signals to UPDRS values. *Random Forest Regressor / Extra Trees Regressor* model were able to perform better than other models after applying robust feature selection algorithms. A table depicting the same is shown below.

Model Results using all features				
	Mean Absolute Error	**Root Mean Square Error**	**R2 Score**	**Explained Variance**
Total_UPDRS	0.837271916	1.657768674	0.976573	0.9742
Motor_UPDRS	0.7006393	1.399442429	0.970812709	0.970984557

Model Results using Feature Selection				
	Mean Absolute Error	**Root Mean Square Error**	**R2 Score**	**Explained Variance**
Total_UPDRS	0.7124	1.5354	0.9799	0.9799
Motor_UPDRS	0.5802	1.1737	0.9794	0.9795

*Fig***.** *Model Metrics on Original Features and features selected using Recursive Feature Elimination*

2 | Related work:

Authors of *Non-invasive telemonitoring of Parkinson's disease, Tsanas et al.*[1] selected the most parsimonious model with a feature selection algorithm using Lasso, and statistically map the selected subset of features to UPDRS using linear and nonlinear regression techniques, which include classical least squares and non-parametric classification and regression trees (CART). Mean Absolute Error metric is used for evaluation of model and authors were able to achieve best MAE of 5.8 for Motor UPDRS and 7.5 for Total UPDRS scores on test data with CART algorithm.

However, Feature selection using Nonlinear methods with more performant feature selection algorithms like Recursive Feature Elimination and Principal Component Analysis can be used to better the original results.

We outperformed the original results by a large margin with Motor UPDRS MAE of 0.58, and Total UPDRS of 0.71 on test data using Extra Trees Regressor combined with RFE algorithm.

3 | Methods:

The problem is to map the features to UPDRS scores using regression techniques. The steps followed to build a ML model is:

1. Data Exploration
2. Data Cleaning
3. Train-Test Split (70-30 split)
4. Feature Normalization
5. Feature Selection
    a. Principal Component Analysis
    b. Feature Selection using univariate regression techniques
    c. Feature Selection using Recursive Feature Elimination
6. Model training
    a. With original features
    b. With Principal components
    c. With features from F-Test & Mutual Information
    d. With features from RFE
7. Model Prediction on test set and Selection
8. Hyperparameter Tuning on best model using cross validation
9. Comparison of current model with best model and iterating to step 1

Assumptions:

1. All recordings for each subject are of weekly interval.
2. Weekly UPDRS are linearly interpolated from baseline, 3-month and 6-month UPDRS scores from physical visit to clinic.
3. Importance to predictive ability of model than to interpretability.

4| Experimental Setup and Data Description:

In this study, 42 persons with Parkinson disease (PWP) were recruited for 6-month trial using At home testing device (AHTD). The AHTD contains a docking station for measuring tremor, paddles and pegboards for assessing upper body dexterity, a high-quality microphone headset for recording patient voice signals and a USB data stick to store test data. An LCD displays instructions for taking the tests. Typical audible prompts instruct the patient to undertake tasks to measure tremor, bradykinesia, complex coordinated motor function, speech and voice. As part of a trial to test the effectiveness of the AHTD system in practice, PWP were recruited and trained to use the device. Subsequently, an AHTD was installed in their home and they performed tests on a weekly basis.

The main aim of this study is to analyze vocal features and representing their characteristics and map these features to UPDRS using regression methods. Main assumption in this study is that vocal performance deterioration is solely due to Parkinson Disease and not any other pathology. All subjects remained un-medicated for the six-month duration of the study.

There are total of 5875 observations with each patient having approximately 200 recordings each.

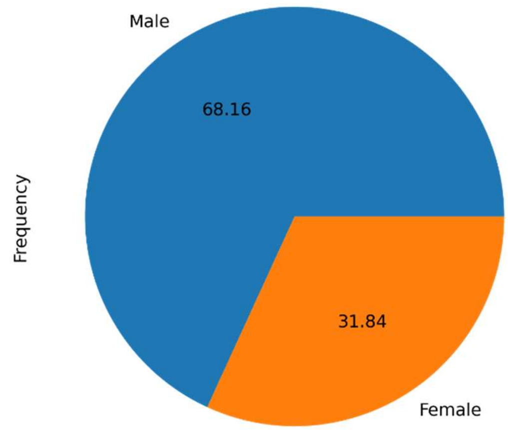

Fig.: Frequency of participants ratio

The pie chart above describes the participants ratio. Male: 68.16% and Female: 31.84%.

