

Diagnosing Parkinson's disease through artificial intelligence using sound samples

Yunjin Chen^{1,*}

¹Faculty of Science, University of New South Wales, Sydney, Australia

*Corresponding author: z5565791@ad.unsw.edu.au

Abstract:

Phonological changes in PD reflect early neurodegenerative changes that affect speech motor control. This paper reviews the development of remote monitoring methods from early acoustic analysis to modern deep learning-based methods. As a non-invasive, low-cost and sensitive tool, speech-based artificial intelligence biomarkers can be used for screening, early diagnosis and longitudinal monitoring. Its advantage lies in its ability to detect subtle changes that are difficult for humans to detect and its potential for high-frequency remote assessment via smartphones. However, some problems remain to be solved, such as the difference in recording environment, the limited annotation data set, and the "black box" nature of AI models limiting the clinical interpretability. In recent years, advances in interpretable AI and multimodal data fusion have brought hope to clinical applications. Future research should focus on improving data diversity, model robustness and transparency, and combine with clinically relevant outcome indicators to promote the wide application of voice AI biomarkers in personalized PD management.

Keywords: Parkinson's disease, sound samples, machine learning

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease. Parkinson's disease is pathologically characterized by the presence of Lewy bodies and Lewy neurites, which are mainly composed of aggregated and misfolded α -synuclein, accompanied by neuronal loss in substantia nigra and other brain regions, and therefore classified as synucleinopathy [1]. Early signs of Parkinson's disease include voice changes (including reduced speech volume and rhythm, frequent speech pauses and

shortened speech), reduced facial expressions and gait patterns with reduced arm related movements [2]. These mild signs are often mistakenly identified as normal aging, which can delay the diagnosis of PD and miss the optimal time for intervention [2]. At present, the most effective treatment for motor symptoms of Parkinson's disease is still the use of Carbidopa but given the heterogeneity in the progression of motor and non-motor features, personalized treatment is important [3]. Early identification of PD and early treatment training can effectively improve the physical activity and non-motor symptoms of PD

patients. For example, patients who received Lee Silverman Voice Therapy (LSVT) have effectively improved their scores and speech intelligibility on the Parkinson's Disease Rating Scale (UPDRS-III) [4].

At present, the diagnosis of PD is still generally the use of clinical observation, because of the lack of clear diagnostic tests, it can only be judged by paying close attention to the relevant characteristics [3]. As an early symptom of Parkinson's disease, voice changes are expected to be a marker for early diagnosis. Machine learning (ML) was used to classify voice samples, and Support Vector Machine (SVM) was used to identify key features in voice samples and analyze speech recordings at different frequencies, to build a model that can diagnose Parkinson [5]. This non-invasive, low-cost, and remote sampling method can be used to diagnose PD.

2. Speech Analysis in Parkinson's Disease Diagnosis

2.1 . Evolution of Speech-Based PD Diagnosis

The traditional method for diagnosing Parkinson's disease has always been based on clinical observations of patients. Through clinical assessment and examination of past medical history, cases of mimicking Parkinson's symptoms are excluded to diagnose Parkinson's disease. Now, by searching for biomarkers as precursory symptoms, a method for early diagnosis is being developed and through positron emission tomography (PET), the failure of the dopaminergic system in the substantia nigra and striatum can be proven 10 years before the onset of the disease, thus enabling early diagnosis of Parkinson's [6]. Unfortunately, although PET is currently the most accurate detection method, due to its high cost, this method cannot be used for preventive examinations of healthy individuals [6]. Therefore, using changes in voice as early markers and combining with AI analysis would be a good way for remote, low-cost preventive examinations of healthy individuals in the future. Recording language samples using smartphones to capture speech disorders in high-risk PD populations is expected to improve the rate of early PD diagnosis [7,8]. This idea is currently still in the experimental stage and there is relatively little research in this area. The basic research is based on using SVM for classification and machine learning analysis of the voice characteristics of PD patients and healthy individuals to train AI [7,9].

A study utilized a dataset from the UCI Machine Learning Repository with 756 voice recordings from 252 individuals (188 PD patients and 64 controls), extracting features

such as jitter, shimmer, Mel-Frequency Cepstral Coefficients (MFCC), and wavelet transforms. After evaluating the supervised machine learning classifiers capable of identifying patients with Parkinson's disease, a support vector machine (SVM) model was applied for feature selection. Then, based on these selected features and the determined combination of machine learning algorithms, the patients with Parkinson's disease were classified. The accuracy rate reached 85.09%, indicating the potential of speech biomarkers in early detection [5].

Recent research results have also identified Parkinson's patients through various smartphone functions (voice, finger tapping, gait), among which speech features such as pause percentage and pitch variation have enabled an integrated model using an SVM analyzer to achieve the best diagnostic value of 0.86 in comparing Parkinson's patients with unaffected control groups, outperforming single-modal models [7].

Additionally, comparisons between ML and DL for voice assessment in PD patients off and on treatment showed ML (e.g., SVM) achieving up to 85% accuracy in binary classifications, while DL on Mel-spectrograms reached 69.75%, suggesting ML's suitability for smaller datasets [9]. These developments illustrate a shift from basic acoustic analysis to sophisticated AI-driven remote monitoring, with ongoing improvements in accuracy and applicability.

2.2 . Pathophysiological Basis of Speech Changes

The reason why patients with Parkinson's disease exhibit low-pitched voices, reduced volume, monotonous tones, and breathy sounds is that the glottis does not close completely, and the vocal cords are bent due to the stiffness or sluggishness of the laryngeal muscles [10]. Phonation is the result of the secondary sound generation of vocal fold vibration and air pressure change. Vocal muscles are responsible for regulating the length of vocal folds, to achieve the effect of regulating pitch [10]. These muscles are innervated by branches of the vagus nerve, and dopamine loss in Parkinson's disease interferes with motor control pathways, thereby affecting these regions, resulting in stiffness of vocal cord muscles and causing subtle changes in speech [10,11]. Longitudinal studies have shown that in the pre-Parkinson's disease stage, the fundamental frequency of speech changes, which is related to early brainstem involvement and dopamine function loss [11]. Acoustic analysis reveals an increase in tremors and twitches, reflecting the speech disorder caused by damage to nigral neurons [10]. Induced animal models with dopamine deficiency also exhibit abnormal speech, supporting

these mechanisms [10].

2.3 . Benefits of AI-Driven Speech Analysis

Smartphone based speech acquisition and AI analysis have significant advantages in PD diagnosis. First, the method is non-invasive, low-cost, and easy to replicate. Lim et al. proposed that the voice recording technology of smartphones can achieve cost-effective large-scale screening, especially in regions where there is a shortage of neurologists, because the equipment required is only ordinary equipment and does not need special equipment [2]. Additionally, a comprehensive review by Sahar Shokrpour et al. also pointed out that the method based on voice is low in cost and noninvasive, capable of achieving early detection and without posing risks to patients [12].

Second, AI analytics can identify subtle speech changes that are difficult for humans to perceive, enabling early warning. The research conducted by Mohammad Amran Hossain and Francesco Amenta indicates that machine learning models can identify acoustic features in early Parkinson's disease, such as pitch variations and tremors, which have a high sensitivity for the early stages of human perceptual failure [5]. Moreover, by analyzing speech patterns using deep learning, subclinical changes can be detected years before the onset of motor symptoms, and it has extremely high accuracy in pathological speech diagnosis [12].

Additionally, smartphone technology enables remote, high-frequency voice monitoring, facilitating physicians to continuously track high-risk populations. For instance, through multimodal smartphone applications, real-time remote diagnosis can be achieved, which is applicable to the elderly population and in the context of the epidemic situation, thereby enhancing accessibility [7]. Voice assessment conducted via mobile devices supports remote healthcare, enabling frequent monitoring without the need to visit a clinic, and this has been demonstrated in the evaluation of treatment effects [9].

2.4 . Challenges in Speech-Based AI Models

Although promising, there are potential limitations to this approach. First, the performance of machine learning is highly data dependent. Currently, speech samples that can be used for training are limited, resulting in models that are prone to overfitting and difficult to generalize to different populations [12,13]. The research conducted by Sahar Shokrpour et al. indicates that the limited nature of the data sets in speech research often results in the constructed models being unable to generalize across different populations, and a small sample size would undermine their reliability [12].

Secondly, many studies only build models based on a single language, which may produce bias and affect accuracy in a cross-language environment [12]. Meanwhile, the "black box" characteristics of AI models reduce the clinical interpretability, making it difficult for physicians to trust the diagnostic basis of the model [13]. Giovanni Costantini et al. points out that although deep learning has been proven to have excellent performance, this low-interpretable model will hinder clinicians from providing advanced evidence in clinical practice and prevent the application of this technology in the clinical field [9].

In addition, how to obtain real patient data while protecting privacy is also a challenge. The research indicates that people have concerns about the privacy of continuous collection of PD patients' data and data sharing. This method of obtaining sensitive voice data has ethical barriers, and proprietary data sets limit the accessibility of the data [12]. Continuous recording raises concerns about data security, and thus it is quite important to adopt privacy-protecting technologies [7].

2.5 . Clinical Applications of Voice AI

The application of artificial intelligence models in this field includes population screening, early detection of latent or high-risk groups, disease progression monitoring, treatment effect evaluation (including drug and surgical intervention), and supervision of speech therapy compliance [14]. Remote high-frequency monitoring via a smartphone or wearable microphone can continuously collect data to capture subtle longitudinal changes, thereby supporting personalized treatment adjustment [15]. These application methods can be summarized into three categories: population screening and early detection, disease progression monitoring, and treatment effect evaluation.

Population screening and early detection: This method is based on analyzing subtle speech disorder characteristics through AI using speech samples, enabling non-invasive screening of high-risk individuals. For instance, a comprehensive model using smartphone speech features such as pause percentage and pitch variability achieved an AUROC of 0.82 in early PD, facilitating large-scale remote screening [7]. Another study combined speech and facial features and achieved an AUROC of 0.89, suitable for community identification of early PD [2]. The Tap Talk smartphone application's cross-device automated speech analysis assists in early detection with high accuracy [14].

Disease Progression Monitoring: Longitudinal speech monitoring tracks the progression of PD through changes in acoustic features. A study on the progression of speech impairment found that over time, aphonia increased, and features such as shimmer were related to the severity of

the disease, enabling continuous assessment [15]. ML models use speech biomarkers to distinguish between early and late-stage PD with 85% accuracy, supporting progression tracking [9]. Multimodal methods including speech predict the progression of PD through features such as MFCCs [4].

Evaluation of therapeutic effect: The voice AI assessment, such as L-Dopa or speech therapy intervention response. The comparison of voice before and after treatment shows that ML detects the drug effect through features like jitter, and the accuracy rate in distinguishing states is 79% [9]. LSVT improves speech intelligibility and can be measured through post-treatment voice analysis [4]. The DL model evaluates the treatment through speech patterns and provides objective adjustment indicators [12].

3. Discussion

Although speech-based AI biomarkers for PD hold great promise, several challenges need to be addressed before routine clinical adoption.

First, data quality and representativeness are urgent issues. Speech recordings collected in uncontrolled conditions are susceptible to environmental noise, microphone differences, and inconsistent speaking styles, resulting in performance degradation when deployed in the real world. The presence of data imbalances in existing datasets, such as underrepresentation of specific languages, dialects, or socioeconomic groups, may introduce systematic bias and reduce diagnostic accuracy in vulnerable populations.

Second, the clinical translation of AI systems will require rigorous prospective validation and demonstration of their added value over existing clinical assessment tools. Regulators and health care providers have demanded transparent evidence of model interpretability, reliability, and safety. The “black box” nature of many deep-learning models remains a barrier to clinical adoption; An interpretable mapping of acoustic features to physiological changes can help build trust and integrate into clinical workflows.

Thirdly, the ethical and privacy issues associated with large-scale voice monitoring cannot be ignored. The continuous recording of voices has raised concerns about data security and potential abuse. Adopting a privacy-protecting computing framework and complementing it with a comprehensive informed consent mechanism is crucial for maintaining public trust.

Therefore, future research should first address the existing methodological and clinical translation barriers. Increasing the diversity of data sets (language, accent, age, gender, and device type) is the key to improve the generalization ability of the model. Techniques such as anti-noise feature extraction, domain adaptation and federated learning can

mitigate the impact of environmental differences. Incorporating interpretable AI frameworks to provide model decision visualization for clinical staff can help build trust and drive regulatory approval. Prospective, multicenter, longitudinal validation studies should be prioritized to confirm the predictive performance and clinical value.

4. Conclusion

Speech-based AI biomarkers provide a non-invasive, scalable, and potentially sensitive approach for early detection and longitudinal monitoring of PD. Its ability to capture subtle, subclinical changes in speech production makes it a valuable complement to traditional neurological assessment. However, technical, translational, and ethical hurdles must be overcome to realize its full potential. Future work should prioritize the development of diverse, noise-robust, interpretable models that can be validated in prospective, multicenter studies while being consistent with clinically meaningful outcome measures. These advances will set the stage for reliable, personalized, and accessible digital diagnostics, leading to improved patient care and earlier therapeutic interventions.

References

- [1] Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson’s disease. *Lancet Neurol*. 2021;20(5):385–97. Available from: [https://www.thelancet.com/journals/lanneur/article/PIIS1474-4422\(21\)00030-2/fulltext](https://www.thelancet.com/journals/lanneur/article/PIIS1474-4422(21)00030-2/fulltext)
- [2] Lim WS, Chiu SI, Wu MC, Tsai SF, Wang PH, Lin KP, Chen YM, Peng PL, Chen YY, Jang JSR, Lin CH. An integrated biometric voice and facial features for early detection of Parkinson’s disease. *NPJ Digit Med*. 2022;5(1):149. PMID: 36309501; PMCID: PMC9617232. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9617232/>
- [3] Kobylecki C. Update on the diagnosis and management of Parkinson’s disease. *Postgraduate Medical Journal*. 2020;96(1139):489-494. doi:10.1136/postgradmedj-2020-137525. PMID: 32675145; PMCID: PMC7385761. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7385761/>
- [4] Pu T, Huang M, Kong X, Wang M, Chen X, Feng X, Wei C, Weng X, Xu F. Lee Silverman Voice Treatment to improve speech in Parkinson’s disease: a systematic review and meta-analysis. *Parkinsons Dis*. 2021; 2021:3366870. PMID: 35070257; PMCID: PMC8782619. DOI: 10.1155/2021/3366870. Available from: <https://doi.org/10.1155/2021/3366870>
- [5] Hossain MA, Amenta F. Machine learning-based classification of Parkinson’s disease patients using speech biomarkers. *Healthcare (Basel)*. 2023;11(24):3239. PMID: 38160364; PMCID: PMC10836572. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10836572/>

ncbi.nlm.nih.gov/articles/PMC10836572/

[6] Ugrumov M. Development of early diagnosis of Parkinson's disease: illusion or reality? *CNS Neurosci Ther.* 2020;26(10):997–1009. PMID: 32597012; PMCID: PMC7539842. DOI: 10.1111/cns.13429. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7539842/>

[7] Lim WS, Fan SP, Chiu SI, Wu MC, Wang PH, Lin KP, Chen YM, Peng PL, Jang JSR, Lin CH. Smartphone-derived multidomain features including voice, finger-tapping movement and gait aid early identification of Parkinson's disease. *NPJ Parkinsons Dis.* 2025; 11:111. Available from: <https://www.nature.com/articles/s41531-025-00953-w>

[8] Rusz J, Hlavnička J, Tykalová T, Novotný M, Dušek P, Šonka K. Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Honolulu, HI: IEEE; 2018. p. 4984–7. Available from: <https://ieeexplore.ieee.org/document/8400578>

[9] Costantini G, Cesarini V, Di Leo P, Amato F, Suppa A, Asci F, Pisani A, Calculi A, Saggio G. Artificial intelligence-based voice assessment of patients with Parkinson's disease off and on treatment: machine vs. deep-learning comparison. *Sensors (Basel).* 2023;23(4):2293. PMID: 36850893; PMCID: PMC9962335. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9962335/>

[10] Ma A, Lau KK, Thyagarajan D. Voice changes in Parkinson's disease: What are they telling us? *J Clin Neurosci.* 2020; 72:1–7. DOI: 10.1016/j.jocn.2019.12.029. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0967586819317631>

S0967586819317631

[11] Harel B, Cannizzaro M, Snyder PJ. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. *Brain Cogn.* 2004;56(1):24–9. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0278262604001393>

[12] Shokrpour S, MoghadamFarid AM, Abkenar SB, Kashani MH, Akbari M, Sarvizadeh M. Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges. *NPJ Parkinsons Dis.* 2025; 11:187. Available from: <https://www.nature.com/articles/s41531-025-01025-9>

[13] Aresta S, Battista P, Palmirotta C, Tagliente S, Lagravinese G, Santacesaria P, Benzini A, Mongelli D, Minafra B, Lunetta C, García AM, Salvatore C. Digital phenotyping of Parkinson's disease via natural language processing. *NPJ Parkinsons Dis.* 2025; 11:182. Available from: <https://www.nature.com/articles/s41531-025-01050-8>

[14] Li R, Huang G, Wang X, Lawler K, Goldberg LR, Roccati E, St George RJ, Aiyede M, King AE, Bindoff AD, Vickers JC, Bai Q, Alty J. 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 (Amst).* 2024;16(1): e70025. DOI: 10.1002/dad2.70025. Available from: <https://doi.org/10.1002/dad2.70025>

[15] Skodda S, Grönheit W, Mancinelli N, Schlegel U. Progression of voice and speech impairment in the course of Parkinson's disease: a longitudinal study. *Parkinsons Dis.* 2013; 2013:389195. PMID: 24386590; PMCID: PMC3872441. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3872441/>