FULL LENGTH ARTICLE · Articles in Press, November 21, 2025

Vocal Biomarkers for Parkinson's Disease Classification through Hybrid Feature Selection with Beluga Whale Optimization

Umashankar Ganesan * · Lumen Christy V † · Sheeba Santhosh ‡ · Hemalatha R. J §

Affiliations & Notes Article Info

SUMMARY

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder that often causes vocal impairments. As a result, voice analysis is a promising avenue for early and objective diagnosis. While voice is a well-established biomarker, the high dimensionality of acoustic feature sets can lead to model overfitting. This necessitates effective feature selection strategies.

Objective

The objective is to validate a hybrid feature selection pipeline that combines statistical filters with a metaheuristic wrapper algorithm. The aim is to identify a

Methods

This study validates a hybrid feature selection pipeline. It combines statistical filters such as Spearman correlation (Corr) and mutual information (MI) with a metaheuristic beluga whale optimization (BWO) wrapper. The framework's efficacy was benchmarked across a diverse suite of machine learning classifiers. Evaluation was conducted using 10-fold stratified group cross-validation on a dataset of 40 PD patients and 40 healthy controls. The process dynamically identifies parsimonious feature subsets within each validation fold.

Results

When paired with a random forest classifier, the Corr+MI+BWO pipeline achieved a peak F1-score of 0.966 using an average of only 21.2 features. This represents a 73% reduction in dimensionality with a minimal 2.4% performance trade-off compared to a standard filter-based approach. The most consistently selected features proved to be both statistically significant ($P < 0.05$) and part of a clinically coherent signature of Parkinsonian dysarthria.

Conclusion

The proposed hybrid framework is an effective methodology that successfully balances model simplicity with high predictive accuracy. This research provides a strong foundation for the development of objective, noninvasive tools for the early detection and monitoring of PD.

Data availability

This study utilized a publicly available dataset of voice recordings from individuals with Parkinson's disease and healthy controls obtained from Figshare, available at the following link: https://figshare.com/articles/dataset/Voice_Samples_for_Patients_with_Parkinson_s_Disease_and_Healthy_Controls/23849127

Keywords

Parkinson's disease · Beluga whale optimization · Machine learning · Voice biomarkers · Hypokinetic dysarthria

References

1. Klempíř, O. · Krupička, R.
Analyzing Wav2Vec 1.0 embeddings for cross-database Parkinson's disease detection and speech features extraction
Sensors. 2024; **24**:5520
Google Scholar

2. Faragó, P. · Ștefănigă, S.-A. · Cordoș, C.-G. …
CNN-based identification of Parkinson's disease from continuous speech in noisy environments
Bioengineering. 2023; **10**:531
Google Scholar

3. Doumari, S.A. · Berahmand, K. · Ebadi, M.J.
Early and high-accuracy diagnosis of Parkinson's disease: outcomes of a new model
Comput Math Methods Med. 2023; **2023**, 1493676
Google Scholar

4. Alshammri, R. · Alharbi, G. · Alharbi, E. …
Machine learning approaches to identify Parkinson's disease using voice signal features
Front Artif Intell. 2023; **6**, 1084001
Google Scholar

5. Swain, K.P. · Samal, S.R. · Ravi, V. …

Google Scholar

6. Luna-Ortiz, I. · Aldape-Pérez, M. · Uriarte-Arcia, A.V. …

Parkinson's disease detection from voice recordings using associative memories

Healthcare. 2023; **11**:1601

Google Scholar

7. Shastry, K.A.

An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease

Healthc Anal. 2023; **3**, 100181

Google Scholar

8. Müller, F. · Nienstedt, J.C. · Buhmann, C. …

Effect of subthalamic and nigral deep brain stimulation on speech and voice in Parkinson's patients

J Neural Transm. 2025; **132**:419-429

Google Scholar

9. Zhao, S. · Dai, G. · Li, J. …

An interpretable model based on graph learning for diagnosis of Parkinson's disease with voice-related EEG

NPJ Digit Med. 2024; **7**:3

Google Scholar

10. dos Santos, V.B. · Ravazio, R. · Teixeira-dos-Santos, D. …

Speech subtypes are associated with worsened tremor and axial symptoms in Parkinson's disease patients

Clin Parkinsonism Relat Disord. 2025; **13**, 100373

Google Scholar

11. Xie, J.C. · Gan, Y. · Liang, P.

Google Scholar

12. Anter, A.M. · Mohamed, A.W. · Zhang, M. …

A robust intelligence regression model for monitoring Parkinson's disease based on speech signals