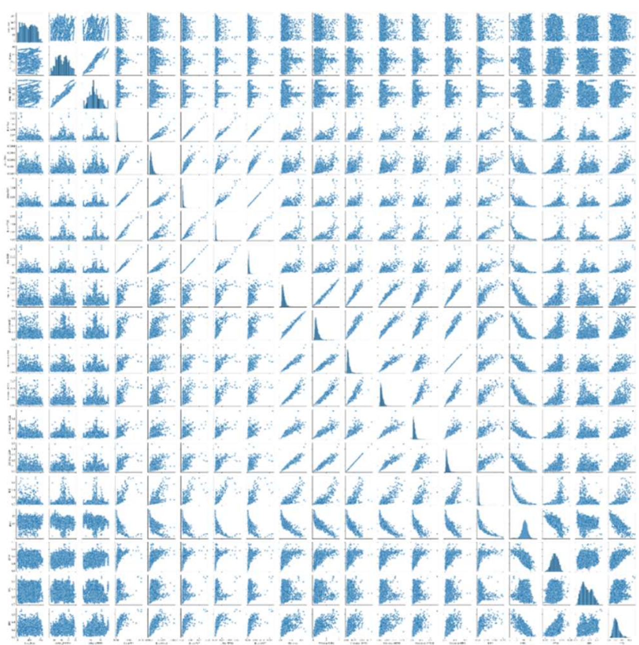

Fig.: Correlation between each pair using Pair plots

The above pair plot is alternative way to a correlation heat map, showing relationship between each pair of attributes and helps in downstream modeling process.

The correlation between two variables is a number that represents how closely values for those two variables tend to be aligned. A correlation of 1 means that when one variable increases, the other variable also increases. A correlation of -1 means that when one variable increases, the other decreases linearly.

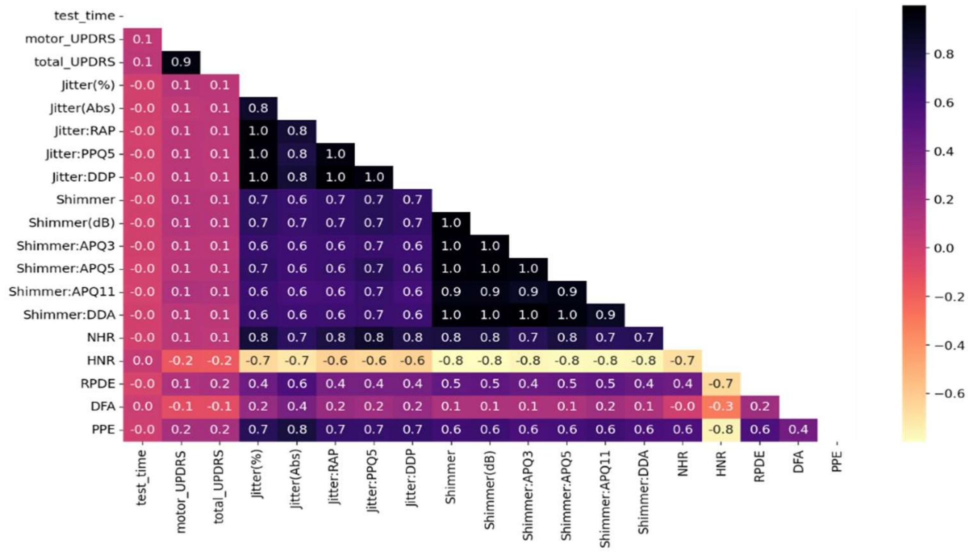

Fig.: Correlation Heatmap of all the features

A correlation heatmap shows the correlation between features of a data set. The plot is a matrix, where cells are correlation values between features. The color of cells represents the strength of correlation, with black indicating a strong positive correlation, yellow indicating a strong negative correlation, and purple indicating no correlation.

Shimmer variables are correlated with each other, and Jitter is also the same. Other variables are not highly corelated except HNR which is negatively correlated with other features.

Dataset statistics

Number of variables	22
Number of observations	5875
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1009.9 KiB
Average record size in memory	176.0 B

Variable types

Numeric	21
Categorical	1

From the above figure, we can see that total number of variables are 22 where 2 variables of them are Total UPDRS and Motor UPDRS. 1 categorical variable is sex and 16 variables are voice measures and other variables are subject# and time of recording from the date of admission.

