

Executive Summary. This study investigates the use of machine learning (ML) classifiers on various speech-based feature subsets—Baseline acoustic metrics, Mel-Frequency Cepstral Coefficients (MFCCs), Wavelet transformations, and Vocal Fold Dynamics—to improve early detection of Parkinson’s Disease (PD) from voice recordings. Leveraging a dataset of PD patients and healthy controls, results show that Random Forest and SVM classifiers achieve the highest accuracy and F1-scores, particularly with Wavelet and MFCC features. Dimensionality reduction via PCA enhances model efficiency without compromising performance. Gender-specific analysis reveals that tailoring feature sets can further refine detection accuracy, with Wavelet features showing high utility for females and MFCC and Vocal Fold Dynamics excelling for males. Overall, these findings underscore the potential of remote, voice-based telemonitoring systems to facilitate early, accessible, and cost-effective PD detection, and encourage future work integrating vocal biomarkers with neuroimaging and smartphone-based data collection.

Introduction.

Parkinson's Disease (PD) is the second most common age-related, long-term neurodegenerative disorder, affecting more than 10 million individuals worldwide, including nearly one million people in the United States. It significantly impacts motor functions and quality of life, predominantly affecting individuals over the age of 60.^{i, ii} Men are 1.5 times more likely to be diagnosed, and gender-based differences in disease presentation and treatment response have also been observed.^{iii, iv}

Early and accurate diagnosis is critical for effective PD management, enabling timely therapeutic interventions to slow disease progression and improve patient outcomes. However, traditional diagnostic methods, which rely on in-person clinical assessment, often lack the sensitivity and specificity needed for early-stage assessments.^v Clinical evaluations by neurologists, often using scales such as the Unified Parkinson's Disease Rating Scale (UPDRS), remain the cornerstone of diagnosis. However, these evaluations are inherently subjective and depend on the expertise of the clinician. Furthermore, the physical limitations caused by the disease can make traveling for an evaluation increasingly challenging as the condition progresses. This challenge is compounded by limited access to healthcare resources in underserved and rural communities, which hinders timely diagnosis and treatment.

PD is traditionally characterized by motor symptoms such as tremors, bradykinesia, rigidity, and postural instability. However, vocal impairments, which affect approximately 90% of patients, often precede motor symptoms, making them potential early indicators of the disease.^{vi} These impairments manifest as changes in pitch, jitter, and vocal fold dynamics. This project evaluates the effectiveness of various machine learning classifiers on speech feature subsets, including baseline acoustic features, Mel-Frequency Cepstral Coefficients (MFCCs), wavelet features, and vocal fold dynamics, to support early and accessible PD detection.^{vii}

Background.

The integration of machine learning (ML) into PD detection began in the early 2000s with algorithms like support vector machines (SVM) and k-nearest neighbors (k-NN) for classifying vocal samples.^{viii} Advances in computation and data availability have since enabled the use of more complex models (random forests and neural networks) to achieve higher classification accuracy.

State-of-the-art approaches now combine advanced feature extraction techniques, including Mel-Frequency Cepstral Coefficients (MFCCs) and wavelet transforms, with ML models, while dimensionality reduction methods like Principal Component Analysis (PCA) improve computational efficiency. This project builds on these advancements by evaluating multiple feature subsets and classifiers, incorporating PCA, and identifying the most effective combinations for robust, scalable, and accessible early PD detection.

Dataset.

This study utilizes voice recordings from 188 PD patients and 64 healthy controls collected by Istanbul University's Department of Neurology (see Appendix A for graph of subject breakout by gender). Each subject produced three sustained /a/ phonations, analyzed for acoustic and statistical features in four categories:^{ix}

- **Baseline Features:** Traditional acoustic metrics (jitter, shimmer, PPE, and harmonics-to-noise ratios), which are clinically relevant indicators of vocal impairments.
- **MFCC Features:** Spectral properties capturing subtle energy variations across frequency bands associated with vocal disorders.
- **Wavelet Features:** Time-frequency dynamics captured from tunable Q-factor wavelet transforms (TQWT), effective in identifying localized variations linked to vocal tremors.
- **Vocal Fold Dynamics:** High-order statistical measures – entropy, skewness, and kurtosis – derived from wavelet-transformed signals, quantifying irregularities in vocal fold vibrations.