Future Gener Comput Syst. 2023; **147**:316-327

Google Scholar

13. Diao, Y. · Xie, H. · Wang, Y. …

Acoustic assessment in mandarin speaking Parkinson's disease patients and disease progression monitoring and brain impairment within the speech subsystem

NPJ Parkinsons Dis. 2024; **10**:115

Google Scholar

14. Mishra, S. · Jena, L. · Mishra, N. …

PD-DETECTOR: a sustainable and computationally intelligent mobile application model for Parkinson's disease severity assessment

Heliyon. 2024; **10**, e34593

Google Scholar

15. Wen, P. · Zhang, Y. · Wen, G.

Intelligent personalized diagnosis modeling in advanced medical system for Parkinson's disease using voice signals

Math Biosci Eng. 2023; **20**:8085-8102

Google Scholar

16. Kadhim, M.N. · Al-Shammary, D. · Sufi, F.

A novel voice classification based on gower distance for Parkinson disease detection

Int J Med Inform. 2024; **191**, 105583

Google Scholar

Comput Mater Continua. 2022; **71**:5639-5658

Google Scholar

18. Soumaya, Z. · Taoufiq, B.D. · Benayad, N. …

The detection of Parkinson disease using the genetic algorithm and SVM classifier

Appl Acoust. 2021; **171**, 107528

Google Scholar

19. Oliveira, G.C. · Paha, N.D. · Ngo, Q.C. …

A pilot study for speech assessment to detect the severity of Parkinson's disease: an ensemble approach

Comput Biol Med. 2025; **185**, 109565

Google Scholar

20. Iyer, A. · Kemp, A. · Rahmatallah, Y. …

A machine learning method to process voice samples for identification of Parkinson's disease

Sci Rep. 2023; **13**, 20615

Google Scholar

21. Hussain, M.M. · Weslin, D. · Kumari, S. …

Enhancing Parkinson's disease identification using ensemble classifier and data augmentation techniques in machine learning

Clin Ehealth. 2023; **6**:150-158

Google Scholar

22. Sadiq, M. · Khan, M.T. · Masood, S.

Attention-based deep learning model for early detection of Parkinson's disease

Comput Mater Continua. 2022; **71**:5183-5200

Google Scholar

J Voice. 2024; **38**:967.e9-967.e17

Google Scholar

24. Li, R. · Huang, G. · Wang, X. …

Smartphone automated motor and speech analysis for early detection of Alzheimer's disease and Parkinson's disease: validation of TapTalk across 20 different devices

Alzheimers Dement. 2024; **16**, e70025

Google Scholar

25. Bottalico, P. · Codino, J. · Cantor-Cutiva, L.C. …

Reproducibility of voice parameters: the effect of room acoustics and microphones

J Voice. 2020; **34**:320-334

Google Scholar

26. Kodali, M. · Kadiri, S.R. · Alku, P.

Automatic classification of the severity level of Parkinson's disease: a comparison of speaking tasks, features, and classifiers

Comput Speech Lang. 2024; **83**, 101548

Google Scholar

27. Majda-Zdancewicz, E. · Potulska-Chromik, A. · Nojszewska, M. …

Speech signal analysis in patients with Parkinson's disease, taking into account phonation, articulation, and prosody of speech

Appl Sci. 2024; **14**:11085

Google Scholar

28. Almasoud, A.S. · Eisa, T.A.E. · Al-Wesabi, F.N. …

Parkinson's detection using RNN-Graph-LSTM with optimization based on speech signals

Comput Mater Continua. 2022; **72**:871-886

Google Scholar

Voice biomarkers as prognostic indicators for Parkinson's disease using machine learning techniques

Sci Rep. 2025; **15**, 12129

Google Scholar

30. Srinivasan, S. · Ramadass, P. · Mathivanan, S.K. …

Detection of Parkinson disease using multiclass machine learning approach

Sci Rep. 2024; **14**, 13813

Google Scholar

31. Contreras-Ruston, F. · Castillo-Allendes, A. · Saavedra-Garrido, J. …

Voice self-assessment in individuals with Parkinson's disease as compared to general voice disorders

Parkinsonism Relat Disord. 2024; **123**, 106944

Google Scholar

32. Favaro, A. · Tsai, Y.-T. · Butala, A. …

Interpretable speech features vs. DNN embeddings: what to use in the automatic assessment of Parkinson's disease in multi-lingual scenarios

Comput Biol Med. 2023; **166**, 107559

Google Scholar

33. Costantini, G. · Cesarini, V. · Di Leo, P. …

Artificial intelligence-based voice assessment of patients with Parkinson's disease off and on treatment: machine vs. deep-learning comparison

Sensors. 2023; **23**:2293

Google Scholar

34. Ali, A.M. · Salim, F. · Saeed, F.

Parkinson's disease detection using filter feature selection and a genetic algorithm with ensemble learning

Diagnostics. 2023; **13**:2816

Beluga whale optimization: a novel nature-inspired metaheuristic algorithm

Knowl Based Syst. 2022; **251**, 109215

Google Scholar

36. Hossain, M.A. · Amenta, F.

Machine learning-based classification of Parkinson's disease patients using speech biomarkers

J Parkinsons Dis. 2024; **14**:95-109

Google Scholar

37. Kumar, A. · Singh, J.P. · Paygude, P. …

Advanced comparative analysis of machine learning algorithms for early Parkinson's disease detection using vocal biomarkers

Digit Health. 2025; **11**:1-13

Google Scholar

38. Malekroodi, H.S. · Madusanka, N. · Lee, B.-i …

Leveraging deep learning for fine-grained categorization of Parkinson's disease progression levels through analysis of vocal acoustic patterns

Bioengineering. 2024; **11**:295

Google Scholar

39. Al Mudawi, N.

Developing a model for Parkinson's disease detection using machine learning algorithms

Comput Mater Continua. 2024; **79**:4945-4962

Google Scholar

40. Valarmathi, P. · Suganya, Y. · Saranya, K.R. …

Enhancing parkinson disease detection through feature based deep learning with autoencoders and neural networks

Sci Rep. 2025; **15**:8624

Analysis of voice, speech, and language biomarkers of Parkinson's disease collected in a mixed reality setting

Sensors. 2025; **25**:2405

Google Scholar

42. Roland, V. · Huet, K. · Harmegnies, B. …

Vowel production: a potential speech biomarker for early detection of dysarthria in Parkinson's disease

Front Psychol. 2023; **14**, 1129830

Google Scholar

43. Shen, M. · Mortezaagha, P. · Rahgozar, A.

Explainable artificial intelligence to diagnose early Parkinson's disease via voice analysis

Sci Rep. 2025; **15**, 11687

Google Scholar

44. Alalayah, K.M. · Senan, E.M. · Atlam, H.F. …

Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method

Diagnostics. 2023; **13**:1924

Google Scholar

45. Kadhim, M.N. · Al-Shammary, D. · Mahdi, A.M. …

Feature selection based on Mahalanobis distance for early Parkinson disease classification

Comput Methods Programs Biomed Update. 2025; **7**, 100177

Google Scholar

46. Ali, L. · Javeed, A. · Noor, A. …

Parkinson's disease detection based on features refinement through L1 regularized SVM and deep neural network

Sci Rep. 2024; **14**:1333

47. Santhosh, K. · Dev, P.P. · Jose, A.B. …

A modified Gray Wolf Optimization algorithm for early detection of Parkinson's disease

Biomed Signal Process Control. 2025; **109**, 108061

Google Scholar

48. Rana, A. · Dumka, A. · Singh, R. …

An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features

Electronics. 2022; **11**:3782

Google Scholar

49. Ahmed, S.Z. · Khan, M.U.

Feature-based voice analysis for Parkinson's prediction using ML classifiers

Int J Multidiscip Res Growth Eval. 2025; **6**:1597-1605

Google Scholar

50. Bashir, A. · Singh, Y. · Zadoo, S. …

Parkinson's disease detection: a machine learning based model

Procedia Comput Sci. 2025; **259**:1783-1792

Google Scholar

51. Deepa, P. · Khilar, R.

Parkinson's disease detection and classification: leveraging voice features and ensemble methods with feature selection and ERT classifier

Procedia Comput Sci. 2024; **235**:1695-1706

Google Scholar

52. Escobar-Grisales, D. · Ríos-Urrego, C.D. · Orozco-Arroyave, J.R.

Deep learning and artificial intelligence applied to model speech and language in Parkinson's disease

Diagnostics. 2023; **13**:2162

53. Yang, S. · Wang, F. · Yang, L. …

The physical significance of acoustic parameters and its clinical significance of dysarthria in Parkinson's disease

Sci Rep. 2020; **10**, 11776

Google Scholar

54. Rahmatallah, Y. · Kemp, A.S. · Iyer, A. …

Pre-trained convolutional neural networks identify Parkinson's disease from spectrogram images of voice samples

Sci Rep. 2025; **15**:7337

Google Scholar

55. Wang, M. · Zhao, X. · Li, F. …

Using sustained vowels to identify patients with mild Parkinson's disease in a Chinese dataset

Front Aging Neurosci. 2024; **16**, 1377442

Google Scholar

56. Mahesh, T.R. · Kumar, V.V. · Bhardwaj, R. …

An artificial intelligence-based decision support system for early and accurate diagnosis of Parkinson's disease

Decis Anal J. 2024; **10**, 100381

Google Scholar

57. Warule, P. · Mishra, S.P. · Deb, S.

Time-frequency analysis of speech signal using Chirplet transform for automatic diagnosis of Parkinson's disease

Biomed Eng Lett. 2023; **13**:613-623

Google Scholar

58. Ahmadi, H. · Huo, L. · Arji, G. …

Early diagnosis of Parkinson's disease using a hybrid method of least squares support vector regression and fuzzy clustering

Biocybern Biomed Eng. 2024; **44**:569-585

59. Tougui, I. · Zakroum, M. · Karrakchou, O. ...

Transformer-based transfer learning on self-reported voice recordings for Parkinson's disease diagnosis

Sci Rep. 2024; **14**, 30131

Google Scholar

60. Di Cesare, M.G. · Perpetuini, D. · Cardone, D. ...

Machine learning-assisted speech analysis for early detection of Parkinson's disease: a study on speaker diarization and classification techniques

Sensors. 2024; **24**:1499

Google Scholar

61. Al-Najjar, H. · Al-Rousan, N. · Al-Najjar, D.

Hybrid grey wolf and whale optimization for enhanced Parkinson's prediction based on machine learning models using biomedical sound

Inform Med Unlocked. 2024; **48**, 101524

Google Scholar

62. Singh, S. · Xu, W.

Robust detection of Parkinson's disease using harvested smartphone voice data: a telemedicine approach

Telemed J E Health. 2020; **26**:327-334

Google Scholar

63. Mills, J. · Duffy, O. · Pedlow, K. ...

Exploring the perceptions of voice-assisted technology as a tool for speech and voice difficulties: focus group study among people with Parkinson disease and their carers

JMIR Rehabil Assist Technol. 2025; **12**, e75316

Google Scholar

64. Penttilä, N. · Tavi, L. · Hyppönen, M. ...

Prosodic features in Finnish-speaking adults with Parkinson's disease

65. Solana-Lavalle, G. · Rosas-Romero, R.
Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation
Biomed Signal Process Control. 2021; **66**, 102415
Google Scholar

66. Houle, N. · Feaster, T. · Mira, A. ...
Sex differences in the speech of persons with and without Parkinson's disease
Am J Speech Lang Pathol. 2024; **33**:96-116
Google Scholar

67. Lim, W.S. · Chiu, S.-I. · Peng, P.-L. ...
A cross-language speech model for detection of Parkinson's disease
J Neural Transm. 2025; **132**:579-590
Google Scholar

68. Saleh, S. · Alouani, Z. · Cherradi, B. ...
AIoT-based embedded systems optimization using feature selection for Parkinson's disease diagnosis through speech disorders
Intell Based Med. 2024; **10**, 100184
Google Scholar

69. Capobianco Guido, R.
Wavelets behind the scenes: practical aspects, insights, and perspectives
Phys Rep. 2022; **985**:1-23
Google Scholar

70. Capobianco Guido, R.
A tutorial on signal energy and its applications
Neurocomputing. 2016; **170**:264-282

71. Capobianco Guido, R.

A tutorial review on entropy-based handcrafted feature extraction for information fusion

Inf Fusion. 2018; **41**:161-175

Google Scholar

Article metrics

29

Captures

PLUMX

View details

Related articles (40)

Enhancing Parkinson's Disease Detection by Combining SMOTE and Feature Selection for Improved Machine Learning Classification Using Voice Recordings

Jihad A. Qadir
Journal of Voice, December 16, 2025

Vocal Biomarkers for Parkinson’s Disease Classification Using Audio Spectrogram Transformers

Madusanka et al.
Journal of Voice, December 10, 2024

Classification of Infant Cry Based on Hybrid Audio Features and ResLSTM

Qiu et al.
Journal of Voice, September 20, 2024

Show more

View full text