4.2 Features Definition:

Vocal Feature	Description
Jitter(%)	Average absolute difference between consecutive periods, divided by the average period.
Jitter(Abs)	Average absolute difference between consecutive periods which gives information about the cycle-to-cycle variation of fundamental frequency given in seconds.
Jitter:RAP	Relative Average Perturbation (RAP), which is the average absolute difference between a period and the average of it and its two neighbours, divided by the average period.
Jitter:PPQ5	Five-point Period Perturbation Quotient, computed as the average absolute difference between a period and the average of it and its four closest neighbours, divided by the average period.
Jitter:DDP	Average absolute difference between consecutive differences between consecutive periods, divided by the average period.
Shimmer	Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.
Shimmer(dB)	Average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20. It gives information about the variability of the peak-to-peak amplitude in decibels.
Shimmer:APQ3	Three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.
Shimmer:APQ5	Five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours, divided by the average amplitude.
Shimmer:APQ11	11-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbours, divided by the average amplitude.
Shimmer:DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods.
Noise to Harmonics Ratio (NHR)	Amplitude of noise relative to tonal components. It quantifies the noise which occurs due to turbulent airflow, resulting from incomplete vocal fold closure in speech pathologies.
Harmonics to Noise Ratio	Amplitude of tonal relative to noise components. It has the same aim as NHR.
Recurrence period density entropy	Addresses the ability of the vocal folds to sustain stable vocal fold vibrations, quantifying the deviations from exact periodicity
Detrended fluctuation analysis	Quantifies the self-similarity of the noise present in the speech caused by the turbulent air flow
Pitch period entropy	Measures the impaired control of stable pitch during sustained phonations

https://doi.org/10.1371/journal.pone.0182428.t001

4.3 Feature Transformation:

We removed subject# from the feature set and applied normalization to bring the features to standard normal scale.

4.3.1 Before Feature Transformation:

	age	test_time	Jitter(%)	Jitter(Abs)	Jitter:RAP	Jitter:PPQ5	Jitter:DDP	Shimmer	Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	Shimmer:APQ11	Shimmer:DDA	NHR	HNR	RPDE	DFA	PPE
count	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104
mean	64.78411	92.85115	0.006154	4.41E-05	0.002988	0.003279	0.008965	0.034034	0.310937	0.017151	0.020103	0.027434	0.051451	0.031913	21.66957	0.542175	0.654144	0.220496
std	8.822715	53.07276	0.005642	3.53E-05	0.003136	0.003744	0.009409	0.025776	0.229764	0.013228	0.016463	0.019978	0.039684	0.057765	4.301168	0.101161	0.070919	0.092671
min	36	0.39653	0.00084	2.25E-06	0.00033	0.00045	0.00098	0.00306	0.026	0.00161	0.00194	0.00249	0.00484	0.000286	1.659	0.20929	0.51404	0.021983
25%	58	47.36275	0.00357	2.23E-05	0.00157	0.00182	0.00472	0.01913	0.175	0.009268	0.01078	0.015648	0.027793	0.010912	19.349	0.471225	0.596928	0.154823
50%	65	91.397	0.00493	3.49E-05	0.00226	0.00249	0.00678	0.02743	0.253	0.0137	0.015895	0.022525	0.04111	0.018577	21.9185	0.542695	0.644965	0.20618
75%	72	137.83	0.00681	5.4E-05	0.00331	0.00349	0.00992	0.03987	0.36625	0.02061	0.02381	0.032933	0.061833	0.031797	24.5065	0.615053	0.71229	0.26677
max	85	215.49	0.09999	0.000396	0.05754	0.06956	0.17263	0.23915	1.97	0.16267	0.16702	0.27546	0.48802	0.74826	37.875	0.96608	0.8656	0.73173

4.3.2 After Feature normalization:

	age	sex	test_time	Jitter(%)	Jitter(Abs)	Jitter:RAP	Jitter:PPQ5	Jitter:DDP	Shimmer	Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	Shimmer:APQ11	Shimmer:DDA	NHR	HNR	RPDE	DFA	PPE
count	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104	4104
mean	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
std	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
min	-3.26	-0.69	-1.74	-0.94	-1.19	-0.85	-0.76	-0.85	-1.2	-1.24	-1.17	-1.1	-1.25	-1.17	-0.55	-4.65	-3.29	-1.98	-2.14
25%	-0.77	-0.69	-0.86	-0.46	-0.62	-0.45	-0.39	-0.45	-0.58	-0.59	-0.6	-0.57	-0.59	-0.6	-0.36	-0.54	-0.7	-0.81	-0.71
50%	0.02	-0.69	-0.03	-0.22	-0.26	-0.23	-0.21	-0.23	-0.26	-0.25	-0.26	-0.26	-0.25	-0.26	-0.23	0.06	0.01	-0.13	-0.15
75%	0.82	1.46	0.85	0.12	0.28	0.1	0.06	0.1	0.23	0.24	0.26	0.23	0.28	0.26	0	0.66	0.72	0.82	0.5
max	2.29	1.46	2.31	16.63	9.96	17.4	17.71	17.4	7.96	7.22	11	8.93	12.42	11	12.4	3.77	4.19	2.98	5.52

5 | Results:

5.1 Feature Selection:

5.1.1 F-Test:

Feature selection using F-test is a univariate linear regression test returning F-statistic and p-values. Features are ranked according to their F-statistic. This test measures linear dependence of each feature w.r.t UPDRS independently to other features.

