MPOWER VOICE ACTIVITY MONITORING AND CLASSIFICATION FOR PARKINSON'S DIAGNOSIS

QIUSHUI ZHANG

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

UNIVERSITY OF WASHINGTON

2018

COMMITTEE:

MARK WHIPPLE

GANG LUO

PROGRAM AUTHORIZED TO OFFER DEGREE:

BIOMEDICAL AND HEALTH INFORMATICS

University of Washington

Abstract

mPower Voice Activity Monitoring and Classification for Parkinson's Diagnosis

Qiushui Zhang

Chair of the Supervisory Committee:

Mark Whipple, Professor

Biomedical Informatics and Medical Education

Background: Parkinson's disease patients' voice data collected via the mPower application can be classified into three groups: Just after taking medication (at their best condition), Immediately before taking medication (at their worst condition), and somewhere in between medication doses (neither best nor worst condition).

Objectives: Our goal for this investigation is to validate voice as an accurate classifier of medication status in patients with Parkinson's Disease.

Methods: After data pre-processing, logistic regression, support vector machines (SVM), decision trees, Gaussian Naïve Bayes and Multi-layer Perceptron (MLP) is applied for model training.

Results: Comparing across the entire groups, the accuracy is relatively low at 0.51 on average for just best and worst condition and it increases to 0.55 for Gaussian Naïve Bayes if the condition between best and worst is also included. If we just consider the data for a single PD patient, the performance of the model can increase to 0.82.

Conclusions: The result shows that there is connection between voice and Parkinson's Disease conditions. However, the difference between the condition for the whole population might not be larger than the difference between each individual.

Introduction:

Parkinson's disease (PD) is a neurodegenerative disorder that affects mainly the motor system. It was first discovered in 1817 by James Parkinson, who described the pattern of motor signs he observed as "shaking palsy" (Parkinson, 2002). Later, more clinical manifestations for PD patients were found, including behavioral alterations, memory impairment, rest tremor, rigidity, and balance impairment (Savitt et al., 2006). Genetically, genome-wide significant association studies have identified twenty-six loci for the occurrence of PD (Nalls et al., 2014) (Chang et al., 2017). To characterize this disease pathologically, scientists have found the relationship between the nosologic classification of Parkinson's syndrome and the concentration of dopamine (DA) in the striatum (Bernheimer et al., 1973). Although scientists have discovered the annual decreasing uptake-rate of DA in the striatum is around 6%-13% for PD patients comparing with 0-2.5% for

control groups of the same age group (Parkinson Study Group, 2002), it can be difficult for clinicians to assess if the disease is progressing within a short period (Harel et al., 2004).

Since PD is a chronic disease, patients may not be aware of the severity of the symptom progression. Current clinical care for patients involves regular, but infrequent, visits with their specialist on the order of every 3-12 months (Guneysel et al., 2008). At these visits, based on the history and clinical exam, changes to medications and the treatment plan may be recommended (Cosentino et al., 2005). Patients present for their scheduled appointments typically regardless of their unique symptom course (APRN, 2017).

Hypokinetic dysarthria is known as the speech production problem of PD, which affect the intensity, pitch and duration of patients' voice (Schulz and Grant, 2000)(Canter, 1963). Scientist have already discovered the changes in fundamental frequency (F_0) for a single PD patient's voice five years prior to diagnosis (Harel et al., 2004). This suggests that voice variation may be one of the indicators for detecting the symptom severity of PD patients.

Our goal is to explore novel features, specifically in voice, to use in detection and management of PD signs and symptoms. These features can then be used to build up models for distinguishing the voice data of PD patients before they take the medicine from after they take the medicine. Based on these differences, we are able to manage and monitor treatment. The primary aim of this work was to validate use of voice features in detection of medication effect.

In the following sections of this paper our methods for data pre-processing, feature extraction, model selection, and model evaluation are discussed. The goal of this investigation is to demonstrate feasibility for using voice as a measureable marker for medication status in PD patients.

Methods:

The raw voice data was collected from synapse where both the recording of PD patients and the sound during the countdown are included. There are 5826 participants and 65,022 samples in total. The participants are divided into two groups: PD participants and non-PD participants. The medication timing for non-PD participants is “I don't take Parkinson medications”, while for PD patients, the situation for their medication timing is more complex. Based on whether these PD patients take medication, they are divided into three categories: “I don't take Parkinson medications” for those who does not take Parkinson medication, “nan” for those who have no idea about this question, and those who take Parkinson medication. For those who take Parkinson medication, they are further classified based on when their audio sample was submitted relative to their medication timing: “Immediately before Parkinson medication”, where patients are classified as the time at which their symptoms were at their worst; “Just after Parkinson medication (at your best)”, where patients are classified as the time at which their symptoms were at their best; “Somewhere in between medication doses”, where the patients submitted the audio neither at their best nor their worst.

To extract audio features from the raw audio data, an openly available utility was used. It extracted features directly from raw audio using the Python Audio Analysis Library, which was capable of extracting 34 audio features, shown in table 1 (Giannakopoulos, 2015). By applying a frame size of 50 milliseconds and a frame step of 25 milliseconds (50% overlap), the features for an audio signal could be extracted.

Feature ID	**Feature Name**	**Description**
1	Zero Crossing Rate	The rate of sign-changes of the signal during the duration of a particular frame.
2	Energy	The sum of squares of the signal values, normalized by the respective frame length.
3	Entropy of Energy	The entropy of sub-frames' normalized energies. It can be interpreted as a measure of abrupt changes.
4	Spectral Centroid	The center of gravity of the spectrum.
5	Spectral Spread	The second central moment of the spectrum.
6	Spectral Entropy	Entropy of the normalized spectral energies for a set of sub-frames.
7	Spectral Flux	The squared difference between the normalized magnitudes of the spectra of the two successive frames.
8	Spectral Rolloff	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.
9-21	MFCCs	Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.
22-33	Chroma Vector	A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing).
34	Chroma Deviation	The standard deviation of the 12 Chroma coefficients.

Table 1. Short-term feature extraction using audioFeatureExtraction.py.

After feature extraction, we first extracted the data of Parkinson's disease patients based on the medication timing: those who submitted audio "Immediately before Parkinson medication", and "Just after Parkinson medication (at your best)". The length of the audio vectors for each patient was slightly different due to the variation in phonation length. At a sampling rate of 40 per second, each vector contained 400 data points if the phonation was a full 10 seconds. However, for a minority of samples, phonation was much shorter leading to vectors with fewer than 400 data points, and these samples were discarded. For the remaining samples with at least 400 data points, everything beyond the first 400 data points were discarded.

To identify pertinent voice features from those extracted and enhance generalization by reducing overfitting, Support vector machine Recursive feature elimination (SVM-RFE) with cross-validation was applied for feature extraction (Formisano et al., 2008). It trained on the whole datasets and found the importance of each feature by applying criteria derived from the coefficients in SVM models (Yan and Zhang, 2015). By recursively removing features with the smallest importance and re-train, desired number of features were selected. This process was recursively repeated until the desired number of features were selected. For the label of our voice data was binary, stratified K-fold cross validation was selected, with 10 folds (Pedregosa et al., 2011).

Next, for each selected feature, the datasets for patients with medication timing of "Immediately before Parkinson medication", and "Just after Parkinson medication (at your best)" were merged

and feature scaling was applied. Machine learning algorithms did not perform well when the attributes of the input numeric had very different scales. The standardization package in scikit-learn was applied to this combined dataset, which subtracted the mean value and then divided by the variance, so the resulting distribution would have unit variance (Géron, 2017) (Pedregosa et al., 2011).

Once the most discriminant features were chosen and feature scaling was done, we applied the data for PD patients to build various machine learning models, evaluating these models based on medication time point labels by calculating the ratio of correct predictions. Initially, only the samples collected Immediately before and after medication administration are pooled together and used to provide a labelled training and testing set.

We first trained on six several models: logistic regression, support vector machine (SVM), decision trees, Gaussian Naïve Bayes and Multi-layer Perceptron (MLP). By applying stratified K-fold cross validation, with 10 folds, we could calculate the accuracy scores of these models by making predictions on each fold for the selected model by training the remaining folds (Pedregosa et al., 2011). We also drew the receiver operating characteristic (ROC) curve for these models, which ploted the true positive rate against the false positive rate. The area under the curve (AUC) using the trapezoidal rule was also calculated.

Next, we explored the effect of adding the "Somewhere in between medication doses point" group to our models. To get a brief idea about this medication time points, we applied the

matplotlib package to visualize the mean value of spectral flux line for these three datasets based on time points to get a brief idea about the “Somewhere in between medication doses point”.

We trained all the selected six models using all three groups (“Immediately before Parkinson medication”, “Just after Parkinson medication (at your best)”, and “Somewhere in between medication doses point”) after feature scaling and analyzing them in two different combinations. First we combined those samples submitted having taken medications at “Somewhere in between medication doses point” with those who are submitted “Immediately before Parkinson medication”. Then, we combine those samples submitted having taken medications at “Somewhere in between medication doses point” with those who are submitted “Just after Parkinson medication (at your best)”. By applying this method, we could tell the relationship between these three medication time points.

Finally, we applied the six models to two PD patients, one male and one female, for their audio files submitted “Immediately before Parkinson medication” and “Just after Parkinson medication (at your best)”. We first applied six models on a single feature, which performed best for the population datasets. After selecting the model with the highest accuracy score, we used more features for each patient instead of just applying all the selected features for the whole population. By using the feature and model with best performance for each PD patient, we randomly selected audio samples in descending order, which allowed us to find the desired number of audio files needed for predicting the medication time points for each single patient.

Results:

Feature selection:

The desired number of selected features, and the name for these features are shown in Figure 1 and Table 2. Figure 1 is a plot of number of features vs. the cross-validation scores for all features for audio files submitted "Immediately before Parkinson medication" and "Just after Parkinson medication (at your best)" using the Support vector machine Recursive feature elimination (SVM-RFE) with cross-validation, with 10 folds. From the plot, five features have the highest cross-validation score.

For the ranking of these 34 features in Table 2, we selected the following five features: the 1st and 6th MFCC, the 4th, and 6th of Chroma Vector, and Spectral Flux. Mel Frequency Cepstral Coefficients (MFCCs) was first described by Davis and Mermelstein in 1980 (Davis and Mermelstein, 1980). It is a kind of spectrum-based feature, which is able to provide specific information for the speaker (Murty and Yegnanarayana, 2006). Chroma Vector was first introduced by Fujishima in 1999 (Fujishima, 1999). Each parameter of the Chroma vector represents a spectral energy of a semi-tone on the chromatic scale (Oudre et al., 2011). Spectral Flux is a feature for measuring the degree of variation of the power spectrum of a signal. Since voice is a slow-changing signal, it will not change rapidly from one frame to another. This could be an important quality for distinguishing the voice of a participants from another (Sadjadi and Hansen, 2013).

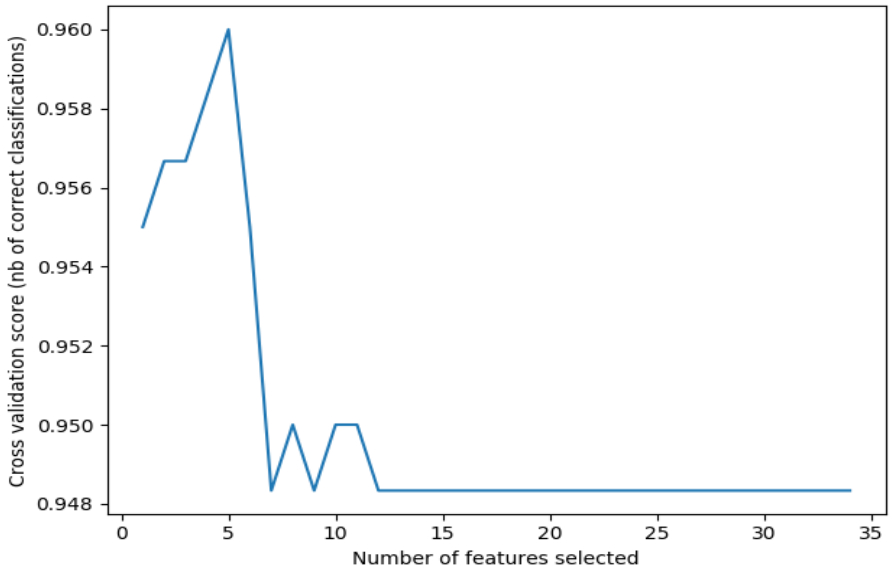

Figure 1: the cross-validation score corresponding to the number of selected features

Feature Name	Ranking	Feature Name	Ranking
Spectral Spread	11	Zero Crossing Rate	17
Spectral Entropy	20	Energy	25
Spectral Flux	1	Entropy of Energy	12
Spectral Rolloff	14	Spectral Centroid	10
MFCC 1	1	Chroma Vector 1	30
MFCC 2	2	Chroma Vector 2	6
MFCC 3	5	Chroma Vector 3	3
MFCC 4	24	Chroma Vector 4	1
MFCC 5	29	Chroma Vector 5	15
MFCC 6	1	Chroma Vector 6	1
MFCC 7	21	Chroma Vector 7	8
MFCC 8	9	Chroma Vector 8	28

MFCC 9	16	Chroma Vector 9	27
MFCC 10	26	Chroma Vector 10	22
MFCC 11	23	Chroma Vector 11	18
MFCC 12	4	Chroma Vector 12	13
MFCC 13	7	Chroma Deviation	19

Table 2: the ranking of 34 features based on Recursive feature elimination with cross-validation, the selected features (estimated as best) were assigned as 1

Evaluation model for two types of medication time points:

For the whole population, we first considered the audio files submitted "Immediately before Parkinson medication" (n = 4991) and "Just after Parkinson medication" (n = 5369) only. By applying cross-validation for stratified K-fold cross validation, with 10 folds, the accuracy scores for the six selected models are shown in Table 3. The receiver operating characteristic (ROC) curve is given in Figure 2.

From the Table 3 and Figure 2, we found that there was no significant difference between these models as well as the selected features. The accuracy scores for almost all of them are around 0.52. Among these five features, spectral flux has a relatively higher accuracy score. From the ROC curve, we found that for all the models, they just performed slightly better than random guessing.

	MFCC 1	**MFCC 6**	**Spectral Flux**	**Chroma Vector 4**	**Chroma Vector 6**

SVM	0.49	0.50	0.50	0.51	0.52
Random Forest	0.50	0.51	0.51	0.51	0.50
Logistic Regression	0.51	0.53	0.55	0.52	0.52
Decision Tree	0.52	0.52	0.53	0.52	0.52
Gaussian	0.51	0.53	0.55	0.52	0.52
MLP	0.53	0.52	0.54	0.53	0.51

Table 3: accuracy score for standardized datasets collected when medication time points were "Immediately before Parkinson medication" and "Just after Parkinson medication"

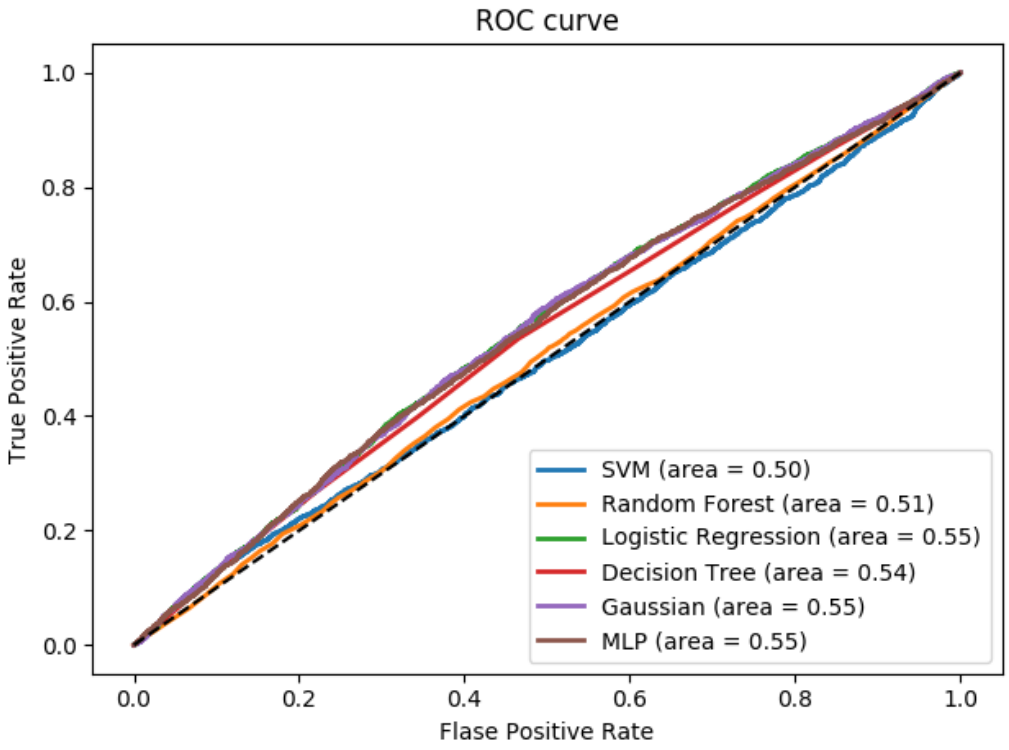

Figure 2: receiver operating characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) for the selected six models

Evaluation model for three types of medication time points:

To get a brief idea about “Somewhere in between medication doses” (n = 10229), the mean value of spectral flux based on time was calculated. The feature spectral flux was chosen since it had the highest accuracy score when only the audio files submitted “Immediately before Parkinson medication” and “Just after Parkinson medication” were considered.

Figure 3 shows the changes of spectral flux with time during the 10 seconds of recording for three different medication statuses. A clear trend can be seen from Figure 3, where the average values of spectral flux for PD patients with different medication time points increases rapidly within the first second and then drops down from second one to five. Later, it increases slightly for the last five seconds. We can see a clear gap between the average value of spectral flux for audios submitted “Immediately before Parkinson medication” and the rest two situations. This indicates that comparing with the voice audios submitted “Immediately before Parkinson medication”, voice audios submitted “Somewhere in between medication doses” seem to relate more to the voice audios submitted “Just after Parkinson medication”. At the same time, PD patients who recorded the audio files neither immediately before nor after Parkinson medication have a worse performance after recording for five seconds comparing with PD patients who recorded immediately after taking the Parkinson medication, even though these two groups of patients share almost the same performance in the first five seconds.

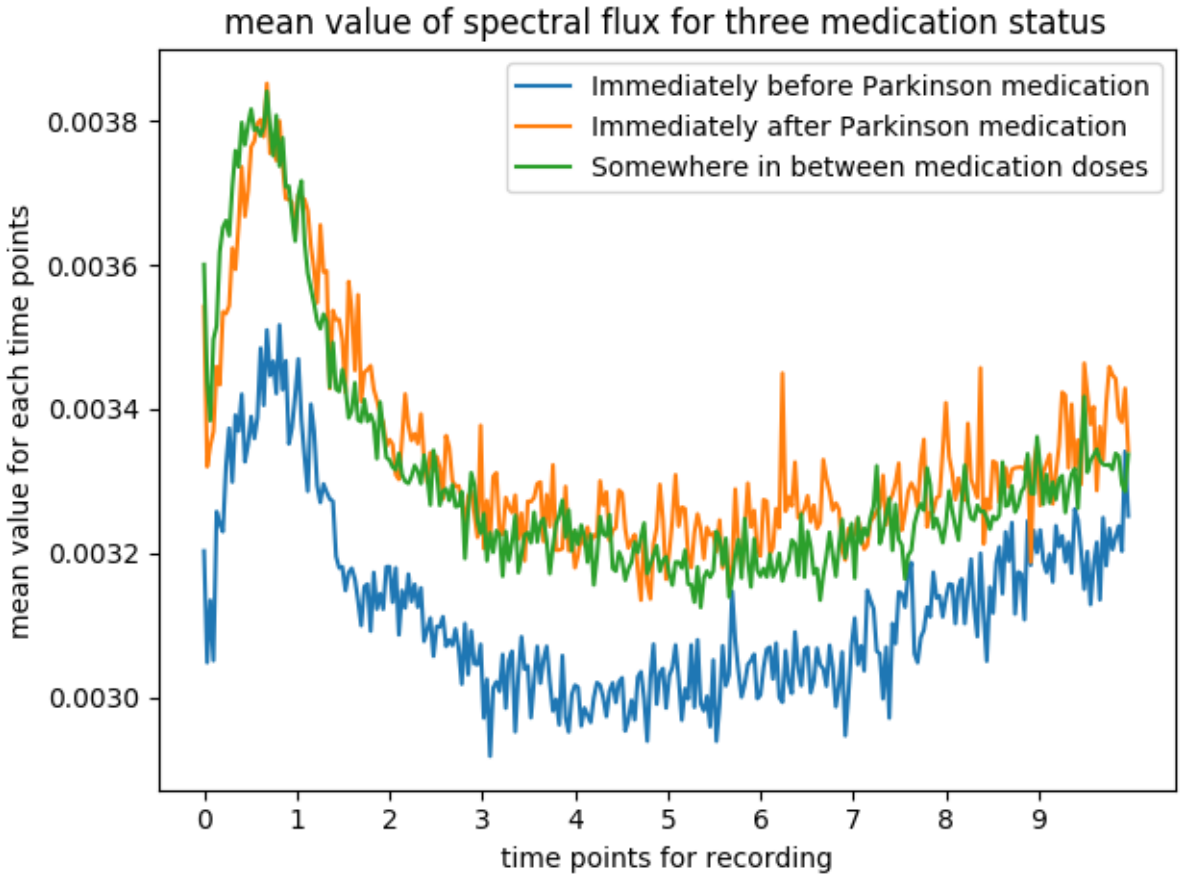

Figure 3: mean value of spectral flux for three groups: "Somewhere in between medication doses", "Immediately before Parkinson medication", and "Just after Parkinson medication"

However, the accuracy scores shown in Table 4 and 5 do not imply significantly that the voice files for PD patients who submitted the recording "Somewhere in between medication doses" has a closer relationship with the voice files submitted "Just after Parkinson medication". In Table 4, we combined the voice data recorded "Somewhere in between medication doses" with the data recorded "Just after Parkinson medication" together, and compared with "Immediately before Parkinson medication". For each group, we randomly selected the same number of vocal files, making it an un-skewed dataset. The accuracy scores are around 0.53. In Table 5, we combined

the voice data recorded “Somewhere in between medication doses” with the data recorded “Immediately before Parkinson medication” together, and compared with “Just after Parkinson medication”. The accuracy scores are around 0.51 and the results are slightly better than random guessing. This is the same situation for considering only data recorded “Immediately before Parkinson medication” and “Just after Parkinson medication”. Therefore, the accuracy scores for Table 4 is slightly better than the scores for Table 5. Although this might not be a significant improvement, this might indicate a closer relationship between audio submitted “Somewhere in between medication doses” and the audio submitted “Just after Parkinson medication” together.

	MFCC 1	**MFCC 6**	**Spectral Flux**	**Chroma Vector 4**	**Chroma Vector 6**
SVM	0.50	0.50	0.52	0.51	0.51
Random Forest	0.50	0.51	0.51	0.52	0.51
Logistic Regression	0.50	0.53	0.53	0.53	0.53
Decision Tree	0.50	0.50	0.54	0.52	0.51
Gaussian	0.51	0.54	0.53	0.53	0.55
MLP	0.51	0.54	0.53	0.53	0.53

Table 4: accuracy score for standardized datasets collected when medication time points were “Immediately before Parkinson medication” and combined datasets for “Somewhere in between medication doses” and “Just after Parkinson medication”

	MFCC 1	**MFCC 6**	**Spectral Flux**	**Chroma Vector 4**	**Chroma Vector 6**

SVM	0.50	0.51	0.50	0.49	0.51
Random Forest	0.50	0.50	0.51	0.51	0.50
Logistic Regression	0.51	0.52	0.53	0.51	0.52
Decision Tree	0.51	0.51	0.52	0.50	0.52
Gaussian	0.50	0.50	0.51	0.52	0.51
MLP	0.50	0.51	0.52	0.50	0.51

Table 5: accuracy score for standardized datasets collected when medication time points were "Just after Parkinson medication" and combined datasets for "Somewhere in between medication doses" and "Immediately before Parkinson medication"

The accuracy result for the total population did not work well. Therefore, we tried to calculate the accuracy score for a single PD patient. We selected two PD patients, one male and one female, for building a binary classifier.

The accuracy scores shown in Table 6 imply that different models have various performance for different individuals. By considering voice data recorded "Immediately before Parkinson medication" and "Just after Parkinson medication" only, for this single male, Gaussian Naïve Bayes has the highest accuracy score; while for this single female, Random Forest has the highest accuracy score.

	Male	**Female**
SVM	0.60	0.56
Random Forest	0.50	0.83
Logistic Regression	0.67	0.63
Decision Tree	0.60	0.60

Gaussian	0.71	0.62
MLP	0.66	0.57

Table 6: accuracy scores for two PD patients' spectral flux data collected when medication time points were "Just after Parkinson medication", and "Immediately before Parkinson medication"

Therefore, we applied Gaussian Naïve Bayes for this single male PD patient and Random Forest for this single female PD patient. In Table 7, the accuracy scores of Gaussian Naïve Bayes for the voice features of this male patient and the accuracy scores of Random Forest for the female patient are shown. For this male patient, vocal feature "entropy of energy" has the best performance; and for this female patient, spectral flux has the highest accuracy score.

	Male	**Female**
Entropy of Energy	0.82	0.78
Spectral Rolloff	0.56	0.55
Spectral Entropy	0.7	0.36
Spectral Flux	0.71	0.83
MFCC 1	0.68	0.51
MFCC 2	0.7	0.67
MFCC 3	0.56	0.44
MFCC 4	0.6	0.57
MFCC 5	0.62	0.27
MFCC 6	0.51	0.51
MFCC 12	0.73	0.41
Chroma Vector 2	0.62	0.46
Chroma Vector 3	0.69	0.77
Chroma Vector 4	0.6	0.48
Chroma Vector 6	0.56	0.4

Table 7: accuracy scores of Gaussian Naïve Bayes for the male patient and Random Forest for the female patient

To understand the relationship between number of samples collected and the accuracy of classification within an individual, we randomly selected a decreasing number of audio samples for both medication time points ("Just after Parkinson medication" and "Immediately before Parkinson medication"). Figure 4 shows a plot of medication classification accuracy for each number of samples. After sample number selection for both medication time points, we applied stratified K-fold cross validation, with 3 folds (due to the limited number of samples for each patient). This procedure was repeated 100 times and the mean value of the accuracy score was calculated. Figure 4 represents the variation of accuracy for each number of voice samples. Fewer number of samples included in the analysis yields lower accuracy scores. A relative maximum accuracy is found using 12 voice samples for the male subject, and 11 for the female subject, for both before and after medications.

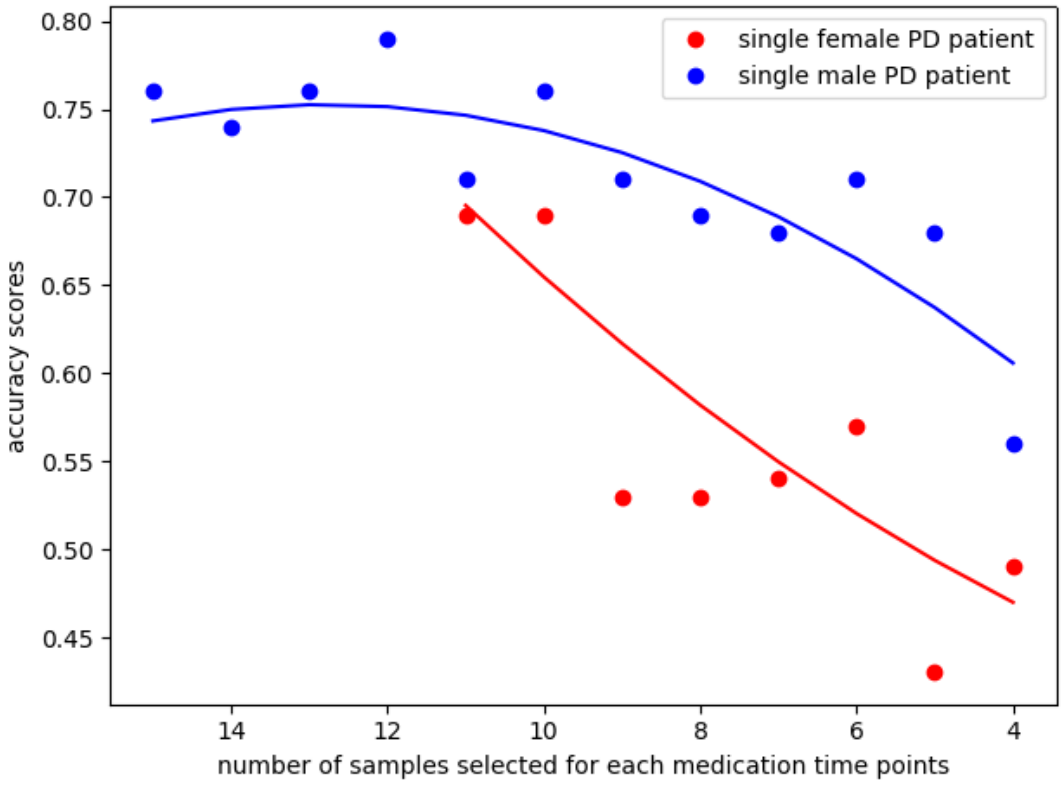

Figure 4: variation of accuracy scores for single PD patients with selected features and models with decreasing number of input datasets

Discussion:

From the results for the medication time points of "Just after Parkinson medication", and "Immediately before Parkinson medication" for the whole population, the models do not predict the group of medication time points for the testing datasets.

Due to the poor model accuracy result, we tried to find out if there existed some control variables that influenced the correlation between the medication status and the voice features. One of them might be gender. Thus, we separated the PD patients by male (2723 files recorded "Immediately before Parkinson medication", and 2978 files recorded "Just after Parkinson medication") and female (2268 files recorded "Immediately before Parkinson medication", and 2391 files recorded "Just after Parkinson medication") and re-calculated the accuracy of the six models: logistic regression, support vector machine (SVM), decision trees, Gaussian Naïve Bayes and Multi-layer Perceptron (MLP). However, the result did not improve significantly. For the male population, the highest accuracy score was 0.56. This score was achieved by applying logistic regression model for the male spectral flux feature. For the female population, the highest accuracy score was 0.53, which was calculated from the vocal feature MFCC6 by applying decision trees and Gaussian Naïve Bayes.

The major limitation of this study was the large number of control variables that would affect the relationship between medication status and value of selected vocal features. The only control variable we had considered was gender. There might be some other control variables including the severity of the disease, the age, the distance between the recording equipment and the patient, the emotion of the participants during recording, the smoking situation of the patient, and the type of voice recording equipment (Sorensen and Horii, 1982) (Van den Stock et al., 2007). By considering all the possible control variables, a high accuracy score might be achieved. However, our data did not include these variables and we could not divide the participant based on these variables.

Therefore, instead of training on the whole population, we built models for single PD patients. Different patients had different novel features and models. This is the reason for the low accuracy result for the whole population of PD patients.

Despite the limitations, our study confirmed the vocal differences in individual PD patients who recorded “Immediately before Parkinson medication”, and “Just after Parkinson medication” at a theoretical level. To get more than 75% of accuracy, at least twelve data files for each medication status were needed. This indicates that the voice for PD patients may be helpful for them to identify their physical condition at that time point and determine if they need to take medication or not. For further study, categorical models are needed to predict the possibility for this single patient to take medication.

Conclusion:

Our findings suggest that the medication status of a PD patient will influence their vocal features. If we are considering building models for the whole population, more control variables are needed besides gender. Based on current findings, we can figure out that different people have different novel vocal feature and model for medication time point prediction. Thus, continuing research on single PD patient is more likely to lead to new insights into measuring the current severity of disease for this PD patient and assist medical decision-making.

References:

1. APRN, A.A., MS, CNL, 2017. A Practical Guide to Parkinson's Disease: Diagnosis and Management. Springer Publishing Company.
2. Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K., Seitelberger, F., 1973. Brain dopamine and the syndromes of Parkinson and Huntington Clinical, morphological and neurochemical correlations. J. Neurol. Sci. 20, 415–455. https://doi.org/10.1016/0022 510X(73)90175-5
3. Canter, G.J., 1963. Speech Characteristics of Patients with Parkinson's Disease: I. Intensity, Pitch, and Duration. J. Speech Hear. Disord. 28, 221–229. https://doi.org/10.1044/jshd.2803.221
4. Chang, D., Nalls, M.A., Hallgrímsdóttir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., Hinds, D., Behrens, T.W., Singleton, A.B., Bhangale, T.R., Graham, R.R., 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat. Genet. 49, 1511–1516. https://doi.org/10.1038/ng.3955
5. Cosentino, M., Martignoni, E., Michielotto, D., Calandrella, D., Riboldazzi, G., Pacchetti, C., Frigo, G., Nappi, G., Lecchini, S., 2005. Medical healthcare use in Parkinson's disease: survey in a cohort of ambulatory patients in Italy. BMC Health Serv. Res. 5, 26. https://doi.org/10.1186/1472-6963-5-26
6. Davis, S., Mermelstein, P., 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process. 28, 357–366. https://doi.org/10.1109/TASSP.1980.1163420
7. Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. Science 322, 970–973.
8. Géron, A., 2017. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.
9. Giannakopoulos, T., 2015. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. PLOS ONE 10, e0144610. https://doi.org/10.1371/journal.pone.0144610

10. Guneysel, O., Onultan, O., Onur, O., 2008. Parkinson's disease and the frequent reasons for emergency admission. Neuropsychiatr. Dis. Treat. 4, 711–714.
11. Harel, B., Cannizzaro, M., Snyder, P.J., 2004. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. Brain Cogn. 56, 24–29. https://doi.org/10.1016/j.bandc.2004.05.002
12. Murty, K.S.R., Yegnanarayana, B., 2006. Combining evidence from residual phase and MFCC features for speaker recognition. IEEE Signal Process. Lett. 13, 52–55. https://doi.org/10.1109/LSP.2005.860538
13. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., Consortium (IPDGC), I.P.D.G., Initiative (PROGENI), P.S.G. (PSG) P.R.T.O.Gen., 23andMe, GenePD, Consortium (NGRC), N.R., Genomics (HIHG), H.I. of H., Investigator, T.A.J.D., Epidemiology (CHARGE), C. for H. and A.R. in G., Consortium (NABEC), N.A.B.E., Consortium (UKBEC), U.K.B.E., Consortium, G.P.D., Group, A.G.A., Ikram, M.A., Ioannidis, J.P.A., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., Singleton, A.B., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. 46, 989–993. https://doi.org/10.1038/ng.3043
14. Oudre, L., Grenier, Y., Fevotte, C., 2011. Chord Recognition by Fitting Rescaled Chroma Vectors to Chord Templates. IEEE Trans. Audio Speech Lang. Process. 19, 2222–2233. https://doi.org/10.1109/TASL.2011.2139205
15. Parkinson, J., 2002. An Essay on the Shaking Palsy. J. Neuropsychiatry Clin. Neurosci. 14, 223–236. https://doi.org/10.1176/jnp.14.2.223
16. Parkinson Study Group, 2002. Dopamine transporter brain imaging to assess the effects of pramipexole vs levodopa on Parkinson disease progression. JAMA 287, 1653–1661.
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825−2830.
18. Sadjadi, S.O., Hansen, J.H.L., 2013. Unsupervised Speech Activity Detection Using Voicing Measures and Perceptual Spectral Flux. IEEE Signal Process. Lett. 20, 197–200. https://doi.org/10.1109/LSP.2013.2237903
19. Savitt, J.M., Dawson, V.L., Dawson, T.M., 2006. Diagnosis and treatment of Parkinson disease: molecules to medicine. J. Clin. Invest. 116, 1744–1754. https://doi.org/10.1172/JCI29178
20. Schulz, G.M., Grant, M.K., 2000. Effects of speech therapy and pharmacologic and surgical treatments on voice and speech in parkinson's disease: A review of the literature. J. Commun. Disord. 33, 59–88. https://doi.org/10.1016/S0021-9924(99)00025-8
21. Sorensen, D., Horii, Y., 1982. Cigarette smoking and voice fundamental frequency. J. Commun. Disord. 15, 135–144. https://doi.org/10.1016/0021-9924(82)90027-2

22. Van den Stock, J., Righart, R., de Gelder, B., 2007. Body expressions influence recognition of emotions in the face and voice. Emotion 7, 487–494. https://doi.org/10.1037/1528-3542.7.3.487
23. Yan, K., Zhang, D., 2015. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens. Actuators B Chem. 212, 353–363. https://doi.org/10.1016/j.snb.2015.02.025