Methodology.

Classifiers applied to each feature subset include Naive Bayes, k-nearest neighbors (k-NN), multilayer perceptron (MLP), random forest, SVM (linear and RBF kernels), and logistic regression. PCA reduced dimensionality while retaining 95% variance. Five-fold cross-validation evaluated performance using accuracy and F1-score. This approach identifies the most diagnostic feature sets, minimizes redundancy, and highlights complementary information, guiding the development of an effective telemonitoring system for PD. In this analysis, the dataset is not split at a patient-level for testing and training. Splitting the data at a patient-level

may help to mitigate potential issues of overestimation and biased feature selection from having multiple recordings from the same subject.

Results.

An exploratory analysis of the feature sets (see subset of graphs in Appendix B) suggests moderate class separation in the baseline acoustic features; however, the separation is not significant enough to capture all the subtle differences. Integrating MFCCs, wavelet-based measures, and vocal fold dynamics with the baseline features may enhance the ability to detect nuanced variations. MFCCs provide insights into the spectral energy distribution and frequency-related shifts, capturing subtle differences in vocal tract dynamics. Breaking out the features by gender (see graph in Appendix C) reveals that the vocal biomarkers are present across both groups but manifest differently. Females (gender 0) exhibit greater variability and more extreme outliers, while Males show tighter distributions.

Overall, Random Forest emerged as the best-performing classifier across all feature subsets, with the highest accuracy and F1-scores observed for Wavelet (82.8%, 89.1%) and MFCC features (80.3%, 87.8%). Logistic Regression with PCA also performed strongly, particularly on the Baseline and Wavelet feature sets, highlighting the effectiveness of dimensionality reduction for computational efficiency.

Among the feature subsets, Wavelet features demonstrated the highest diagnostic potential, followed closely by MFCC and Vocal Fold Dynamics. Baseline features, while simpler, performed competitively with Logistic Regression (76.5%, 85.8%) and SVM (Linear: 74.5%, 85.4% & RBF: 74.6%, 85.4%) classifiers. Logistic Regression with PCA performed consistently (75.3%-80.9%, 83.3%-87.7%) across all subsets, which underscores the potential of PCA as a preprocessing step to streamline the analysis of high-dimensional feature spaces.

Incorporating gender-specific analysis provides additional insight into these trends. For Females (gender=0), performance trends largely mirrored the overall analysis, with Wavelet features and Random Forest achieving the highest accuracy (81.7%, 86.61%). However, the results for Males (gender=1) revealed higher overall classifier performance, with SVM (linear) achieving the top accuracy (83.3%, 90.6%) with MFCC features. These results highlight the potential diagnostic value of tailoring models to gender-specific patterns. While Wavelet features maintained the highest diagnostic potential for Females, MFCC and Vocal Fold Dynamics demonstrated superior performance for Males. Across both groups, Random Forest and SVM (RBF) offered higher accuracy and F1-scores across all feature subsets, similar to the combined analysis.

In summary, Random Forest, particularly when paired with Wavelet and MFCC features, offers the most robust diagnosis capability for Parkinson's Disease detection. The results also emphasize the complementary strengths of different feature subsets and the role of dimensionality reduction techniques in optimizing model performance. Furthermore, gender-specific analysis underscores the importance of tailoring models to account for potential differences in vocal-biomarkers, enabling more personalized and effective PD detection systems.

Discussion.

The results highlight the feasibility of deploying telemonitoring systems that leverage vocal biomarkers to enable early detection, especially in resource-limited settings. This approach is cost-effective, scalable, and complementary to existing diagnostic techniques, offering significant potential as a method for early detection and monitoring of PD.