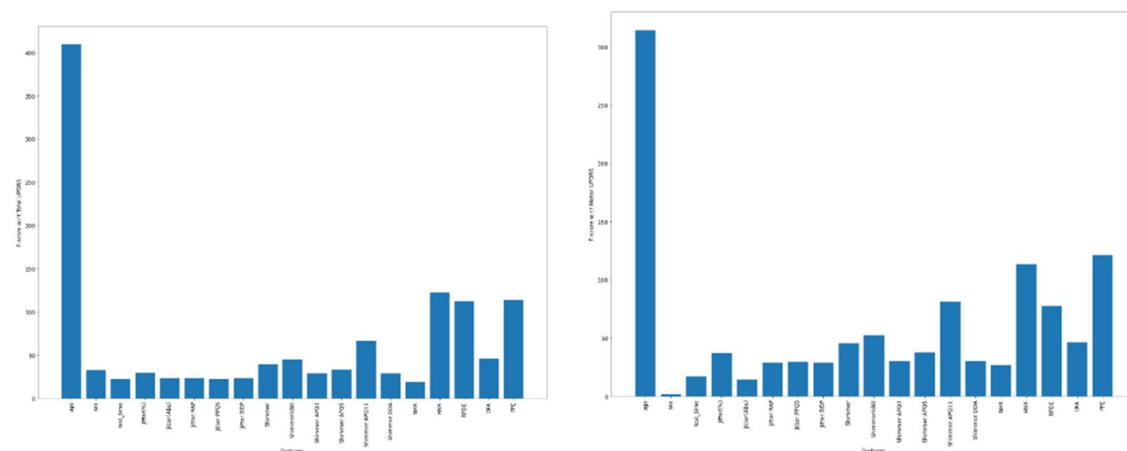

Fig. F-score of features w.r.t Total and Motor UPDRS

5.1.2 Mutual Information Feature Selection:

The k best characteristics are those that result in the most mutual information. This metric is based on information theory and evaluates the reduction of uncertainty for a random variable while considering the value of another random variable that is known. In contrast to F-test, this test measures dependency of individual feature w.r.t response variable and captures complex non-linear relationship between individual feature and response variable.

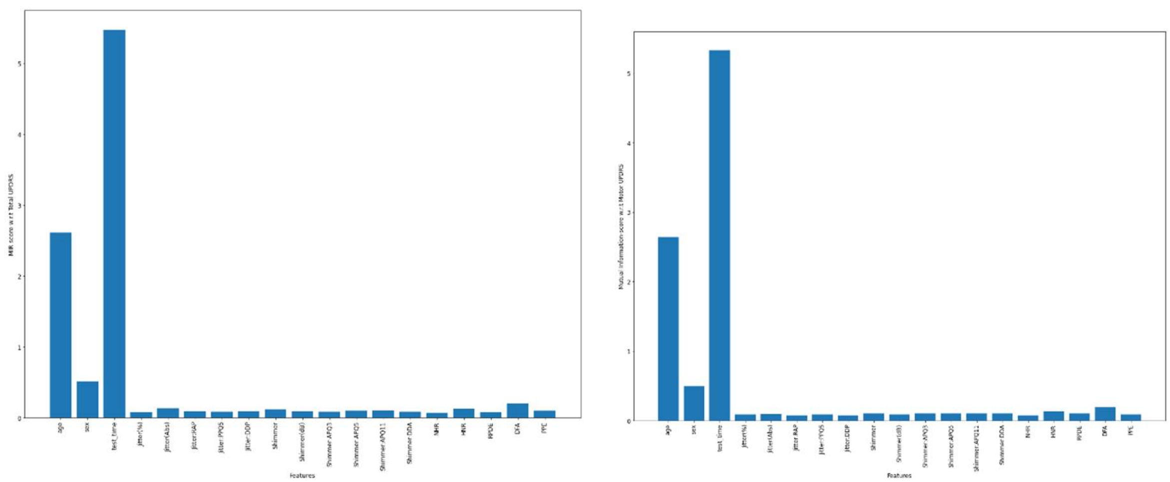

Fig. Mutual Information Feature Selection w.r.t Total and Motor UPRDS

5.1.3 Wrapper Method:

We can also use a machine learning classifier to test it repeatedly on subsets of our data set until we find which one makes the best prediction. This is a more expensive computing method, but it is more accurate than the previous ones. Recursive Feature elimination method iteratively eliminates features by ranking the features using a given ML algorithm to compute feature importance and removes least ranked feature until the given number of features remain.

5.1.3.1 RFE

We used the Recursive Feature Elimination (RFE) method to select the best features with 50% features selected i.e., 9 features.

Selected Features are:

a. age
b. sex
c. test_time

d. Jitter(Abs)
e. Shimmer:APQ5
f. HNR
g. RPDE
h. DFA
i. PPE

5.1.4. Principal Component Analysis:

PCA is linear dimensionality reduction technique using singular value decomposition (SVD) of data and projects it to lower dimensional space and tries to capture the variance in original data using minimal number of variables.

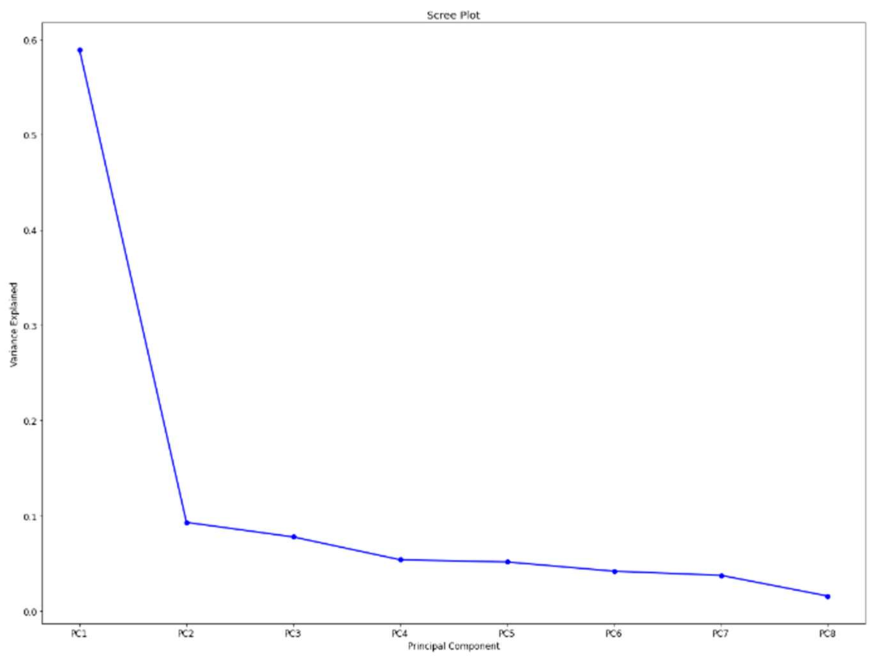

Fig. PCA vs Explained Variance

5.2. Model Results:

In this section model results of only the best performing models are plotted in all features case and subset of features using RFE as well.

5.2.1. Models Performance on Test set with Original Features

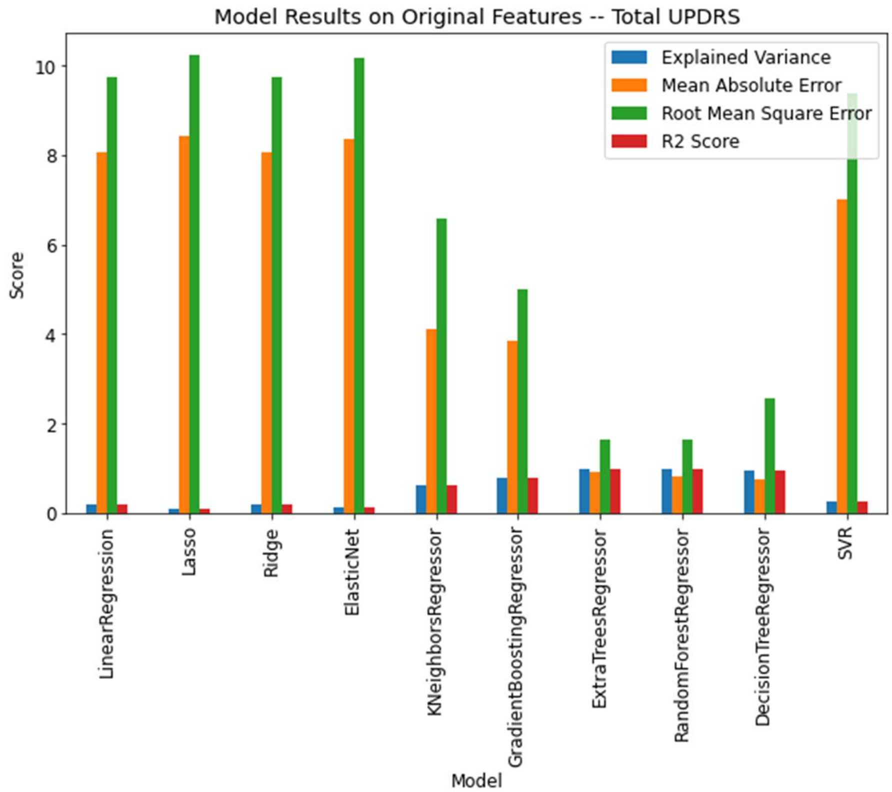

Fig.: Bar Plot for Total UPRDS

	Model	Explained Variance	Mean Absolute Error	Root Mean Square Error	R2 Score
0	LinearRegression	0.19016135	8.06510841	9.74824058	0.18994179
1	Lasso	0.10844032	8.42084643	10.22900683	0.10807017
2	Ridge	0.18990689	8.06717730	9.74976674	0.18968812
3	ElasticNet	0.11856623	8.35368565	10.17067089	0.11821449
4	KNeighborsRegressor	0.63207807	4.12656986	6.57108807	0.63192330
5	GradientBoostingRegressor	0.78826309	3.83581591	4.98727883	0.78797306
6	ExtraTreesRegressor	0.97714982	0.92918526	1.64058090	0.97705653
7	RandomForestRegressor	0.97694565	0.82296356	1.64668408	0.97688551
8	DecisionTreeRegressor	0.94369971	0.73845611	2.57515208	0.94347123
9	SVR	0.27498341	7.00340449	9.38410501	0.24932929

Fig. Metrics for Total UPRDS

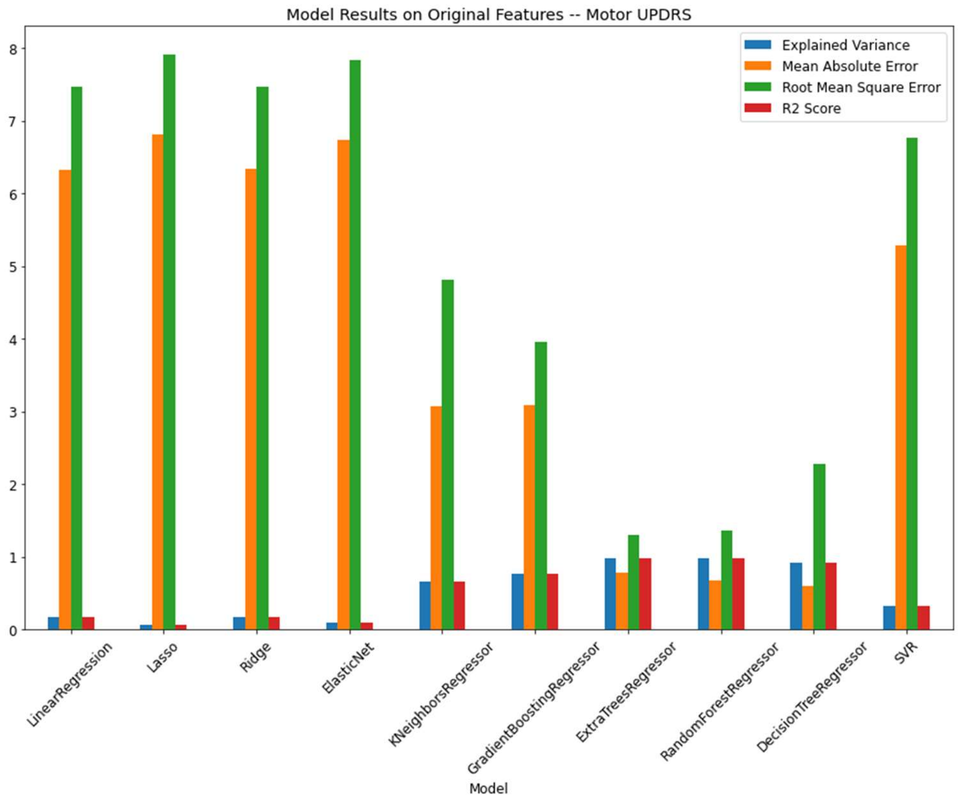

Fig. Bar plot for Motor UPRDS

	Model	Explained Variance	Mean Absolute Error	Root Mean Square Error	R2 Score
0	LinearRegression	0.17013387	6.33011067	7.46362259	0.16979936
1	Lasso	0.06883951	6.81799067	7.90643752	0.06836583
2	Ridge	0.16996417	6.33131443	7.46438961	0.16962872
3	ElasticNet	0.08670607	6.73398760	7.83018146	0.08625003
4	KNeighborsRegressor	0.65624721	3.07071752	4.80621335	0.65573743
5	GradientBoostingRegressor	0.76804441	3.09202083	3.95013492	0.76745469
6	ExtraTreesRegressor	0.97526482	0.77211137	1.29119160	0.97515351
7	RandomForestRegressor	0.97257093	0.67902775	1.36057814	0.97241133
8	DecisionTreeRegressor	0.92316883	0.59795196	2.27063243	0.92316176
9	SVR	0.32736160	5.28617773	6.76749246	0.31744244

