JOURNAL OF
EMERGING INVESTIGATORS

Article

Using advanced machine learning and voice analysis features for Parkinson's disease progression prediction

Deetya Narsipur[1], Pranav Narsipur[1], Keshav Narsipur[1]
[1] Redmond High School, Redmond, Washington

SUMMARY

Parkinson's disease (PD) is a neurodegenerative disease that is important to diagnose early for appropriate treatment. Decline in voice quality is an early symptom of PD, and prior studies have used machine learning models to analyze audio clips and monitor for progression of PD. We hypothesized that the voice features Harmonics to Noise Ratio (HNR), Detrended Fluctuation Analysis (DFA), and Jitter Absolute (JA) would be most useful in detecting PD progression represented by a Unified Parkinson's Disease Rating Scale (UPDRS) and that the relationship between these features and UPDRS scores would be nonlinear given the complexity of PD. We used the publicly available PD telemonitoring dataset from the UCI (University of California Irvine) Machine Learning Repository to validate our hypothesis. After controlling for age, we identified DFA, HNR and JA as the best factors to predict motor and total UPDRS scores, with accuracy measured by MAE and MSE. DFA was the most effective to use in advanced machine learning to accurately predict high motor and total UPDRS scores. We also identified random forest as the best possible model to predict PD progression. We concluded that random forest performed the best as it was a non-linear model, emphasizing the importance of non-linear relationships found in voice features of PD patients, given that PD is a complex disease impacting multiple neural systems with varying degrees of progression. Our study has potential to help clinicians identify progression of PD and manage diagnosis in a non-invasive manner while also providing insights into diagnosis.

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disease affecting millions of patients worldwide, where neurons that produce dopamine die or become impaired due to unknown causes (1). PD is characterized by motor symptoms such as tremor, rigidity, and slowed movement as well as non-motor symptoms including cognitive and speech difficulties (1). As PD progresses, symptoms worsen and impact quality of life (2). As PD is a progressive disease, it is important to diagnose it in a timely manner. However, it requires consultations with medical specialists, understanding of the patient's medical history, clinical exams, and invasive tests (3). After the assessment and tests are done by a medical specialist, the procedure for PD is to assign scores using the Unified Parkinson's Disease Rating Scale (UPDRS) to represent the severity and progression of PD (4). Total UPDRS values range from 0 to 199, where 0 represents no disability or symptoms of PD and 199 indicates the most advanced stage of PD with significant impairment (4). The UPDRS consists of four main parts: I) thinking, behavior, and mood; II) activities of daily living; III) motor examination; and IV) complications of therapy (4). Speech is assessed in both Part II as part of daily living activities and Part III as part of the motor examination (4). The motor UPDRS score (Part III), which measures motor symptoms including speech, facial expression, tremor, rigidity, slowness of movement, gait, and balance problems, ranges from 0 (no motor symptoms) to 108 (severe motor symptoms) (4).

Given that an accurate diagnosis of PD requires the patient to have physical and financial access to medical specialists and clinical tests, there has been growing interest in alternative methods of predicting PD progression, with voice analysis showing one possible approach (5). PD is diagnosed by doctors based on medical history, a consideration of symptoms and a physical and neurological exam, which may not be accessible to all (1). Speech impairment such as changes in voice quality, articulation, and the stress and intonation in language have been observed in PD patients as some of the first symptoms to develop (6). For patients with PD, a complex dysfunctional interaction of respiratory, laryngeal, and articulatory functions results in these changes (7). Thus, research has begun to focus on tracking the voice attributes objectively, remotely, and in a non-invasive manner to monitor PD progression (8).

Different voice features have been evaluated for their relationship with PD progression. One such voice feature is Detrended Fluctuation Analysis (DFA), a measure of self-similarity of voice over time, which can uncover changes in speech timing and reveal patterns in speech rhythm. A study

https://doi.org/10.59720/24-284

used DFA to classify PD in patients and found that the DFA values were generally higher for PD patients than for controls due to impaired muscle control (9). Another voice feature is Harmonics to Noise Ratio (HNR) is a ratio which represents the voice quality in audio data, with higher values indicating clearer voices and lower values indicating breathiness or hoarseness. A study characterizing voice quality in early PD using HNR reveals that higher noise from breathiness is due to incomplete vocal fold closure (10). This can be used to track PD through telemonitoring and for classification of PD (11,12). Jitter Absolute (JA) measures frequency instability in voice data and can be used to differentiate healthy voices from dysphonic voices due to vocal tremor and reduced muscle control, classifying PD and remote monitoring (10-12). Additionally, Recurrence Period Density Entropy (RPDE) is used to detect the presence of aperiodic, irregular patterns in the voice (11).

These three voice features have been evaluated and shown potential for predicting PD progression individually. We sought to determine if these three voice features could be used in combination to predict PD progression. Further, the relationship between these voice features and clinical measures for PD progression like UPDRS is poorly understood. Models used in prior research studies have been mostly linear in nature; however, the nature of PD raises the question if the relationship may be non-linear (13).

In this study, we evaluated the potential of HNA, DFA and JA, to predict PD progression. We hypothesized that the voice features HNR, DFA, and JA are the most significant voice-related features to detect PD progression represented by a UPDRS, and the relationship between these features and UPDRS scores will be nonlinear given the complexity of PD. To evaluate this hypothesis, we utilized the publicly available PD Telemonitoring dataset from the University of California Irvine (UCI) Machine Learning Repository (5). We employed a mix of linear and non-linear modelling approaches and to evaluate the performance of our models, we used the coefficient of determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). We also conducted time series analysis to evaluate how the predictive power of these features changes over the course of PD progression.

We found that voice features DFA, HNR, and JA can all help track PD progression, with DFA being a feature that has the strongest correlation with total and motor UPDRS scores with the exception of age. This research may limit the number of invasive and expensive tests like dopamine transporter scan and magnetic resonance imaging patients with PD face when monitoring progression. This research contributes to the growing collection of research on using voice features for PD progression which could result in better patient outcomes.

RESULTS

We used advanced machine learning to predict progression in PD from speech data. We started with a dataset from UCI and data was split into training (80%) and testing (20%) sets using a random seed of 42 for reproducibility. After that, we performed feature engineering. Feature engineering was performed to create interaction terms between age and voice features, as well as polynomial features (up to degree 2) for DFA, HNR, and JA. These engineered features were added to the original feature set to capture potential non-linear relationships between features and UPDRS scores. Subsequently, feature selection was conducted using RFE, RFECV, and random forest importance to identify the most predictive features. We then performed feature selection for features using recursive feature elimination with cross-validation (RFECV) to select the most important features in the UCI dataset. We then trained the model to predict motor UPDRS values. Finally, using the trained model, we assessed which features were most important for predicting motor UPDRS scores and the overall accuracy of the model and calculated the accuracy in measuring motor UPDRS scores.

Correlation analysis

DFA had a weak negative correlation with motor UPDRS ($r = -0.116$, $p < 0.001$) and total UPDRS ($r = -0.113$, $p < 0.001$). HNR also had a weak negative correlation with motor UPDRS ($r = -0.157$, $p < 0.001$) and total UPDRS ($r = -0.162$, $p < 0.001$). JA had a weak positive correlation with both motor UPDRS ($r = 0.051$, $p < 0.001$) and total UPDRS ($r = 0.067$, $p < 0.001$). Contrary to our hypothesis, PPE (Pitch Period Entropy), which measures variability or unpredictability in the pitch period of speech, had a positive correlation ($r = 0.162$, $p < 0.001$ with motor UPDRS) that was stronger than anticipated. Its correlation strength was similar to that of HNR ($r = -0.157$) but in the opposite direction. Based on our background knowledge, we had expected PPE to show a weaker correlation than the three primary voice features (DFA, HNR, and JA) we had identified in our hypothesis. Previous studies had emphasized the diagnostic and monitoring potential of DFA, HNR, and JA, with less attention given to PPE as a primary biomarker for PD progression (9-12). This could indicate that larger PPE could be correlated to greater UPDRS scores. RPDE also had a positive correlation motor UPDRS ($r = 0.129$, $p < 0.001$) and total UPDRS ($r = 0.157$, $p < 0.001$). This shows RPDE's association with the progression of PD regarding correlation analysis.

Importance of features in predicting motor UPDRS

By analyzing feature importance using random forest, age was found as the best predictor for motor UPDRS importance (0.674). DFA was the next most important (0.101) feature followed by JA (0.054) and HNR (0.023), which were less important than expected. Recurrence Period Density Entropy (RPDE) had the same as HNR (0.023) (**Figure 1**).

Linear regression models

We created multiple linear regression models using all the voice features with age as predictors. The R^2 value for motor UPDRS was 0.112 and 0.138 for total UPDRS in this model. We compared the performance of models using all features versus models using only selected features from our RFE and RFECV analyses. The model with all features achieved slightly higher R^2 values than models with reduced feature sets, despite potential multicollinearity concerns. These low R^2 values indicate linear regression models don't accurately represent the variance in UPDRS scores. All predictors were standardized during preprocessing using sklearn's StandardScaler, which transformed each feature to have zero mean and unit variance before fitting the regression models. Age had the greatest positive coefficient at 1.768 for motor UPDRS, making it the most important predictor. PPE and RPDE also had positive coefficients, which relates to their prior positive correlations. DFA, HNR, and JA all had

https://doi.org/10.59720/24-284

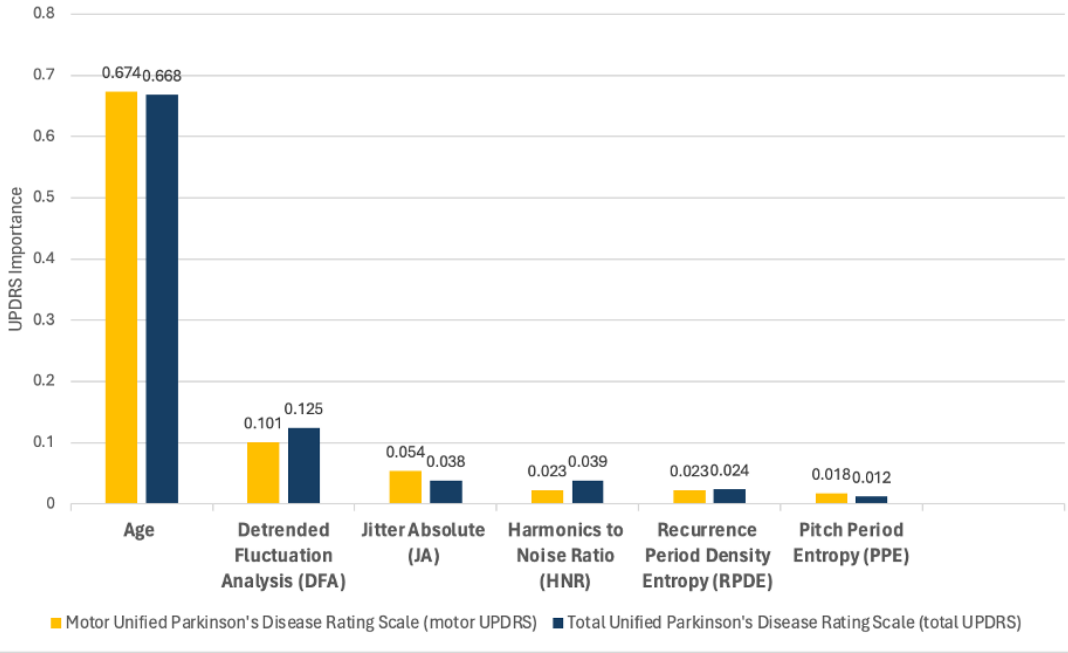

Figure 1: Voice features importance in predicting motor and total UPDRS scores. Importance of various voice features in a Random Forest model in predicting UPDRS scores, a key indicator of PD progression. Features included in the model include age, Detrended Fluctuation Analysis (DFA), Jitter Absolute (JA), Harmonics to Noise Ratio (HNR), Recurrence Period Density Entropy (RPDE), and PPE (Pitch Period Entropy). Actual values are reported above each bar. Importance is evaluated using random forest regressor with 100 trees.

negative coefficients. However, PPE had unexpectedly large coefficients (motor UPDRS: 1.666; total UPDRS: 1.632), showing its potential importance (**Table 1**).

Non-linear models/random forest

Random forest models significantly outperformed linear regression models, achieving an R^2 of 0.892 for motor UPDRS and R^2 of 0.902 for total UPDRS compared to 0.112 and 0.138 respectively for linear regression. Random forest had an R^2 of 0.892 for motor UPDRS and R^2 of 0.902 for total UPDRS (**Table 1**). Using random forest, an MAE of 1.879 for motor UPDRS and 2.343 for total UPDRS was reported. The better performance of random forest models suggests that the relationship between voice features and PD progression is non-linear and more complex than can be captured by simple linear models.

MAE and MSE

Random forest had an R^2 of 0.892 for motor UPDRS and R^2 of 0.902 for total UPDRS (**Table 1**). Using random forest, we achieved an MAE of 1.879 for motor UPDRS and 2.343 for total UPDRS. These error metrics further support the superior performance of random forest compared to linear regression, which had significantly higher error values (MAE of 6.493 and 8.275 for motor and total UPDRS, respectively). Lower MAE values indicate that the random forest predictions were closer to the actual UPDRS scores.

Time series analysis

The UCI dataset collected voice data for the patients over a period of 5.3 to 5.8 months. The frequency of collection was approximately weekly, with additional measurements in between for many patients. Analyzing the predictive powers in various voice features over time displayed differing patterns (**Figure 2**). The dataset was divided into ten equally spaced time intervals spanning the 5.3-5.8 month study period. At each time interval, we built separate single-feature linear regression models for each voice feature (age, DFA, HNR, JA, RPDE, and PPE), with each model using only one feature to predict either motor or total UPDRS scores. For each model at each time point, we calculated the R^2 value to see how well that single feature predicted UPDRS scores at that particular stage of the disease progression. We then plotted these R^2 values across the 10 time points to visualize how each feature's predictive power changed throughout the observation period (**Figure 2**).

Our analysis showed that age had the greatest predictive power, and its R^2 increased slightly 1.34 times from 0.064 to 0.086. DFA shows the most substantial increase with its R^2 increasing 32.5 times from 0.0004 to 0.0130. HNR initially had a large predictive power when starting with R^2 being 0.071 at time 0 and decreased immediately after the initial assessment 2.84 times from 0.071 to 0.025. JA had low predictive power that further decreased 2.75 times from 0.011 to 0.0040. PPE also had a predictive power that decreased 2.92 times from 0.070 to 0.025. RPDE had a predictive power that decreased 1.75 times from 0.035 to 0.02. Changes in predictive power over time can show the importance of different features during different stages of PD or signify useful features for short- term or longitudinal studies. Further, no specific timepoints stood out which showed an inflexion on disease progression. This also highlights the need to do more longitudinal and adaptive modelling research.

Variance

To demonstrate the effectiveness of three voice features in the random forest model, we compared model performance across different sets of voice features: all features, a subset

https://doi.org/10.59720/24-284

	Motor UPDRS linear regression	Motor UPDRS random forest	Total UPDRS linear regression	Total UPDRS random forest
MAE	6.493	2.043	8.275	2.594
MSE	58.332	9.132	97.5886	13.662
R^2	0.086	0.857	0.119	0.877

Table 1: Linear regression and random forest models predictions for motor and total UPDRS scores. Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2).

comprising of age and three selected voice features (which we termed "original features"), and age alone. We compare the predictive power for different features between motor and total UPDRS scores R^2 values (**Table 2**). Using all features explained the highest variance (R^2 = 0.892 for motor, 0.902 for total). Age alone had a lower predictive power (R^2 = 0.631 for motor, 0.703 for total). The features DFA, HNR, JA helped improve predictions greater than age alone but are slightly less powerful than using all features (R^2 = 0.857 for motor, 0.877 for total).

In summary, our analysis of the voice features as predictors of PD progressions resulted in key observations: DFA, HNR, JA, RPDE and PPE showed promise as important voice features in predicting PD progression. The features had weak but statistically significant correlation with UPDRS scores. Specifically, DFA demonstrated weak negative correlations with both motor UPDRS and total UPDRS. HNR similarly showed weak negative correlations with motor UPDRS and total UPDRS. In contrast, JA exhibited weak positive correlations with both measures. RPDE and PPE also showed weak positive correlations with both motor and total UPDRS. Importantly the combination of the three voice features, provided significant additional predictive power to PD progression beyond age. Finally, the superior performance of non-linear models particularly random forest suggests complex, non-linear relationships, between voice features and PD progression.

	Motor UDPRS	Total UDPRS
R^2 all features	0.892	0.902
R^2 age only	0.631	0.703
R^2 original features	0.857	0.877
Variance all vs age	0.261	0.199
Variance age only vs original features	0.226	0.174

Table 2: Comparison of R^2 values and variance explained by different feature sets for motor and total UPDRS predictions. R^2 values for models Linear Regression and Random Forest using all features including age, age only, and original features DFA, JA, and HNR. It also presents the difference in R^2 values between models (labeled as 'variance'): 'All vs age' shows the additional variance explained when using all features compared to using age alone, while 'age only vs original features' shows the improvement gained by adding the three original voice features (DFA, HNR, JA) to the age-only model. Results are shown separately for motor UPDRS and total UPDRS to allow for comparison between these two.

DISCUSSION

Overall, our results partially support the hypothesis that voice features can help predict PD progression, measured by UPDRS scores. We found that voice characteristics are complex and non- linear in disease progression, with certain features, such as age, being identified as more important during feature selection.

Among the voice features, DFA shows the second strongest correlation with both motor and total UPDRS scores, following age. This finding is in line with our hypothesis. The negative correlation between DFA and UPDRS scores indicates that as the disease progresses, the self-similarity of the voice decreases, reflecting increased voice instability. Contrary to our hypothesis, HNR and JA showed weaker correlation than we expected due to the weakness of the correlation and small direction. This suggests that their predictive

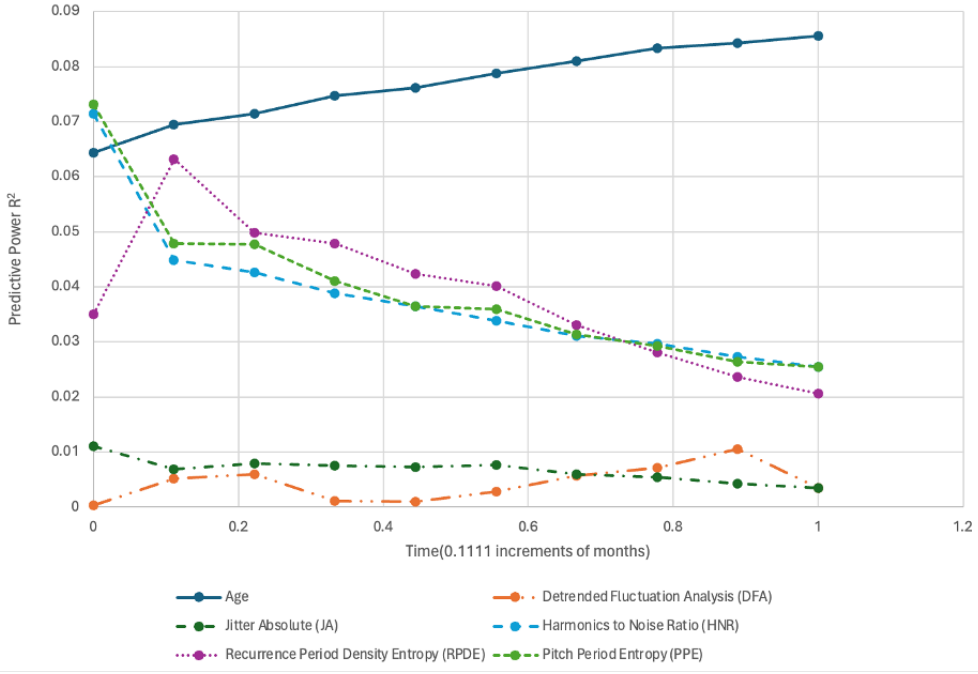

Figure 2: Predictive power of age, DFA, JA, HNR, RPDE, PPE over time. R^2 values from Random Forest for age (dark blue line), DFA (orange line), JA (dark green line), HNR (light blue line), RPDE (purple line), and PPE (light green line) measured at time increments of 0.1111.

https://doi.org/10.59720/24-284

capabilities might be more limited than the currently available research indicates. However, other voice features not initially included in our hypothesis, such as PPE and RPDE, show unexpected importance in our models. This suggests there are a bigger set of voice features are relevant for predicting PD progression. Thus, expanding the voice feature set from the three to five, including PPE and RPDE may be important in further research.

The stark contrast in performance between linear and nonlinear models shows the dynamics between voice features and UPDRS scores. Linear regression models only explained about 9 – 12% of variance in UPDRS scores, but non-linear random forest models explained about 86 – 88% of variance. This supports that the relationships between voice features and PD predictions are non-linear. The superior performance of random forest models suggests that the relationship between voice features and PD progression is likely non-linear. While random forest's better performance could potentially be attributed to other factors such as its ensemble nature, robustness to outliers, or handling of multicollinearity, the substantial performance gap supports our hypothesis that voice biomarkers relate to UPDRS scores in complex, non-linear ways. This aligns with the known complexity of PD as a progressive neurodegenerative disorder affecting multiple neural systems. This reflects prior research which shows the PD is a complex disease, impacting multiple neural systems with varying degrees of progression (14). This also reiterates the importance of using advanced non-linear modelling techniques to analyze voice data for PD progression.

Additionally, the time series analysis showed how the predictive power of varying voice features differed over the course of PD progression. PD progresses differently in everyone so the time period of 5.3-5.8 months would have been sufficient to see how in some PD worsens but in others it does not greatly progress (3). HNR's predictive power decreased over time while DFA's power slightly increased, suggesting that different voice features may be less or more informative in different stages of the disease. This finding could help longitudinal clinical studies, allowing them to better optimize when to use specific features in different stages.

However, our study also has its limitations. Random forest may still overfit on smaller datasets and it may place more importance on features that have many unique values. Our study uses a relatively small data set of data being 5875 observations across 42 people who have early-stage PD. This limited scope might regulate the applicability of our recommended approach. Future studies using a large data set and using longitudinal data over a longer duration would provide more robust insights into how voice features change time with individuals. Increased availability of public voice data sets for patients with PD would allow the training and validation of more robust models. Enriching the input data set with other noninvasive data like a patient's gait could also improve the model. There is also opportunity to use clinical data along with voice and gait to improve progression prediction accuracy. From a feature standpoint, we could incorporate more robust features, implement more cross validation, and apply ensemble methods that combine predictions from multiple models. There are opportunities to perform advanced signal processing like cepstral analysis, fundamental frequency, and format analysis to further improve predictions in disease progression.

In conclusion, we found that our methods of analyzing voice features (DFA, HNR and JA) were able to predict PD progression with significantly higher accuracy (R^2 = 0.877) compared to previous approaches or single feature models. Our results demonstrates that voice features such as DFA and HNR help to predict UPDRS scores, validating our hypothesis. DFA, HNR, and UPDRS scores have weak but significant correlations observed between them. We had previously indicated that DFA is higher in patients with PD when compared to patients without PD. This seemingly contradictory finding can be explained in two ways: I) while DFA might be higher in PD patients compared to healthy controls, it may decrease as the disease progresses (higher UPDRS scores); and II) the relationship between DFA and PD might not be linear throughout the disease course. Our research also reveals a non-linear relationship between voice features and UPDRS scores evidenced by the random forest model outperforming linear regression approach. As a follow up to this research, we used voice features and random forest model identified in this paper, applied Expectation Maximalization clustering, Principal Component Analysis and Support Vector Regression to predict motor UDPRS with an achieve a MAE of 0.47. This surpasses the performance of recent studies, further validating our findings (10). Overall, this work contributes to the growing body of evidence supporting the use of voice analysis in PD assessment.

MATERIALS AND METHODS

The PD dataset was downloaded from UCI (5). The data set contains a total of 5,875 observations from 42 patients with PD. Data was preprocessed by first checking for and handling missing values. Features were then standardized using sklearn's StandardScaler to ensure all variables had zero mean and unit variance, preventing features with larger scales from dominating the analysis. Subject-level data was also separated for time series analysis. Data was loaded preprocessed using Python (v3.8) with pandas (v1.2.4). Feature columns included age and 16 voice measures, with a focus on five features: JA, HNR, RPDE, DFA, and PPE. Additional voice measures were included in the model but showed less significance: Jitter(%), Jitter:RAP, Jitter:PPQ5, Jitter:DDP, Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA, and NHR. Features were standardized using sklearn's StandardScaler (v0.24.2).

Feature engineering was performed to create interaction terms between age and voice features, as well as polynomial features (up to degree 2) for DFA, HNR, and JA since PD progression often exhibits nonlinear patterns (13). It was integrated in our linear regression and random forest models which would make linear regression a linear model with non-linear features and random forest non-linear as it is a non-linear model by design. These engineered features were added to the original feature set. Feature selection was conducted using three methods: recursive feature elimination with cross-validation (RFECV), recursive feature elimination (RFE), and feature importance from random forest. The RFECV method determined that seven features (age, DFA, HNR, JA), RPDE, PPE, and Shimmer) provided the best predictive power. While we retained all features in our models for comparison purposes, our analysis and discussion focused primarily on these top predictors. This feature

https://doi.org/10.59720/24-284

selection process assisted in our comparison between the 'all features' model and the 'original features' model (age, DFA, HNR, JA), helping quantify the additional predictive value of the expanded feature set (**Table 2**). RFE and RFECV used linear regression as the estimator, while random forest regressor used 100 trees. The top features selected by each method were recorded and compared. RFE method ranked features by recursively removing the least important features based on their coefficients in the linear regression model. We configured it to select the top three features, which consistently identified age, DFA, and HNR as most important. Recursive feature elimination with cross-validation (RFECV) identified the optimal number of features by maximizing cross-validated performance. We used 5-fold cross-validation with negative mean squared error (neg_MSE) as the scoring metric. This method determined that seven features (age, DFA, HNR, JA, RPDE, PPE, and Shimmer) provided optimal predictive power, with performance plateauing or declining when additional features were included. Random Forest Feature Importance quantified each feature's contribution to reducing prediction error, calculated as the mean decrease in impurity (Gini importance) across all trees in the forest. Features were ranked based on their normalized importance scores from 0 to 1. We performed correlation analysis using pandas and seaborn (v0.11.1). A correlation matrix was generated to visualize relationships between three types of relationships: features and motor UPDRS, among different voice features, and between motor and total UPDRS. Random Forest Regressor from sklearn was used to analyze feature importance, with 100 decision trees in the forest.

Two regression models were built and compared: linear regression and Random Forest Regressor, both from sklearn. Data was split into training (80%) and testing (20%) sets. The analysis was performed twice, once with all features and once with only the original features (age, DFA, JA, and HNR). Model performance was evaluated using R^2, MAE, and MSE. These metrics were calculated using sklearn's metrics module.

To test the hypothesis that DFA, HNR, and JA are significant predictors beyond age, models were built using only these features plus age, and their performance was compared to models using all features. Specific tests were conducted to assess the relationships between these features and UPDRS scores. Variance explanation analysis was conducted by comparing R^2 scores of models with age only and models with all features.

A time series analysis was also done where for each subject a linear trend was calculated (slope, intercept, R^2) for all features and UPDRS scores over time. These time series were further divided into ten equally spaced normalized time points (ranging from 0 to 1, where 0 represents the beginning of data collection and 1 represents the end) for a predictive power change analysis. At each time point, single-feature linear regression models were built separately for each voice feature to predict UPDRS scores, and the resulting R^2 values were calculated. This showed how the R^2 of each feature changes over time.

Multiple regression analysis was conducted to test the combined effect of features on the target variables. All statistical analyses were performed using scipy (v1.6.2) and statsmodels (v0.12.2). A p-value < 0.05 was considered statistically significant for all tests. Data visualizations, including correlation heatmaps, feature importance plots, and time series trends, were created using matplotlib (v3.3.4) and seaborn (v0.11.1).

Received: August 11, 2024
Accepted: November 1, 2024
Published: August 6, 2025

REFERENCES

1. "Parkinson's Disease." *Mayo Clinic*. www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055. Accessed 5 Aug. 2024.
2. Martinez-Martin, Pablo, et al. "Health-Related Quality-of-Life Scales in Parkinson's Disease: Critique and Recommendations." *Movement Disorders*, vol. 26, no. 13, 2011, pp. 2371-2380. https://doi.org/10.1002/mds.23834.
3. "Diagnosis -Parkinson's Disease." *NHS Choices*. www.nhs.uk/conditions/parkinsons-disease/diagnosis/. Accessed 5 Aug. 2024.
4. Goetz, Christopher G., et al. "Movement Disorder-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results." *Movement Disorders*, vol. 23, no. 15, 2008, pp. 2129-2170. https://doi.org/10.1002/mds.22340.
5. Tsanas, Athanasios et al. "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests." *IEEE transactions on bio-medical engineering*, vol. 57, no. 4, 2010, pp. 884-93. https://doi.org/10.1109/TBME.2009.2036000.
6. Ho, Aileen K., et al. "Speech Impairment in a Large Sample of Patients with Parkinson's Disease." *Behavioural Neurology*, vol. 11, no. 3, 1999, pp. 131-137. https://doi.org/10.1155/1999/327643.
7. Duffy, Joseph R. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. 4th ed., Mosby, 2019.
8. Arora, Siddharth, et al. "Smartphone Motor Testing to Distinguish Idiopathic REM Sleep Behavior Disorder, Controls, and PD." *Neurology*, vol. 91, no. 16, 2018, pp. e1528-e1538. https://doi.org/10.1212/WNL.0000000000006366.
9. Viswanathan, Rekha, et al. "Complexity Measures of Voice Recordings as a Discriminative Tool for Parkinson's Disease." *Biosensors*, vol. 10, no. 1, 2020, p. 1. https://doi.org/10.3390/bios10010001.
10. Rusz, Jan et al. "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease." *The Journal of the Acoustical Society of America*, vol. 129, no. 1, 2011, pp. 350-67. https://doi.org/10.1121/1.3514381.
11. Little, Max A, et al. "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease." *IEEE transactions on bio-medical engineering*, vol. 56, no. 4, 2009, p. 1015. https://doi.org/10.1109/TBME.2008.2005954.
12. Tsanas, Athanasios, et al. "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease." *IEEE transactions on bio-medical engineering*, vol. 59, no. 5, 2012, pp. 1264-71. https://doi.org/10.1109/TBME.2012.2183367.
13. Tsanas, Athanasios, et al. "Nonlinear Speech Analysis

https://doi.org/10.59720/24-284

Algorithms Mapped to a Standard Metric Achieve Clinically Useful Quantification of Average Parkinson's Disease Symptom Severity." *Journal of the Royal Society Interface*, vol. 8, no. 59, 2011, pp. 842-855. https://doi.org/10.1098/rsif.2010.0456.

14. Poewe, Werner, et al. "Parkinson Disease." *Nature Reviews Disease Primers*, vol. 3, no. 1, 2017, p. 17013. https://doi.org/10.1038/nrdp.2017.13.