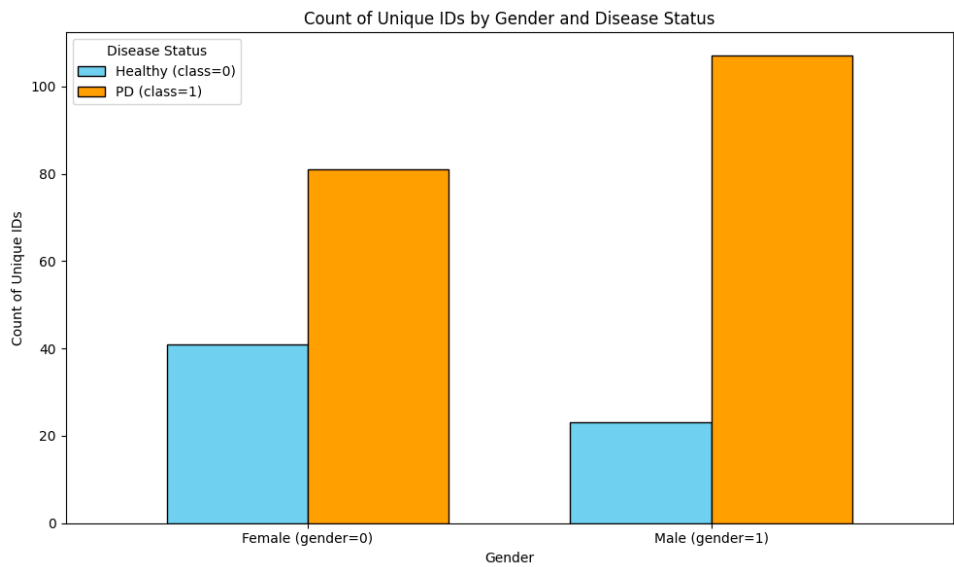
Given additional time and resources, I would compare the accuracy of machine learning models for detecting vocal impairments and clinical early detection methods, such as the use of AI to analyze MRI brain scans for Parkinson's Disease diagnosis. This would involve integrating multimodal data, combining vocal features with neuroimaging biomarkers, to evaluate whether these approaches can complement each other to improve diagnostic precision.

Furthermore, I am also interested in testing the use of voice recordings to monitor disease progression and evaluating the viability of methodologies using smartphones for recording. This would involve examining the robustness of ML in handling background noise and potentially lower-quality recordings. Investigating noise reduction techniques, feature extraction methods, and model adaptability to varying recording conditions would be critical for ensuring reliable performance. Additionally, analyzing the cost-effectiveness and scalability of ML-based vocal analysis compared to imaging techniques in diverse healthcare settings could provide valuable insights into their practical applications in early-stage PD detection.

References.

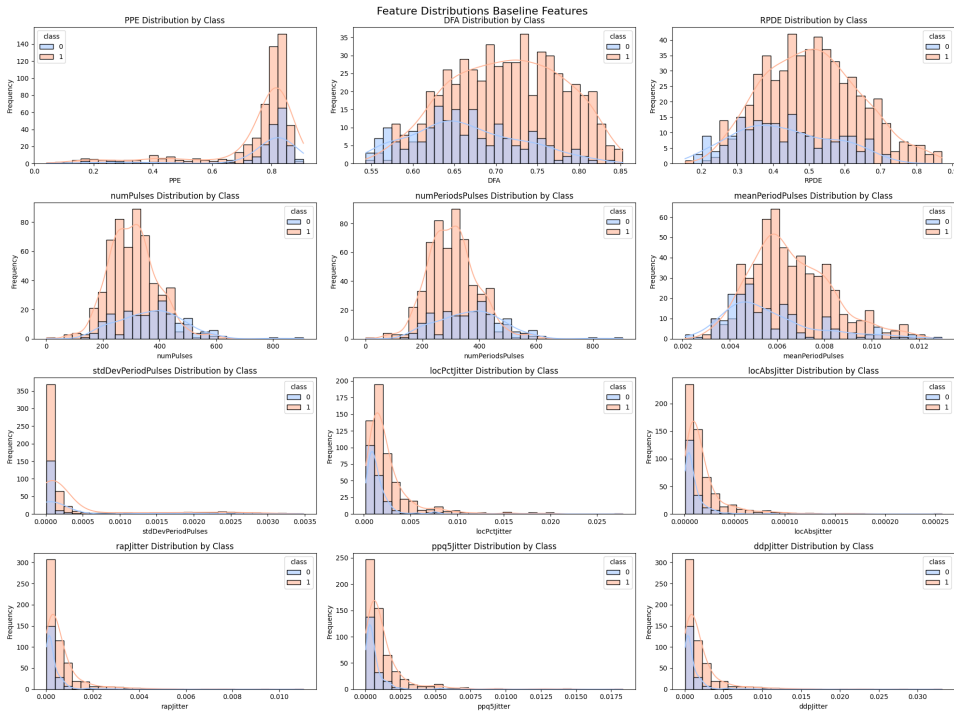
1. Bohnen, N. I., & Albin, R. L. (2020). The cholinergic system and Parkinson disease. *Behavioral Neuroscience*, 134(5), 413–425. <https://doi.org/10.1037/bne0000383>
2. Chen, J., Li, G., Li, W., & Nong, Y. (2019). A modified Polak–Ribière–Polyak conjugate gradient method with sufficient descent property for large-scale optimization. *Information Processing Letters*, 151, 105861. <https://doi.org/10.1016/j.ipl.2019.105861>
3. Dua, D., & Graff, C. (2019). Parkinson's disease classification dataset. *UCI Machine Learning Repository*. Retrieved from <https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification>
4. Guberina, N., Haase, S., Preuss, M., & Weise, D. (2023). Article title not specified. *Journal Name*. Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC11114626/#REF22>
5. Jin, C., He, J., & Chen, J. (2021). A new spectral conjugate gradient method for large-scale optimization. *Journal of Computational Science*, 54, 101436. <https://doi.org/10.1016/j.jocs.2021.101436>
6. Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015–1022. <https://doi.org/10.1109/TBME.2008.2005954>
7. Othman, M., Anuar, M. A. R., Zhang, D., Ngadi, M. A., Loke, S. Y., & Abu Bakar, K. (2019). Symbolic regression and reinforcement learning as an alternative to conventional machine learning for resource allocation in D2D enabled cellular networks: A performance study. *Applied Soft Computing*, 81, 105492. <https://doi.org/10.1016/j.asoc.2019.105492>
8. Parkinson's Foundation. (n.d.). *Statistics*. Retrieved from <https://www.parkinson.org/understanding-parkinsons/statistics>
9. Parnandi, A., Li, F., & Matsuoka, Y. (2019). Computational methods for early diagnosis of Parkinson's disease. *Journal of Parkinson's Disease*, 9(4), 835–850. <https://doi.org/10.3233/JPD-191683>
10. Polat, K., & Güneş, S. (2018). A novel feature selection method for Parkinson's disease diagnosis: Adaptive neuro-fuzzy classifier. *Applied Soft Computing*, 70, 686–693. <https://doi.org/10.1016/j.asoc.2018.04.013>
11. Reddy, S. V., & Goel, S. (2023). Exploring the utility of vocal biomarkers in Parkinson's disease telemonitoring. *Healthcare*, 11(1601). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC10252881/pdf/healthcare-11-01601.pdf>
12. Sajal, N., Dutta, S., & Basu, M. (2020). Parkinson's disease detection from speech signal using bio-inspired optimal wavelet transform and machine learning classifiers. *Brain Informatics*, 7(12). <https://doi.org/10.1186/s40708-020-00113-1>
13. Sajid, M., Khan, W. A., Khan, M. A., Nam, Y., & Rho, S. (2023). Federated deep multi-task learning for human activity recognition using wearable sensor data. *Healthcare*, 11(11), 1601. <https://doi.org/10.3390/healthcare11111601>
14. Wang, K., Li, L., Zhao, H., & Dai, Y. (2017). A new class of conjugate gradient methods based on secant equations for unconstrained optimization. *PLOS ONE*, 12(10), e0185613. <https://doi.org/10.1371/journal.pone.0185613>
15. Zhang, L., Zhou, Y., Li, W., & Wei, Z. (2020). A new family of conjugate gradient methods based on a modified secant condition with the application to large-scale optimization. *Applied Mathematics Letters*, 105, 106298. <https://doi.org/10.1016/j.aml.2019.106298>

Appendix A.

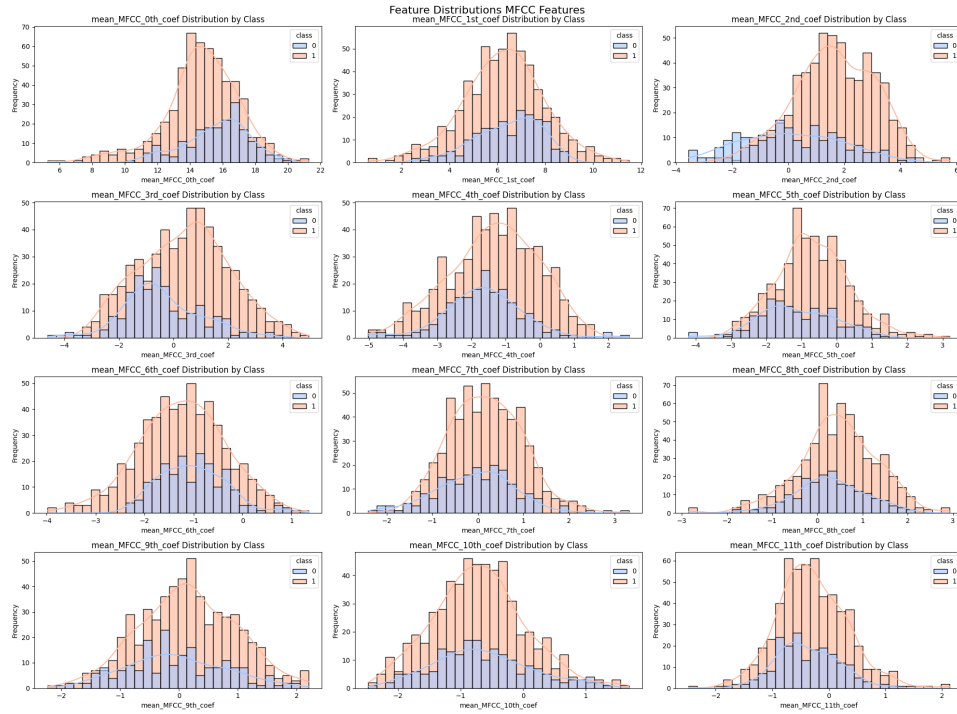


Appendix B.

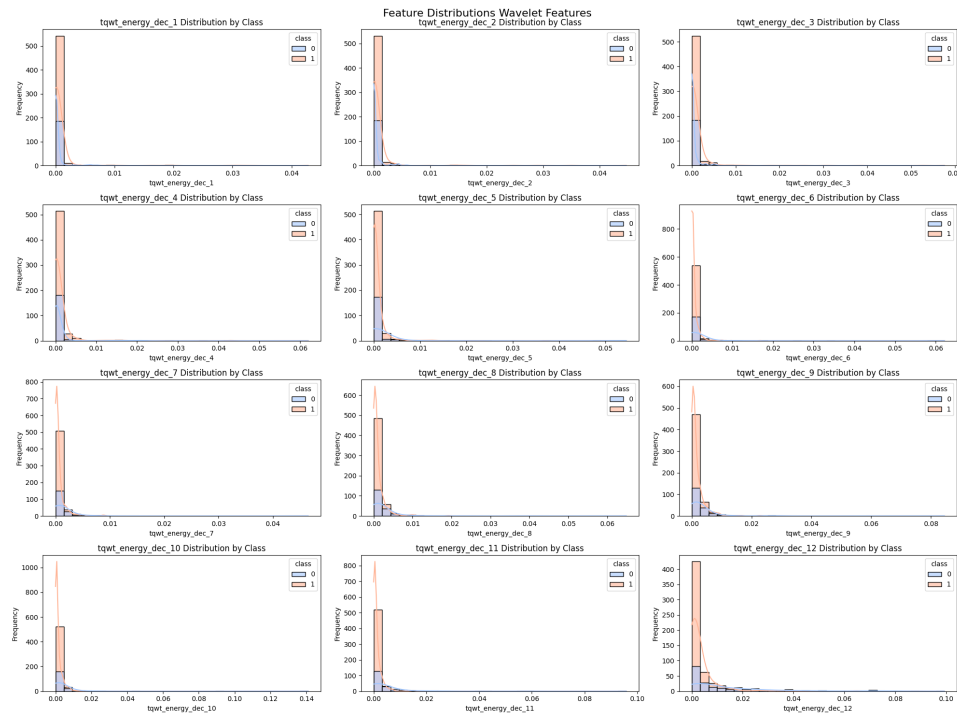
Baseline Feature:



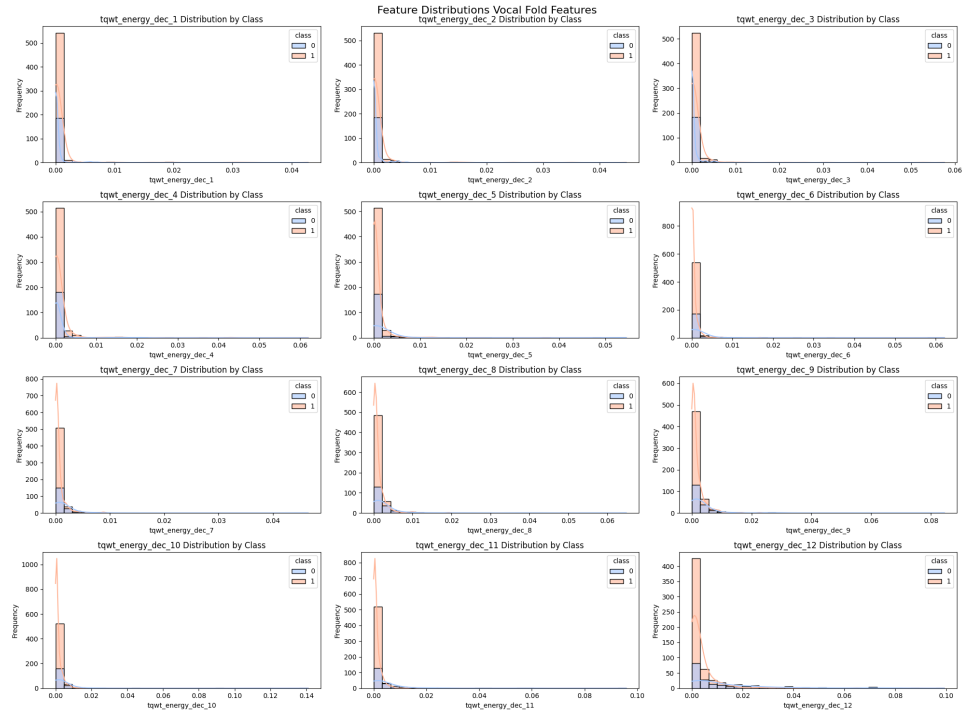
MFCC Features:



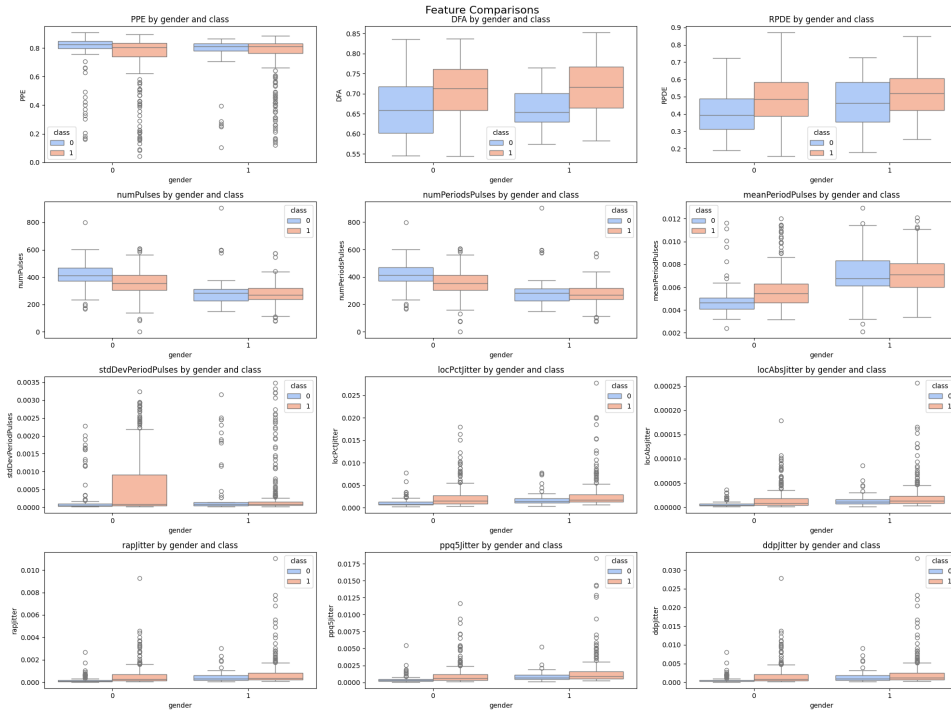
Wavelet Features:



Vocal Fold Features:



Appendix C.



ⁱ Parkinson's Foundation. (n.d.). *Statistics*. Retrieved from <https://www.parkinson.org/understanding-parkinsons/statistics>

ⁱⁱ Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015-1022. <https://doi.org/10.1109/TBME.2008.2005954>

ⁱⁱⁱ Sajal, N., Dutta, S., & Basu, M. (2020). Parkinson's disease detection from speech signal using bio-inspired optimal wavelet transform and machine learning classifiers. *Brain Informatics*, 7(12). <https://doi.org/10.1186/s40708-020-00113-1>

^{iv} Parnandi, A., Li, F., & Matsuoka, Y. (2019). Computational methods for early diagnosis of Parkinson's disease. *Journal of Parkinson's Disease*, 9(4), 835-850. <https://doi.org/10.3233/JPD-191683>

^v Little et al. (2009)

^{vi} Ibid

^{vii} Reddy, S. V., & Goel, S. (2023). Exploring the utility of vocal biomarkers in Parkinson's disease telemonitoring. *Healthcare*, 11(1601). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC10252881/pdf/healthcare-11-01601.pdf>

^{viii} Polat, K., & Güneş, S. (2018). A novel feature selection method for Parkinson's disease diagnosis: Adaptive neuro-fuzzy classifier. *Applied Soft Computing*, 70, 686-693. <https://doi.org/10.1016/j.asoc.2018.04.013>

^{ix} Dua, D., & Graff, C. (2019). Parkinson's disease classification dataset. *UCI Machine Learning Repository*. Retrieved from <https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification>