ISSN: 2582-6751
Journal of Clinical Medical Research

ATHENAEUM SCIENTIFIC PUBLISHERS

Research Article

AI-Based Quality of Voice Analysis Models for Clinical Use, Insights of Quality of Models from 19 Parkinson's Disease Studies (2013-2023)

Pedersen M[1*], Meiner VG[2]

[1]Ear-Nose-Throat Specialist, Head and Neck Surgeon, Fellow of The Royal Society of Medicine UK, Danish Representative of the Union of European Phoniatricians, The Medical Center, Østergade 18, 1100 Copenhagen, Denmark

[2]Student, Scientist in Computer Science, IT-University of Copenhagen, Rued Langgaardsvej 7, 2300 Copenhagen Denmark

***Correspondence author:** Mette Pedersen, Ear-Nose-Throat Specialist, Head and Neck Surgeon, Fellow of The Royal Society of Medicine UK, Danish Representative of the Union of European Phoniatricians, Copenhagen, Denmark; Email: m.f.pedersen@dadlnet.dk

Citation: Pedersen M, et al. AI-Based Quality of Voice Analysis Models for Clinical Use, Insights of Quality of Models from 19 Parkinson's Disease Studies (2013-2023). Jour Clin Med Res. 2025;6(1):1-8.

https://doi.org/10.46889/JCMR.2025.6107

Received Date: 06-02-2025

Accepted Date: 08-03-2025

Published Date: 15-03-2025

Abstract

Voice analysis, powered by Artificial Intelligence (AI) and Machine Learning (ML), has emerged as a valuable tool for detecting and monitoring voice disorders. By identifying vocal biomarkers, AI-driven models can facilitate early diagnosis, track disease progression and support clinical decision-making. This study systematically evaluates the effectiveness and quality of various ML models applied in the 19 studies of AI-related voice analysis in Parkinson's' Disease retrieved from The Royal Society of Medicine Library UK, spanning the period from 2013 to 2023. The models assessed include Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), Random Forest (RF) and hybrid CNN-LSTM architectures. Their performance is examined based on accuracy, sensitivity, specificity and error metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Findings indicate that SVM consistently delivers high accuracy (up to 96%) and is particularly effective for small to medium-sized voice-related datasets with pre-engineered datasets. CNNs achieve superior performance (up to 97%) on large, feature-rich datasets; however, their computational demands and limited validation constrain scalability. Random forest models demonstrate robustness in handling imbalanced datasets, while CNN-LSTM hybrids show potential by integrating spatial and temporal feature extraction, though they require further validation.

A critical limitation identified in the analyzed studies is the lack of detailed dataset descriptions, diversity and real-world applicability, which restricts comparison with other studies and generalizability. This paper highlights the strengths and limitations of current models for AI-driven voice analysis approaches and emphasizes the need for standardized, diverse datasets and enhanced evaluation metrics to advance AI applications in voice disorder diagnostics and monitoring.

Keywords: Artificial Intelligence; Vector Machines; Convolutional Neural Networks; Random Forest

Introduction

Background on Parkinson's Disease (PD)

This article is a continuation of the article 'Artificial Intelligence for Screening Voice Disorders: Aspects of Risk Factors'. We looked at the deficiencies of articles used for AI (Artificial Intelligence) analysis of the acoustic measurements of voice [1,2]. The continuation revolves around voice-related AI in software models and their quality. Parkinson's Disease in the literature seems to be the disorder used the most frequently for testing and evaluating various software models on acoustical voice measurements.

Importance of Early Detection and Non-Invasive Methods

To outline the theme, we have earlier searched the library of The Royal Society of Medicine UK for voice-related AI, during the period: 2013 - 2023. 24 papers were researched on voice analysis of Parkinson's Disease. 5 out of 24 articles were reviews. We did

an overview of risk factors in the software [2]. There were many details of deficiencies in the articles, especially with a focus on evaluation metrics, which are necessary to correct in the future. It is already clear that the dataset must be extremely well defined, not only for vowels including /m/ compared to that of /a/ and /u/, but certainly age, gender and race are necessary to be added to the discussion and the stage/level of the disorder in question itself. For comparison to other papers, a crucial factor is whether a study focuses on diagnostics or treatment, particularly in the context of treatment approaches.