Fig. Metrics for Motor UPRDS

5.2.2. Model Results using Selected Features from RFE

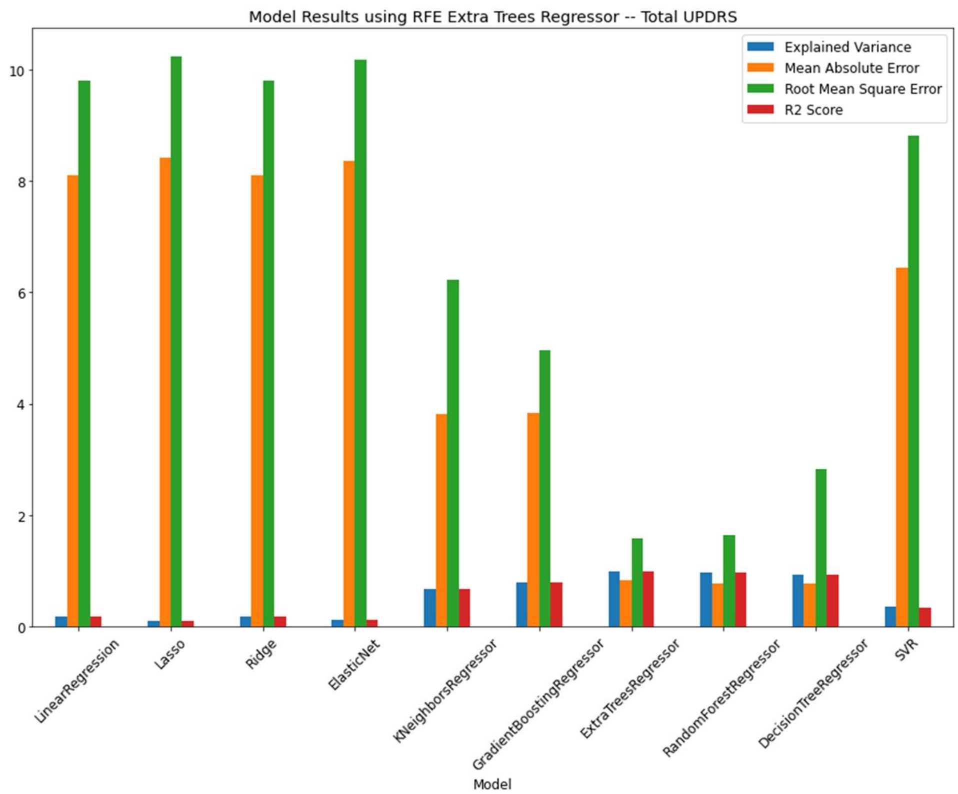

Fig. bar plot for Total UPRDS

	Model	Explained Variance	Mean Absolute Error	Root Mean Square Error	R2 Score
0	LinearRegression	0.15605461	6.39416644	7.52712136	0.15561296
1	Lasso	0.06883951	6.81799067	7.90643752	0.06836583
2	Ridge	0.15604211	6.39427794	7.52717662	0.15560057
3	ElasticNet	0.08663825	6.73555819	7.83048529	0.08617912
4	KNeighborsRegressor	0.68683893	2.83453989	4.58696659	0.68642973
5	GradientBoostingRegressor	0.75511425	3.17859641	4.05815957	0.75456191
6	ExtraTreesRegressor	0.9790304	0.58063808	1.1885482	0.97894684
7	RandomForestRegressor	0.97531192	0.60017734	1.2896214	0.9752139
8	DecisionTreeRegressor	0.9348591	0.50590443	2.09418046	0.93464002
9	SVR	0.43159047	4.71601046	6.20752333	0.4257243

Fig. Metrics for Total UPRDS

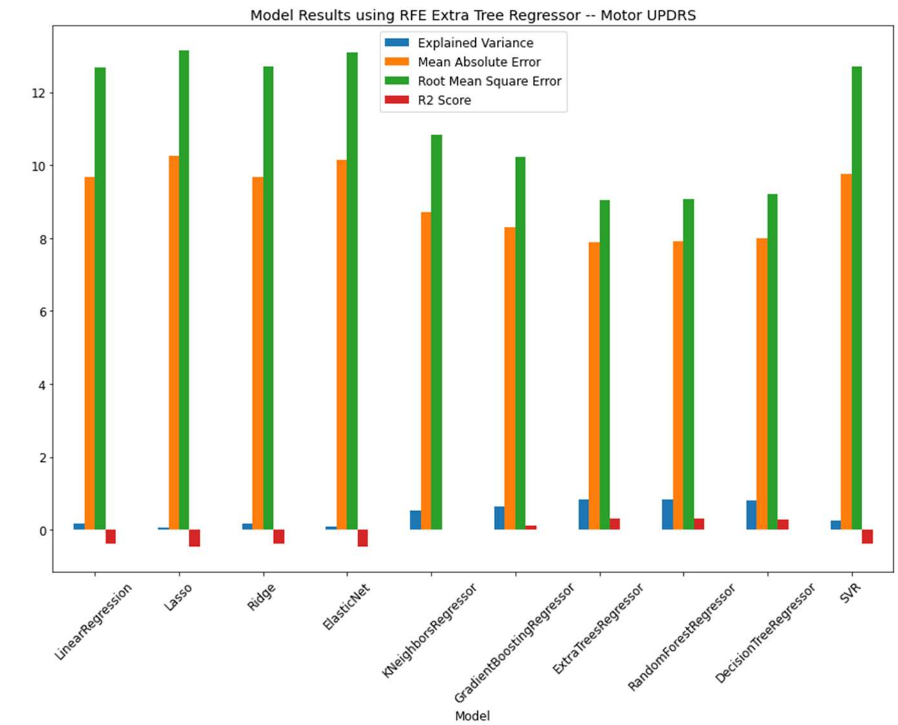

Fig. Bar Plot for Motor UPRDS

	Model	Explained Variance	Mean Absolute Error	Root Mean Square Error	R2 Score
0	LinearRegression	0.15605461	6.39416644	7.52712136	0.15561296
1	Lasso	0.06883951	6.81799067	7.90643752	0.06836583
2	Ridge	0.15604211	6.39427794	7.52717662	0.15560057
3	ElasticNet	0.08663825	6.73555819	7.83048529	0.08617912
4	KNeighborsRegressor	0.68683893	2.83453989	4.58696659	0.68642973
5	GradientBoostingRegressor	0.75511425	3.17859641	4.05815957	0.75456191
6	ExtraTreesRegressor	0.9790304	0.58063808	1.1885482	0.97894684
7	RandomForestRegressor	0.97531192	0.60017734	1.2896214	0.9752139
8	DecisionTreeRegressor	0.9348591	0.50590443	2.09418046	0.93464002
9	SVR	0.43159047	4.71601046	6.20752333	0.4257243

Fig. Metrics for Motor UPRDS

From the above graphs and results, we conclude that the ExtraTreesRegressor and RandomForest Regressor outperforms all the other models. The table below shows RandomForest Regressor for All features and Extra Trees Regressor for selected features using RFE.

Model Results using All Features				
	Mean Absolute Error	Root Mean Square Error	R2 Score	Explained Variance
Total_UPDRS	0.837271916	1.657768674	0.976573273	0.976648157
Motor_UPDRS	0.7006393	1.399442429	0.970812709	0.970984557

Model Results using Feature Selection Algorithm				
	Mean Absolute Error	Root Mean Square Error	R2 Score	Explained Variance
Total_UPDRS	0.743245848	1.613987611	0.977794316	0.977849405
Motor_UPDRS	0.604390726	1.279626524	0.975596608	0.975724054

6 | Discussion:

The results support the hypothesis that the vocal features can be used to predict the UPDRS scores. The results reveal interesting insights about existing medical practices. The results show that the vocal features can be used to predict the UPDRS scores. Detecting early stage of Parkinson's Disease would highly impact for early diagnosis which leads to address early medications for improving the quality of the patient's life.

The limitation was that the dataset which we have used in this project was conducted on 42 patients, if there was more data collected from various kinds of patients, who were exposed to multiple diseases and also Parkinson's Disease would give different attributes and observations. Though the UPDRS scores are linearly interpolated with the assumption of progression of disease being linear over time is not a robust assumption and may be more dynamic in practice.

7 | References:

1. https://www.nature.com/articles/npre.2009.3920.1.pdf
2. https://www.cdc.gov/genomics/hugenet/casestudy/parkinson/parkcoffee_view.htm
3. https://pubmed.ncbi.nlm.nih.gov/19932995/

8 | Relevant Papers:

1. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng. 2010 Apr;57(4):884-93. doi: 10.1109/TBME.2009.2036000. Epub 2009 Nov 20. PMID: 19932995

2. Little MA, McSharry PE, Hunter EJ, Ramig LO (2009), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease',IEEE Transactions on Biomedical Engineering, 56(4):1015-1022

3. Accurate telemonitoring of Parkinson.s disease progression by non-invasive speech tests', IEEE Transactions on Biomedical Engineering (to appear).

9 | Contributions

9.1 Contributions by Appidi, Harinadh

- Initial Draft
- Project Research Initiative
- Literature Review
- Analysis on Current Project Dataset
- Exploratory Data Analysis
- Documentation Draft 1
- Documentation Draft 2
- Documentation Review 1
- Documentation Review 3
- Documentation Final Review 1
- Data Modeling on Machine Learning Models
- Analysis on the Machine Learning Models
- Creating the Presentation
- Presentation Review 2
- Presentation Final 1

9.2 Contributions by Kasturi, Aditya Karthikeya

- Research on Related Work
- Collecting data from related projects
- Analysis on Related work Datasets
- Project Workflow understanding
- Crafting Project Guidelines
- Documentation Draft 3
- Documentation Draft 4
- Documentation Review 2
- Documentation Final Review 2
- Data Modeling on Feature Selection
- Data Visualization
- Editing the Presentation
- Presentation Review 1
- Presentation Final 2

APPENDIX: