

Gradient Boosting for Parkinson's Disease Diagnosis from Voice Recordings

Ibrahim Karabayir

Loyola University Chicago

Samuel Goldman

University of California San Francisco

Suguna Pappu

Loyola University Chicago

Oguz Akbilgic (✉ oakbilgic@luc.edu)

Loyola University Chicago <https://orcid.org/0000-0003-0313-9254>

Research article

Keywords: Parkinson's Disease, Gradient Boosting, Machine Learning, Artificial Intelligence

DOI: <https://doi.org/10.21203/rs.2.20727/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Parkinson's Disease (PD) is a clinically diagnosed neurodegenerative disorder that affects both motor and non-motor neural circuits. Speech deterioration (hypokinetic dysarthria) is a common symptom, which often presents early in the disease course. Machine learning can help movement disorders specialists improve their diagnostic accuracy using non-invasive and inexpensive voice recordings.

Method: We used "Parkinson Dataset with Replicated Acoustic Features Data Set" from the UCI-Machine Learning repository. The dataset included 44 speech-test based acoustic features from patients with PD and controls. We analyzed the data using various machine learning algorithms including Light and Extreme Gradient Boosting, Random Forest, Support Vector Machines, K-nearest neighborhood, Least Absolute Shrinkage and Selection Operator Regression, as well as logistic regression. We also implemented a variable importance analysis to identify important variables classifying patients with PD.

Results: The cohort included a total of 80 subjects: 40 patients with PD (55% men) and 40 controls (67.5% men). Disease duration was 5 years or less for all subjects, with a mean Unified Parkinson's Disease Rating Scale (UPDRS) score of 19.6 (SD 8.1), and none were taking PD medication. The mean age for PD subjects and controls was 69.6 (SD 7.8) and 66.4 (SD 8.4), respectively. Our best-performing model used Light Gradient Boosting to provide an AUC of 0.951 with 95% confidence interval 0.946-0.955 in 4-fold cross validation using only seven acoustic features.

Conclusions: Machine learning can accurately detect Parkinson's disease using an inexpensive and non-invasive voice recording. Light Gradient Boosting outperformed other machine learning algorithms. Such approaches could be used to inexpensively screen large patient populations for Parkinson's disease.

Background

Parkinson's disease (PD) is a neurodegenerative disorder of largely unknown cause [1]. After Alzheimer's disease, it is the second most common neurodegenerative disease [2]. In 2010, there were approximately 680,000 people over 45 years old with PD in the US in 2010 and this number is expected to rise to 1,238,000 in 2030 [3]. By the time PD becomes clinically apparent, more than 50% of dopaminergic neurons in the substantia nigra have been lost, with a corresponding 80% decline in striatal dopamine levels [4, 5]. Thus, early identification of disease is essential if neuroprotective therapies are to be implemented.

Diagnosis of PD currently relies on clinical examination. The current gold standard is based on motor signs and symptoms (bradykinesia, resting tremor, rigidity, postural reflex impairment) and response to dopaminergic drugs [6, 7]. In addition to the classic motor signs and symptoms, PD is well-recognized to also affect non-motor neural circuits [6, 8, 9]. However, the accuracy of PD diagnosis in practice is only around 80% [10, 11], implying that a large population with PD is undiagnosed or misdiagnosed [12]. Hence, identification of novel motor or non-motor markers of PD or improving the accuracy of currently

available diagnostic tools is important, particularly in early disease. Noninvasive speech tests have been explored as a marker of disease [11, 13], since deterioration of speech is consistently observed in patients with PD [14-16]. Naranjo et al [17, 18] previously showed that patients with PD could be identified with moderately high accuracy using acoustic features extracted from a speech test. In this study, we implemented machine learning methods, specifically Light [19] and Extreme Gradient Boosting [20], to significantly improve PD detection accuracy from acoustic features extracted from voice recordings.