Dysphonia is seen in 3-9% of the adult population. In Europe with a population of 746 million. 960 thousand adults with swallowing and voice problems are diagnosed with Parkinson's Disease. In the United States, it is 800 thousand out of 332 million people. The disorder is thoroughly described in many papers [1]. This leads us to the role of voice analysis. In the 19 papers analyzed, only acoustical phenomena are considered. When it comes to biomarkers, also in Parkinson's Disease, other biomarkers used are Voice Handicap Index (VHI), which is a method for grading the subjective complaints of the client. The GRBAS test is a method used by voice experts to assess the quality of a client's sound production. Airflow measurement is also essential for a comprehensive voice evaluation, with Minimum Phonation Time (MPT) being a fundamental parameter. The basic acoustic measures of voice analysis as biomarkers are the fundamental frequency (F0), jitter, shimmer and Harmonics to Noise Ratio (HNR). These biomarker suggestions are based on the work of the Union of European Phoniatricians (UEP) committee. The biomarker suggestions refer to a consensus report [3]. This consensus is of great value for defining each biomarker and to describe how to measure and perform the data collection for datasets usable to AI. E.g. in the case of laryngoscopy, it was shown in a study that only half of the measurements were usable for AI [4].

The aim of this paper was to compare AI models on voice-related acoustic datasets/features in PD used in 19 papers in the literature from 2013-2023.

In the future foundation models of AI will thereafter be a solution for evaluating client states of voice in clinical practice. Acoustic measures will not be enough, as referred to in the consensus paper [3]. But it is necessary to find a usable AI model for acoustic analysis to be adapted to clinical applications and usable as a part of foundation models. This was the reason for our comparison of models for voice analysis in Parkinson's Disease in the 19 analyzed papers.

Ethical Statement

The project did not meet the definition of human subject research under the purview of the IRB according to federal regulations and therefore, was exempt.

Methodology

We analyzed the 19 papers for the quality of various AI models usable for voice-related clinical applications. Some of the papers are referred to when relevant in the discussion of the models. The various models included traditional machine learning models like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Random Forest (RF), Gradient Boosting Models (e.g., LightGBM, XGBoost) and deep learning approaches such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memories (LSTMs) and hybrid models (for example combining CNN and LSTM). For feature selection and dimensionality reduction techniques, Principal Component Analysis (PCA), Grey Wolf Optimization (GWO) and Lasso Regression were found. The models were found in articles dating back to 2013 and up to 2023.

Results

Which Models are Consistently Performing Well Across Studies?

Based on the analysis of the 19 articles, the models that consistently performed well in accurately identifying and classifying voice features across studies are Support Vector Machines (SVM), Convolutional Neural Networks (CNNs) and Random Forest (RF). These models demonstrated high accuracy, sensitivity and specificity on voice features (Table 1).

SVM was the most frequently used model, evaluated in 11 articles.

SVM achieved accuracies of voice-related parameters ranging from 84% to 96% across various datasets, particularly those focused on sustained vowels such as /a/, /u/ and /m/. It was shown that SVM's ability to handle high-dimensional data and its robustness against overfitting made it a reliable choice [5-7].

CNNs, though tested in only 3 studies, showed strong performance with accuracy between 85% and 97% when applied to feature-rich datasets. Examples in CNNs were particularly effective when used on larger datasets with minimal preprocessing [8,9].

Random Forest (RF) was evaluated in 5 studies and demonstrated consistent performance, with accuracy ranging from 83% to 88%. RF was particularly robust in studies where feature selection techniques like PCA or Grey Wolf Optimization (GWO) were applied [10].

Models like K-Nearest Neighbors (KNN) and LSTM (Long Short-Term Memory) also performed well, but their success depended heavily on the dataset size and quality. For instance, KNN achieved an accuracy of 80%-91% in studies using smaller, well-labeled datasets [11,12].

LSTM models excelled in capturing temporal patterns in voice data, achieving an accuracy of up to 85% [13].

Model	Number of Studies	Accuracy (%) (No. of Studies)	F1-Score (%) (No. of Studies)	Recall/Sensitivity (%) (No. of Studies)	Precision (%) (No. of Studies)	Specificity (%) (No. of Studies)
SVM (Support Vector Machines)	11 out of 19	84-96 (10 out of 19)	80-95 (6 out of 19)	89-95 (8 out of 19)	85-94 (5 out of 19)	87-93 (8 out of 19)
CNN (Convolutional Neural Networks)	3 out of 19	85-97 (3 out of 19)	82-96 (2 out of 19)	89-96 (3 out of 19)	85-92 (2 out of 19)	88-92 (3 out of 19)
Random Forest (RF)	5 out of 19	83-88 (4 out of 19)	81-90 (3 out of 19)	86-91 (4 out of 19)	80-89 (3 out of 19)	85-90 (4 out of 19)

Table 1: Models information across studies.

Table 1 out of the 19 studies, 11 studies used SVM, 10 reported accuracy, 8 sensitivity and 8 specificities. Only 6 included F1-score and 5 precision. Out of the 3 studies that used CNN, all 3 studies reported accuracy, sensitivity and specificity. Only 2 included F1-score and precision. Of the 5 studies that used RF, 4 reported accuracy, sensitivity and specificity. 3 reported F1-score and precision (recall is the same as sensitivity).

Fig. 1 shows the confusion matrix, where scores should be found (Table 1).

The models must achieve high scores to be considered for clinical practice. However, having an accuracy score alone is not sufficient, as it does not provide insight into the balance of the dataset. For example, if a dataset is unbalanced-e.g., 95 out of 100 data points are actual positives-a model could predict all cases as positive and still achieve a 95% accuracy, even though it has failed to identify any true negatives. This is why it is important to also include metrics such as sensitivity, specificity, precision and F1-score. Sensitivity measures the ability to correctly identify true positives, while specificity evaluates the ability to correctly identify true negatives.

Precision highlights how many of the predicted positives are correct and the F1-score balances precision and sensitivity, offering a more comprehensive measure of a model's performance, especially in imbalanced datasets. These metrics collectively provide a fuller picture of how well a model performs and ensure that high accuracy does not mask critical weaknesses in other aspects of detection (Fig. 1).

At most times it is not clinically possible to find large materials referring to one specific disorder under specific well-defined circumstances. In the clinics therefore, the aspect will be using for example four biomarkers in combination (VHI, GRBAS test, basic acoustic measures, MPT) [3]. For this, a foundation model as a solution can be necessary.

The strengths and limitations of the models include the following aspects in the 19 studies:

1. Support Vector Machines (SVM)

Strengths

- High Accuracy and Robustness: SVM consistently achieved high accuracy (84-96%) across 10 of 11 voice-related studies, in examples with smaller datasets and pre-selected features.
- Effective for High-Dimensional Data: SVM excels in handling datasets with many features shown making it suitable for voice-based disorder detection where acoustic features like F0, jitter, shimmer and HNR are analyzed.
- Well-Established Technique: SVM is widely adopted for voice analysis due to its interpretability and relatively lower computational cost compared to deep learning models [5-7]

Limitations

- Limited Scalability: SVM struggles with very large datasets due to computational intensity in solving quadratic optimization problems [7]
- Manual Feature Selection: Its performance heavily depends on high-quality voice feature engineering (data sets), which may limit its adaptability to raw or minimally processed data [5]

2. Convolutional Neural Networks (CNN)

Strengths

- Excellent for Raw Data: CNNs performed exceptionally well (accuracy 85-97%) in datasets where raw voice data were used, due to their ability to automatically learn hierarchical patterns [6,8]
- Generalization Power: CNNs excel at generalizing across larger datasets and diverse voice feature sets, making them particularly effective in studies involving vowel sounds [9]
- End-to-end Learning: Unlike traditional models, CNNs do not require manual voice feature selection, reducing dependency on domain expertise [9]

Limitations

- High Computational Requirements: CNNs require significant computational resources, which can make them impractical for smaller voice-related research groups or real-time applications [10]
- Overfitting Risk: Due to their complexity, CNNs are prone to overfitting on smaller datasets unless regularization techniques like dropout are employed [9]
- Fewer Studies: CNN was used in only three studies, which exemplifies its limited demonstrated generalizability in the detection of voice parameters; this limitation is further highlighted in the references [6,9]

3. Random Forest (RF)

Strengths

- Robust and Flexible: RF exhibited consistent performance for voice analysis (accuracy 83-88%) across datasets of varying sizes and feature qualities, as shown [7,10]
- Handles Imbalanced Data well: RF's ensemble nature allows it to handle imbalanced voice-related datasets effectively [7]
- Interpretable Results: It was shown that voice-related feature importance rankings provided by RF make it easier to understand which acoustic features are most relevant [10]

Limitations

- Dependency on voice feature selection: An example is given of RF's performance relying on effective feature selection techniques, such as PCA, to avoid overfitting and optimize its classification capabilities [14]
- Moderate Computational Demand: Although RF is less resource-intensive than CNNs, it can still become computationally expensive when applied to very large voice-related datasets [7]

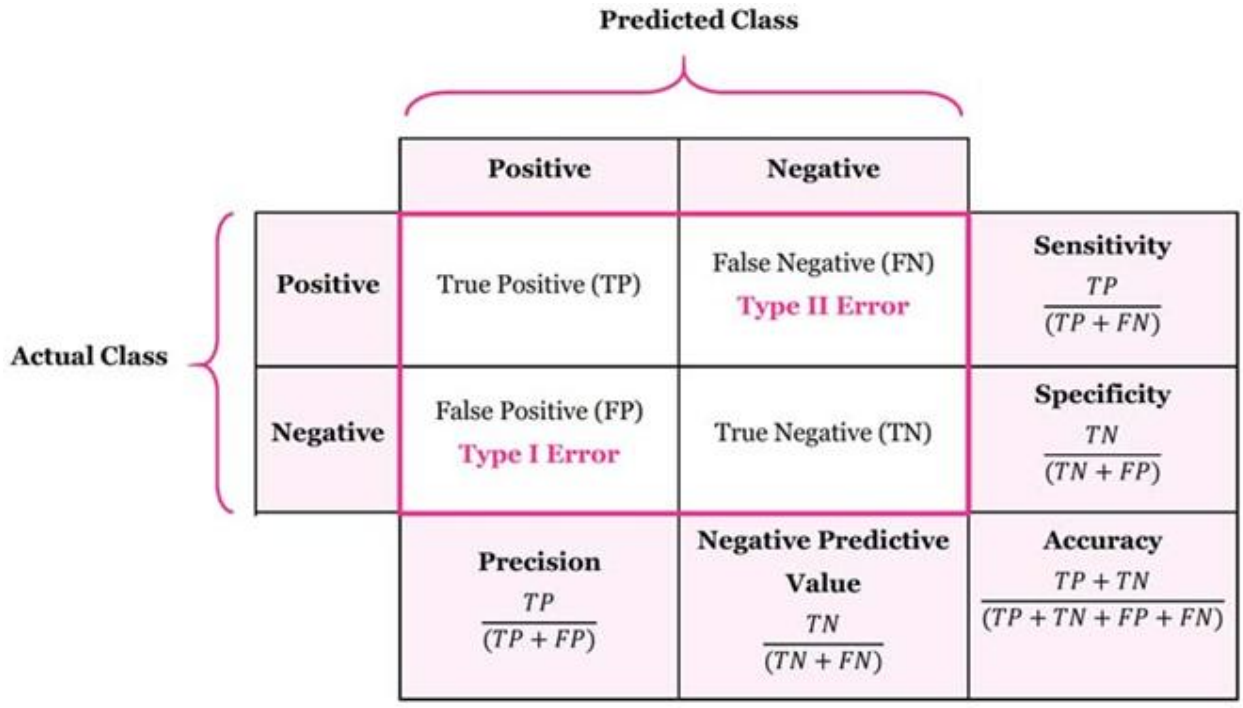

Figure 1: Confusion matrix.

Which Models Can Be Reliably Used for Voice Analysis?

The reliability of machine learning models depends on their ability to generalize across diverse datasets, maintain high-performance metrics and handle different feature types effectively. Beyond consistent performance, reliability also considers practical aspects like computational cost, ease of use and adaptability to various real-world scenarios. The most frequently studied models-Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Random Forest (RF), Hybrid CNN-LSTM models and Ensemble Methods-each demonstrate strengths and limitations that influence their reliability (Table 2).

SVM, the most widely evaluated model, consistently achieved high-performance metrics across studies, particularly for small to medium datasets with engineered acoustic features. Its low computational cost and broad generalizability make it a reliable choice for voice application in both research and clinical settings. CNNs, while evaluated in fewer studies, showed exceptional reliability in datasets with complex or raw features. Their automated feature extraction reduces reliance on domain expertise, though their high computational requirements may limit their applicability in resource-constrained setups. Random Forest models offer robustness and interpretability, performing well in imbalanced datasets and benefiting significantly from feature selection techniques like PCA.

Hybrid CNN-LSTM models present a promising option for datasets requiring both spatial and temporal feature extraction, combining the advantages of CNNs and LSTMs. However, their computational demands and limited validation make them less accessible for widespread use. Finally, Ensemble Methods, including AdaBoost and Gradient Boosting Machines, demonstrated reliability in handling small or imbalanced datasets, with simplicity and robust sensitivity as their key strengths.

Table 2 shows the benefits and costs of using the different models for voice applications. Some might be easy to implement and use but at the cost of computation.

The choice of a reliable model for voice application ultimately depends on the specific characteristics of the dataset/features and the intended application. SVM emerges as the most broadly applicable model due to its low computational requirements and consistent performance across datasets with engineered voice features. CNNs excel in large-scale, feature-rich datasets but require greater computational resources. Random Forest remains a strong option for imbalanced datasets, offering interpretability and flexibility with feature selection techniques. Meanwhile, Hybrid CNN-LSTM models and Ensemble Methods

hold significant promise but are best suited for specific scenarios such as datasets with temporal dependencies or class imbalances.

In addition to accuracy and other classification metrics, error measurements like Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R^2 (Coefficient of Determination) provide critical insights into the performance of machine learning models, particularly when dealing with regression tasks or continuous outputs. These metrics quantify the magnitude of prediction errors, offering a complementary perspective to classification metrics such as sensitivity and specificity (Table 3).

Model	Generalizability	Ease of Use	Computational Cost	Applications
SVM	High for pre-engineered datasets/features	High	Low	Best for small to medium datasets with manual datasets/features.
CNN	High for raw data	Medium	High	Effective for large datasets.
Random Forest	Medium	High	Medium	Reliable for imbalanced datasets, requires feature selection.
Hybrid CNN-LSTM	Medium to High	Medium	High	Suitable for datasets with spatial and temporal features.
Ensemble Methods (e.g., AdaBoost)	Medium	High	Low	Reliable for small datasets with class imbalances.

Table 2: Reliability of models for voice analysis.

Error Metric	Number of Studies Reporting	Range	Relevance
MAE (Mean Absolute Error)	5 out of 19	0.02-0.15	Quantifies the average magnitude of prediction errors without emphasizing larger errors.
RMSE (Root Mean Square Error)	6 out of 19	0.03-0.18	Emphasizes larger errors, providing a better sense of model performance for extreme values.
R^2 (Coefficient of Determination)	4 out of 19	0.75-0.95	Indicates how well the model explains variability in the data, higher values suggest better fit.

Table 3: Summary of error metrics reported across the 19 studies.

Table 3 RMSE, reported in 6 studies out of 19 studies, showed values ranging from 0.03 to 0.18, indicating variability in the models' ability to handle extreme errors. Lower RMSE values were associated with models like CNN and hybrid CNN-LSTM. MAE, reported in 5 studies, ranged from 0.02 to 0.15, highlighting the average error magnitude. This metric was most frequently used with regression-based methods and models like SVM. R^2, reported in 4 studies, ranged from 0.75 to 0.95, showing good variability explanation by some models but also highlighting gaps in generalization for certain datasets.

Models like CNNs and Random Forests tended to report lower error metrics when used with well-processed datasets, such as examples with manually engineered features of voice analysis [6,14]. However, it is shown that smaller voice datasets or datasets

with noise often lead to higher errors, particularly for models relying on raw data inputs [7,10].

Datasets

In these 19 papers, the fundamental requirements outlined in the consensus paper are not comprehensively addressed, particularly regarding sound selection and measurement setup. While some datasets provide methodological clarity and simplicity, the majority fall short of the standards necessary for robust, real-world machine learning applications in voice detection [3].

Discussion

It is crucial to carefully evaluate various factors when selecting the most reliable model for specific objectives, ensuring both robustness and scalability, as demonstrated in PD voice analysis applications. We have earlier described the use of AI and high-speed films (15.732 videos) [4]. The problem was that only half of the films presented the vocal folds sufficiently for AI use. Therefore, we focused on artificial intelligence in voice-related acoustical models and other options. It seems that the data sets used in the 19 articles from 2023 to 2023, also are very insufficiently described and not in conformity with the consensus [3]. Understandably, the consensus does not incorporate AI, except for a single reference to the GRBAS test" [15].

For clinical use, models need to not only achieve high accuracy but also demonstrate generalizability and robustness across diverse datasets. Support Vector Machines (SVM) consistently performed well, achieving accuracies up to 96% and strong sensitivity and specificity. Its low computational cost and reliability on small to medium datasets make it a practical option for clinical settings, though its dependence on manually engineered datasets (features) may limit its scalability. Manually engineered datasets/features are probably nonetheless necessary. Convolutional Neural Networks (CNNs) showed potential with larger, feature-rich datasets, achieving up to 97% accuracy without requiring manual feature selection. However, their high computational demands and limited validation reduce their immediate applicability in clinical practice. Random Forest models, while slightly less accurate, proved robust and interpretable, particularly when dealing with imbalanced datasets, making them useful where dataset selection techniques can be applied. Combining CNN-LSTM models into a single framework showed promise in combining spatial and temporal analysis, but their computational intensity and limited studies hindered their readiness for deployment. Error metrics such as RMSE and MAE highlight these differences further, with CNNs and hybrid models minimizing prediction errors more effectively, but at the cost of complexity [16-2].

Conclusion

Ultimately, the choice of an AI model for voice-related clinical use depends on the specific requirements of the application as shown in the 19 voice-related AI papers of PD. SVM stands out for its simplicity and reliability in smaller datasets, while CNNs could be valuable for larger datasets if computational resources allow. However, the lack of standardized and diverse datasets remains a major limitation, underscoring the need for further development before any of these models can be considered fully ready for clinical implementation.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Financial Disclosure

This research did not receive any grant from funding agencies in the public, commercial or not-for-profit sectors.

Author's Contribution

Mette Pedersen made the conception, design of the work and drafting of the work.

Vitus Girelli Meiner, made substantial contributions to the acquisition, analysis, interpretation of data and drafting of the work.

References

1. Tanner CM, Ostrem JL. Parkinson's disease. New England J Med. 2024;391(5):442-52.
2. Pedersen M. Artificial intelligence for screening voice disorders: Aspects of risk factors. Am J Medical and Clinical Res Rev. 2025;4(2):1-8.
3. Lechien JR, Geneid A, Bohlender JE, Cantarella G, Avellaneda JC, Desuter G, et al. Consensus for voice quality assessment in clinical practice: guidelines of the European Laryngological Society and Union of the European Phoniatricians. European Archives of Oto-Rhino-

Laryngology. 2023;280(12):5459-73.
4. Pedersen M, Larsen CF. Accuracy of laryngoscopy for quantitative vocal fold analysis in combination with AI, a cohort study of manual artefacts. Sch J Otolaryngol. 2021.
5. Viswanathan R, Arjunan SP. Estimation of severity in Parkinson's disease using acoustic features of phonatory tasks. IETE J Research. 2023;69(9):6292-303.
6. Dao SV, Yu Z, Tran LV, Phan PN, Huynh TT, Le TM. An analysis of vocal features for Parkinson's disease classification using evolutionary algorithms. Diagnostics. 2022;12(8):1980.
7. Manor Y, Naor S, Shpunt D, Diamant N, Hillel A, Ezra A, et al. Machine learning classifiers and subjective vocal perception of Parkinson's disease patients and healthy control. Mov Disord. 2019;34(suppl 2).
8. Jain A, Abedinpour K, Polat O, Çalışkan MM, Asaei A, Pfister FM, et al. Voice analysis to differentiate the dopaminergic response in people with Parkinson's disease. Frontiers in Human Neuroscience. 2021;15:667997.
9. Pah ND, Motin MA, Kumar DK. Phonemes based detection of parkinson's disease for telehealth applications. Scientific Reports. 2022;12(1):9687.
10. Bao G, Lin M, Sang X, Hou Y, Liu Y, Wu Y. Classification of dysphonic voices in Parkinson's disease with semi-supervised competitive learning algorithm. Biosensors. 2022;12(7):502.
11. Suppa A, Costantini G, Asci F, Di Leo P, Al-Wardat MS, Di Lazzaro G, et al. Voice in Parkinson's disease: a machine learning study. Front Neurol. 2022;13:831428.
12. Rajasekar SJ, Narayanan V, Perumal V. ParkAI: An AI based tool for detection of parkinson's disease using vocal measurements. Movement Disorders. 2021;36:1.
13. Rajeswari SS, Nair M. Prediction of Parkinson's disease from voice signals using machine learning. J Pharmaceutical Negative Results. 2022;13.
14. Altay EV, Alatas B. Association analysis of Parkinson disease with vocal change characteristics using multi-objective metaheuristic optimization. Medical Hypotheses. 2020;141:109722.
15. Costantini G, Cesarini V, Di Leo P, Amato F, Suppa A, Asci F, et al. Artificial intelligence-based voice assessment of patients with Parkinson's disease off and on treatment: machine vs. deep-learning comparison. Sensors. 2023;23(4):2293.
16. Lim WS, Chiu SI, Wu MC, Tsai SF, Wang PH, Lin KP, et al. An integrated biometric voice and facial features for early detection of Parkinson's disease. NPJ Parkinson's Disease. 2022;8(1):145.
17. Yu Q, Zou X, Quan F, Dong Z, Yin H, Liu J, et al. Parkinson's disease patients with freezing of gait have more severe voice impairment than non-freezers during "ON state". J Neural Transmission. 2022;129(3):277-86.
18. Gaballah A, Parsa V, Cushnie-Sparrow D, Adams S. Improved estimation of parkinsonian vowel quality through acoustic feature assimilation. The Scientific World J. 2021;2021(1):6076828.
19. Park JE, Oh SW, Shin JY, Lee SY, Hong SH, Ahn NH, et al. Say "AH~": Vocal analysis in parkinson's disease and essential tremor. Mov Disord. 2020;35(suppl 1).
20. Da Cruz Morello AN, Beber BC, Fagundes VC, Cielo CA, Rieder CR. Dysphonia and dysarthria in people with Parkinson's disease after subthalamic nucleus deep brain stimulation: Effect of frequency modulation. J Voice. 2020;34(3):477-84.
21. Viswanathan R, Arjunan SP, Bingham A, Jelfs B, Kempster P, Raghav S, et al. Complexity measures of voice recordings as a discriminative tool for Parkinson's disease. Biosensors. 2019;10(1):1.
22. Sheibani R, Nikookar E, Alavi SE. An ensemble method for diagnosis of Parkinson's disease based on voice measurements. J Medical Signals and Sensors. 2019;9(4):221-6.
23. Arora S, Baghai-Ravary L, Tsanas A. Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice. The J Acoustical Soc Am. 2019;145(5):2871-84.