Methods

Data: We utilized “Parkinson Dataset with Replicated Acoustic Features Data Set” that was donated to University of California Irvine Machine Learning repository by Naranjo, et al [17] in April 2019. The publicly available data we used in this study were first presented by Goetz, et al [21], and other than sex, individual-level descriptors are not publicly available. However, they reported that the dataset includes patients with early-stage PD not taking medication. A follow-up study [11] reported that PD duration was 5 years or less for all subjects, with a mean Unified Parkinson’s Disease Rating Scale (UPDRS) score of 19.6 (SD=8.1). The dataset available to us [17] included 44 acoustic features extracted from voice recordings of 40 patients with PD and 40 controls. Recordings of a sustained phonation of the vowel /a/ for 5 seconds were repeated three times (three runs). Digital recordings were implemented at a 44.1 KHz sampling rate and 16 bits/sample [17].

The 44 acoustic features extracted from voice recordings comprised five categories: pitch and amplitude local perturbation, noise, spectral envelope, and nonlinear measures. Four pitch local features (jitter relative, jitter absolute, jitter RAP (relative absolute perturbation)), jitter PPQ (pitch perturbation quotient), and five amplitude perturbation measures (shimmer local, shimmer dB, APQ3 (3 point Amplitude Perturbation Quotient), APQ5 (5 point Amplitude Perturbation Quotient), and APQ11 (11-point Amplitude Perturbation Quotient)) were extracted using a waveform matching algorithm. As measures of relative level of noise in speech [17], five different variants of harmonic-to-noise ratio (HNR) corresponding to different frequency bandwidths (HNR05 [0-500 Hz], HNR15 [0-1500 Hz], HNR25 [0-2500 Hz], HNR35 [0-3500 Hz], HNR38 [0-3800 Hz]) [22]. Glottal-to-Noise Excitation Ratio (GNE), which quantifies the amount of voice excitation, was also calculated. Since PD is known to affect articulation [23], 13 Mel Frequency Cepstral Coefficients (MFCCs) associated with articular position and 13 Delta Coefficients as time dependent derivatives of MFCCs were also extracted. In addition, Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), and Pitch Period Density Entropy (PPE) were also extracted as non-linear measures of voice recordings. Further details of the dataset can be found in Naranjo et al. [17].

Features: Speech deterioration is one of the motor symptoms of PD [14, 24-26]. Patients have reduced pitch variability compared to controls as well as reduced intra-individual variability [27, 28]. As described above, each acoustic feature was calculated three times for different runs of the speech test. Thus, in addition to testing the diagnostic accuracy of our analytic approach, we were also able to investigate intra-individual changes in response from different runs of the test. We considered acoustic features

calculated for all three runs as individual predictors. Moreover, for a given acoustic feature, we created three artificial variables representing the change from one run to another (Figure 1). Therefore, our feature set included 264 acoustic features and sex for 80 subjects.

Classification: We implemented gradient boosting algorithms to distinguish between subjects with PD and controls. Gradient boosting is an ensemble machine learning consisting of several weak models (shallow decision trees rather than overfitting deep ones) and it can be used for both regression and classification problems [19, 20]. Because it uses weak classifiers, it is more robust against overfitting compared to a random forest, a similar method that allows overfitting of individual tree predictors [20, 29, 30]. In our work, we mainly implement 4-fold cross validation to identify any overfitting by randomly splitting data into four distinct folds. We also repeat this process multiple times and present average results. We considered two gradient boosting machines Extreme Gradient Boosting (XGB) and Light Gradient Boosting (LGB), and for comparison, the more traditional machine learning algorithms Random Forest (RF), Support Vector Machines (SVM), K-nearest neighborhood (KNN), Least Absolute Shrinkage and Selection Operator (LASSO) regression to implement regularization, and a statistical approach, Logistic Regression (LR).

Variable Importance Analysis, Feature Selection, and Re-Classification: We first built the gradient boosting model using 265 features with four folds cross-validation and repeated this process 100 times. At each run, for each model built within 4-fold cross validation (4x100 models), we implemented a feature importance analysis that calculates the relative contribution of each feature to the corresponding model. A higher value of this metric for a specific feature implies it as a more important feature than another feature that has lower value of this metric [31]. By averaging the feature importance obtained from 400 individual models, we obtained a ranking of the 265 features. Next, we built new classification models with 4-fold cross-validation by incrementally adding the top 15 most important features selected from the previous step into the model with respect to their importance ranking. We repeated each of these steps 100 times to better estimate the effect of each feature on the model performance when they are introduced into the model. We then identified the step where the model performance started diminishing or stopped increasing. Finally, using the features introduced up to that specific step, we rebuilt gradient boosting models with 4-fold cross validation and report various performance metrics such as specificity, sensitivity, positive predictive value, accuracy, F1 score, and area under the receiver operating characteristics curve (AUC).

Results

Cohort: Our cohort included 40 subjects with PD (55% men) and 40 healthy controls (67.5% men). All subjects were over 50 years of age and the mean age and standard deviation (SD) for PD subjects and controls was 69.6 (SD 7.8) and 66.4 (SD 8.4), respectively. PD diagnosis required at least two of resting tremor, bradykinesia or rigidity [21], and no evidence for other forms of parkinsonism.

Classification: We initially built the classification models with 4-fold cross-validation using the entire set of 265 predictors. We repeated each classification model 100 times by randomly splitting the data into four folds. Various classification performance metrics with their 95% Confidence Intervals (CI) are presented in Table 1. LGB provided the highest F1 score of 0.878 with 95% CI 0.871-0.884, and AUC of 0.951 (95%CI 0.946-0.955).

Table 1 Comparison of alternative machine learning methods. (LGB: Light Gradient Boosting, XGB: Extreme Gradient Boosting, LR: Logistic Regression, SVM: Support Vector Machines, RF: Random Forest, KNN: K-nearest Neighbor, LASSO: Least Absolute Shrinkage and Selection Operator)

| Metrics | | Accuracy Metrics with 95 % CI | | | | | |
|-------------|----------------------|-------------------------------|---------------|---------------|---------------|---------------|----------------|
| | LGB | XGB | LR | SVM | RF | KNN | LASSO |
| F1 | 0.839 | 0.810 | 0.771 | 0.730 | 0.810 | 0.744 | 0.763 |
| | [0.831-0.847] | [0.802-0.819] | [0.762-0.780] | [0.721-0.739] | [0.800-0.819] | [0.735-0.753] | [0.755-0.7723] |
| AUC | 0.898 | 0.891 | 0.839 | 0.838 | 0.884 | 0.841 | 0.870 |
| | [0.892-0.905] | [0.885-0.898] | [0.830-0.847] | [0.830-0.846] | [0.876-0.892] | [0.834-0.848] | [0.863-0.877] |
| Accuracy | 0.841 | 0.816 | 0.771 | 0.744 | 0.818 | 0.760 | 0.761 |
| | [0.833-0.849] | [0.809-0.823] | [0.762-0.780] | [0.735-0.752] | [0.810-0.826] | [0.752-0.768] | [0.753-0.769] |
| Sensitivity | 0.839 | 0.801 | 0.777 | 0.704 | 0.795 | 0.712 | 0.782 |
| | [0.827-0.850] | [0.789-0.813] | [0.765-0.790] | [0.691-0.716] | [0.782-0.808] | [0.699-0.725] | [0.769-0.794] |
| Specificity | 0.844 | 0.830 | 0.764 | 0.784 | 0.841 | 0.807 | 0.741 |
| | [0.832-0.855] | [0.819-0.841] | [0.750-0.778] | [0.771-0.798] | [0.831-0.852] | [0.796-0.818] | [0.729-0.754] |
| PPV | 0.853 | 0.835 | 0.780 | 0.780 | 0.844 | 0.796 | 0.762 |
| | [0.843-0.863] | [0.825-0.845] | [0.769-0.791] | [0.769-0.791] | [0.834-0.854] | [0.786-0.806] | [0.753-0.772] |

Variable Importance Analysis: As described in the Methods section, using the total of 400 models obtained through 100 runs of 4-fold cross-validation, we obtained variable rankings based on their importance in classification in the LGB algorithm. The top 15 variables are shown on the x-axis of Figure 2.

Feature Selection and Re-classification: To obtain a compact model, we repeated our 4-fold classification strategy 15 times by incrementally introducing a new variable into the model based on the order of importance. Figure 2 summarizes the accuracy metrics with associated 95% CIs for each step of this re-classification.

Figure 2 shows that all accuracy metrics gradually increase (F1-score of 0.878 (95%CI 0.871-0.884), AUC of 0.951 (95% CI 0.946-0.955), Overall Accuracy of 0.880 (95% CI 0.873-0.886), Sensitivity of 0.872 (95% CI 0.862-0.882), Specificity of 0.887 (95% CI 0.877-0.896), Positive Predictive Value of 0.892 (95% CI 0.884-0.901)) in the first seven steps of the feature selection protocol and then slightly decline in following steps. In other words, after introducing the top seven variables - Delta3 (Run2), Delta9 (Run3), Delta0 (Run 3), MFCC4 (Run 2), MFCC10 (Run 2), MFCC8 (Run 2), and HNR15 (Run 1) - into the model, additional variables did not improve the classification accuracy. We further implemented a grid search by changing the learning rates and feature and bagging fraction to identify whether the performance could be improved. However, there was no significant difference in AUC values of models with different parameter settings.

Independent sample two-tail t-tests showed that the means of these top seven selected features significantly ($p < 0.05$) differed for PD cases and controls. To identify whether such differences exist for all three runs, we further implemented t-tests for those seven features for all runs. Our results showed the top seven acoustic features significantly ($p < 0.05$) differ for PD cases and controls across all three runs, however, the p-values are smaller for the runs that were listed in top seven features.

Sensitivity Analysis: The main reason for implementing 4-fold cross-validation in our work was to make our results comparable to the work of Naranjo et al. [17, 18], which is the original study utilizing these data. However, using the top seven variables, we also repeated our cross-validation on the compact model for 5- and 10- fold cross-validation for the light gradient boosting model and obtained F1-score of 0.879 (95%CI 0.872, 0.886) and 0.875 (0.867, 0.883), respectively.

The above models analyzed acoustic features from three runs as individual predictors. In a sensitivity analysis we explored whether using the average of acoustic features across the three runs might improve the model. This classification approach performed more poorly, with an F1-score of 0.819 (95% CI 0.812, 0.827) vs. 0.878 (95%CI 0.871, 0.884) for the individual predictor model.

Discussion

We were able to accurately classify persons with Parkinson's disease by analysis of voice recordings using machine learning. Acoustic features extracted from speech test recordings offer a potential application for computerized non-invasive diagnostic tools. The data we used in this study included 44 acoustic features generated separately for three runs of the same speech test task. In their original studies on the same data, Naranjo et al [17, 18] proposed a statistical approach that treated the results of

these runs as repeated measures. The Light Gradient Boosting model presented here outperformed the statistical approach in all metrics: AUC 0.951 vs. 0.879; sensitivity 0.872 vs 0.765; specificity 0.887 vs 0.792; precision 0.887 vs. 0.806; and overall accuracy 0.880 vs 0.779. Moreover, we could reach this level of accuracy using only seven features.

As reported above, Delta3 (Run2), Delta9 (Run3), Delta0 (Run 3), MFCC4 (Run 2), MFCC10 (Run 2), MFCC8 (Run 2), and HNR15 (Run 1) variables were the most important classifiers, and that these features were indeed significantly different for PD cases and controls across all three runs.

It is worth noting that only one of the seven acoustic variables obtained from the first run of the speech test was selected as a predictor in the final model. Four variables were from second run, and two from the third run. None of the variables representing changes from one run to another were selected as one of the top seven variables.

This study demonstrates that machine learning can assist clinicians in the accurate diagnosis of PD. Since the PD subjects in this study were in their early stages of disease, this approach may provide an opportunity for earlier diagnosis of PD. Future work should investigate whether such acoustic patterns exist during the prodromal phase of PD.

Our study has several limitations. The most important limitation is the small sample size. Despite the fact that our carefully designed cross-validation yielded very high accuracy, there is a need to repeat these analyses in a larger cohort. Moreover, the small sample size may also limit inferences of variable importance. Despite the fact that our model performed with high classification accuracy, the feature importance analysis must be cautiously interpreted since the ranks of importance may change when the study is repeated in a larger cohort. Additionally, all PD subjects in the dataset were drawn from a single study. External validation is needed to test the broader generalizability of our model. Another important limitation of our study is that our dataset includes only subjects with PD and controls. It is unclear whether our model can distinguish between subjects with PD and those with other diseases that can affect speech.

Conclusions

Gradient boosting algorithms can be used to identify patients with Parkinson's disease using a simple non-invasive speech test. Further studies are required to determine whether similar results can be obtained from records of normal conversation or phone calls. This approach could be used to screen large patient populations at different stages of Parkinson's disease. The value of this approach to identify early prodromal PD remains to be determined.

Abbreviations

PD: Parkinson's Disease

UPDRS: Unified Parkinson's Disease Rating Scale

SD: Standard Deviation

RAP: Relative Absolute Perturbation

PPQ: Pitch Perturbation Quotient

ABQ: Amplitude Perturbation Quotient

HNR: Harmonic-to-noise Ratio

MFCC: Mel Frequency Cepstral Coefficient

RPDE: Recurrence Period Density Entropy

DFA: Detrended Fluctuations Analysis

PPE: Pitch Period Density Entropy

XGB: Extreme Gradient Boosting

LGB: Light Gradient Boosting

RF: Random Forest

SVM: Support Vector Machines

KNN: K-nearest Neighborhood

LASSO: Least Absolute Shrinkage and Selection Operator

LR: Logistic Regression

AUC: Area Under the Receiver Operator Characteristics Curve

CI: Confidence Interval

Declarations

Ethics approval and consent to participate: NA. This manuscript is based on a publicly available de-identified data from UCI-Machine Learning Data Repository.

Availability of data and material: The data is available at UCI-Machine Learning Data Repository named "Parkinson Dataset with replicated acoustic features Data Set".

Consent for publication: All authors have approved this submission.

Competing interests: No author has any conflict of interest to report.

Funding: There is no funding supported this project.

Authors' contributions: All authors have read and approved the manuscript. OA and SG conceptualized the study. IK and OA carried out the analysis and drafted the first version. IK, SG, SP, OA provided critical feedback and contributed in writing the manuscript.

Acknowledgements: NA

References

- [1] C. M. Tanner and S. M. Goldman, "Epidemiology of Parkinson's disease," (in eng), *Neurol Clin*, vol. 14, no. 2, pp. 317-35, May 1996, doi: 10.1016/s0733-8619(05)70259-0.
- [2] E. R. Dorsey *et al.*, "Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030," (in eng), *Neurology*, vol. 68, no. 5, pp. 384-6, Jan 30 2007, doi: 10.1212/01.wnl.0000247740.47667.03.
- [3] C. Marras *et al.*, "Prevalence of Parkinson's disease across North America," (in eng), *NPJ Parkinsons Dis*, vol. 4, p. 21, 2018, doi: 10.1038/s41531-018-0058-0.
- [4] J. M. Fearnley and A. J. Lees, "Ageing and Parkinson's disease: substantia nigra regional selectivity," (in eng), *Brain*, vol. 114 (Pt 5), pp. 2283-301, Oct 1991, doi: 10.1093/brain/114.5.2283.
- [5] G. W. Ross, R. D. Abbott, H. Petrovitch, C. M. Tanner, and L. R. White, "Pre-motor features of Parkinson's disease: the Honolulu-Asia Aging Study experience," (in eng), *Parkinsonism Relat Disord*, vol. 18 Suppl 1, pp. S199-202, Jan 2012, doi: 10.1016/s1353-8020(11)70062-1.
- [6] P. Rizek, N. Kumar, and M. S. Jog, "An update on the diagnosis and treatment of Parkinson disease," (in eng), *Cmaj*, vol. 188, no. 16, pp. 1157-1165, Nov 1 2016, doi: 10.1503/cmaj.151179.
- [7] O. Suchowersky, S. Reich, J. Perlmutter, T. Zesiewicz, G. Gronseth, and W. J. Weiner, "Practice Parameter: diagnosis and prognosis of new onset Parkinson disease (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology," (in eng), *Neurology*, vol. 66, no. 7, pp. 968-75, Apr 11 2006, doi: 10.1212/01.wnl.0000215437.80053.d0.
- [8] A. E. Lang and A. M. Lozano, "Parkinson's disease. Second of two parts," (in eng), *N Engl J Med*, vol. 339, no. 16, pp. 1130-43, Oct 15 1998, doi: 10.1056/nejm199810153391607.
- [9] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," (in eng), *Biometrics*, vol. 44, no. 3, pp. 837-45, Sep 1988.

- [10] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, "Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis," (in eng), *Neurology*, vol. 86, no. 6, pp. 566-76, Feb 9 2016, doi: 10.1212/wnl.0000000000002350.
- [11] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," (in eng), *IEEE Trans Biomed Eng*, vol. 57, no. 4, pp. 884-93, Apr 2010, doi: 10.1109/tbme.2009.2036000.
- [12] A. Schrag, Y. Ben-Shlomo, and N. Quinn, "How valid is the clinical diagnosis of Parkinson's disease in the community?," (in eng), *J Neurol Neurosurg Psychiatry*, vol. 73, no. 5, pp. 529-34, Nov 2002, doi: 10.1136/jnnp.73.5.529.
- [13] B. Harel, M. Cannizzaro, and P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study," (in eng), *Brain Cogn*, vol. 56, no. 1, pp. 24-9, Oct 2004, doi: 10.1016/j.bandc.2004.05.002.
- [14] W. Maetzler, I. Liepelt, and D. Berg, "Progression of Parkinson's disease in the clinical phase: potential markers," (in eng), *Lancet Neurol*, vol. 8, no. 12, pp. 1158-71, Dec 2009, doi: 10.1016/s1474-4422(09)70291-1.
- [15] L. O. Ramig, C. Fox, and S. Sapir, "Speech treatment for Parkinson's disease," (in eng), *Expert Rev Neurother*, vol. 8, no. 2, pp. 297-309, Feb 2008, doi: 10.1586/14737175.8.2.297.
- [16] S. Skodda, "Effect of deep brain stimulation on speech performance in Parkinson's disease," (in eng), *Parkinsons Dis*, vol. 2012, p. 850596, 2012, doi: 10.1155/2012/850596.
- [17] L. Naranjo, C. J. Pérez, and J. Martín, "Addressing voice recording replications for tracking Parkinson's disease progression," (in eng), *Med Biol Eng Comput*, vol. 55, no. 3, pp. 365-373, Mar 2017, doi: 10.1007/s11517-016-1512-y.
- [18] L. Naranjo, C. J. Pérez, J. Martín, and Y. Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications," (in eng), *Comput Methods Programs Biomed*, vol. 142, pp. 147-156, Apr 2017, doi: 10.1016/j.cmpb.2017.02.019.
- [19] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," presented at the Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017.
- [20] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367-378, 2002/02/28/ 2002, doi: [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- [21] C. G. Goetz *et al.*, "Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device," (in eng), *Mov Disord*, vol. 24, no. 4, pp. 551-6, Mar 15 2009, doi: 10.1002/mds.22379.

- [22] Y. L. Shue, P. Keating, and C. Vicenik, "VOICESAUCE: A program for voice analysis," *The Journal of the Acoustical Society of America*, vol. 126, no. 4, pp. 2221-2221, 2009/10/01 2009, doi: 10.1121/1.3248865.
- [23] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," (in eng), *J R Soc Interface*, vol. 8, no. 59, pp. 842-55, Jun 6 2011, doi: 10.1098/rsif.2010.0456.
- [24] M. A. Horning, J. Y. Shin, L. A. DiFusco, M. Norton, and B. Habermann, "Symptom progression in advanced Parkinson's disease: Dyadic perspectives," (in eng), *Appl Nurs Res*, vol. 50, p. 151193, Dec 2019, doi: 10.1016/j.apnr.2019.151193.
- [25] I. Rektorova *et al.*, "Speech prosody impairment predicts cognitive decline in Parkinson's disease," (in eng), *Parkinsonism Relat Disord*, vol. 29, pp. 90-5, Aug 2016, doi: 10.1016/j.parkreldis.2016.05.018.
- [26] I. Suttrup and T. Warnecke, "Dysphagia in Parkinson's Disease," (in eng), *Dysphagia*, vol. 31, no. 1, pp. 24-32, Feb 2016, doi: 10.1007/s00455-015-9671-9.
- [27] X. Chen *et al.*, "Sensorimotor control of vocal pitch production in Parkinson's disease," (in eng), *Brain Res*, vol. 1527, pp. 99-107, Aug 21 2013, doi: 10.1016/j.brainres.2013.06.030.
- [28] L. K. Bowen, G. L. Hands, S. Pradhan, and C. E. Stepp, "Effects of Parkinson's Disease on Fundamental Frequency Variability in Running Speech," (in eng), *J Med Speech Lang Pathol*, vol. 21, no. 3, pp. 235-244, Sep 2013.
- [29] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. [Online]. Available: www.jstor.org/stable/2699986.
- [30] C. Leistner, A. Saffari, P. M. Roth, and H. Bischof, "On robustness of on-line boosting - a competitive study," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 27 Sept.-4 Oct. 2009 2009, pp. 1362-1369, doi: 10.1109/ICCVW.2009.5457451.
- [31] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016. [Online]. Available: <https://doi.org/10.1145/2939672.2939785>.

Figures

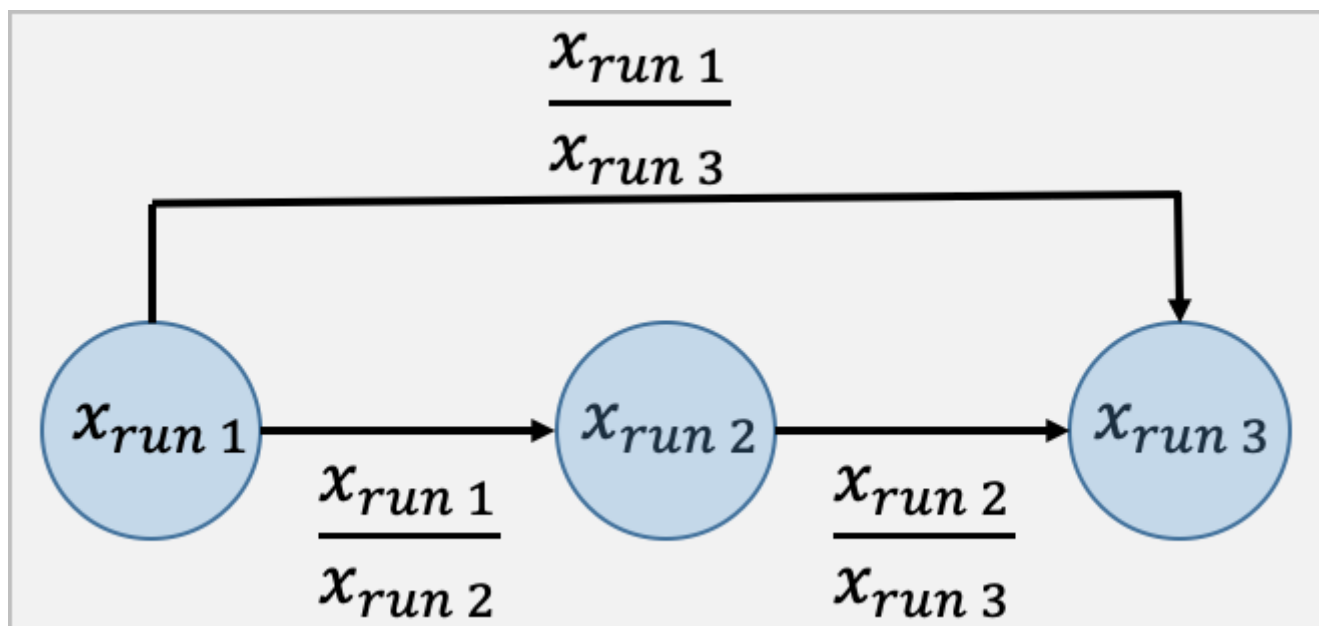


Figure 1

Acoustic features used in modeling



Figure 2

Feature selection and reclassification results for 4-fold cross validation using the LGB model