RESEARCH

Whispers of Wellness: Unveiling Parkinson's Disease through AI-Powered Voice Analysis

Warren May de la Cruz† *, Miguel Lorenzo Singian† and Arvin Eduardo Ymson†
Aboitiz School of Innovation, Technology, and Entrepreneurship, Asian Institute of Management, Makati City, Philippines

*Correspondence:
wdelacruz.MSDS2023@aim.edu
†Equal contributor

Abstract

Parkinson's disease (PD) is a progressive neurological disorder affecting motor and speech functions, often diagnosed through invasive and costly methods. This paper builds on the research of Iyer et al. (2023), using deep learning to analyze personal telephone-collected voice recordings, successfully differentiating between individuals with PD and healthy controls. The study uses an open-access dataset of voice recordings from 50 PD patients and 50 healthy individuals. The methodology involved converting the raw audio files into spectrograms (visual representations of the frequencies in audio), for features to be extracted by a vision-transformer (ViT) model, and fed to a classification head. The model achieved an average ROC-AUC of 93% to the demonstrating the potential for non-invasive, cost-effective diagnosis, and remote monitoring of PD. Further testing on diverse populations and incorporating longitudinal data at different stages of PD could enhance the robustness and applicability of this approach.

Keywords: student employability; mock interview; machine learning; shap; dice

Highlights

- Non-Invasive Diagnosis: Leveraged AI to differentiate between Parkinson's patients and healthy controls using voice recordings.
- High Accuracy: Achieved an average ROC-AUC of 93% on the validation set and 76% on the holdout set, demonstrating strong diagnostic potential.
- Voice Data Utilization: Utilized a dataset of 100 participants, converting raw audio files into spectrograms for feature extraction.
- Hybrid Model: Employed a hybrid Vision Transformer (ViT) model combining CNN for local feature extraction and ViT for capturing long-range dependencies.
- Practical Applications: Showcased the feasibility of remote and non-invasive diagnosis, beneficial for early detection and monitoring in underserved areas.
- Future Directions: Recommended further testing on diverse demographics and early-stage PD patients to enhance robustness and applicability.

1 Introduction

Parkinson's disease (PD) is a progressive neurological disorder that significantly impacts motor and speech functions. Traditional diagnostic methods for PD often involve invasive and costly procedures, such as clinical assessments, neuroimaging, and dopamine transporter scans, which can be burdensome for patients and healthcare systems[1]. By leveraging personal telephone-collected voice recordings, our project employs machine learning models to differentiate between individuals with PD and healthy controls. This approach not only offers a non-invasive diagnostic alternative but also holds promise for remote monitoring and early detection, particularly in underserved areas [2]. Our study utilizes an open-access dataset comprising voice recordings from a collaboration between the University of Arkansas and Georgia Tech. The dataset includes 100 participants: 50 individuals diagnosed with Parkinson's disease and 50 healthy controls. Each participant provided five-second telephone voicemail samples while sustaining an /a/ sound. By analyzing these samples with a deep learning vision-transformer (ViT) model, our aim is to identify distinguishing features of PD and advance the development of accessible diagnostic tools.

Related Works

Parkinson's Effect on Voice

Voice impairments in PD patients are significant and include issues such as reduced pitch range, decreased articulation rate, and variations in voice intensity. These impairments are often due to vocal muscle disorders, swallowing difficulties, and reduced facial muscle control. Notably, voice changes can occur early in the disease's progression, even before other motor symptoms become apparent [3]. Suppa et al. (2022)[3] conducted a study highlighting these voice impairments and used machine learning techniques to analyze the variations in voice characteristics among PD patients.

Diagnosing PD Through Voice Changes Using ML Machine learning has been extensively applied to enhance the accuracy and robustness of PD diagnosis from voice data. Dinesh and He (2017)[4] is one of many studies that used traditional machine learning methods on extracted features from voice recordings of PD patients, achieving an accuracy score of 91-95%.

Iyer et al. (2023)[5] is part of a new wave of research, using deep learning to extract more features from voice recordings. Iyer et al. (2023)[5] used a convolutional neural network on spectrogram representations of voice recordings from of 40 PD patients and 41 healthy controls, achieving high ROC-AUC of 97%. Their dataset

was made available online at https://doi.org/10.6084/m9.figshare.23849127, and it is this same dataset that we used for our study.

2 Data and Methods

2.1 Data Source

Our study utilizes an open-access dataset comprising voice recordings from a collaboration between the University of Arkansas and Georgia Tech. The dataset includes 100 participants: 40 individuals diagnosed with Parkinson's disease (PD) and 41 healthy controls. Each participant provided five-second telephone voicemail samples while sustaining an /a/ sound.

The features of the dataset are listed below, in Table 1:

Table 1 Data Description

Feature Name	Remarks
Audio Recordings /a/	file extension (.wav); 40 PD patients
Audio Recordings /a/	file extension (.wav); 41 Healthy Controls

The dataset can be found on this website.

3 Methodology

1. Data Preparation The audio was preprocessed using the following steps, imitating the preprocessing of Iyer (2023)[5] where possible.

- Noise Reduction: Noise reduction was applied by clipping audio signals within a specific range. For male voices, the range was set between -75 dB and 300 dB, and for female voices, between -100 dB and 600 dB. This method mitigates background noise by constraining the signal to a defined amplitude range.
- Normalization: The amplitude of the audio signals was normalized by rescaling the clipped data to fit within the range of -1 to 1. This normalization ensures consistency in amplitude levels across all samples, allowing for more accurate subsequent analysis.
- Segmentation: The recordings were segmented to focus on the sustained /a/ sound.
- Energy Calculation: Computing the short-time energy of the audio using a 12.5 ms window to identify non-silent regions.
- Trimming Silence: Using energy thresholds, silent portions were removed by retaining sections where the energy exceeded 5
- Duration Check: Ensuring that the trimmed audio was at least 1.5 seconds long. Shorter recordings were excluded from further processing.

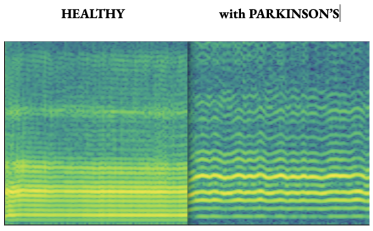

Figure 1 Healthy and with Parkinson's Spectogram

- Extracting Segment: Retaining only the first 1.5 seconds of the trimmed audio for consistency across samples.

3.3 Creation of Spectrograms

Spectrogram Calculation:

- The spectrogram was generated using the following parameters:
  - Window Function: A Hann window was applied to smooth the signal.
  - Segment Length (nperseg): 32 ms segments were used, corresponding to a window length of 32 ms multiplied by the sample rate.
  - Overlap (noverlap): 50% overlap between segments was used, calculated as 16 ms multiplied by the sample rate.
  - Number of FFT Points (nfft): 1024 points were used for the Fast Fourier Transform (FFT) to provide a detailed frequency representation.

  Conversion to Decibels:
  - The spectrogram values were converted to decibels (dB) for better visualization. The values were scaled relative to the maximum absolute value to provide a range from 0 to - dB.
  - The plot was saved as a JPEG file with a high resolution (600 dpi) to ensure clarity and detail in the spectrogram image.

4 Model Training

A hybrid Vision Transformer model with convolutional embedding rather than patch embedding was employed. In order to replicate Iyer et al. (2023), the dataset was first split into a 70% training and 30% validation set. A further holdout study was conducted, with a 70:15:15 split. The model's performance was evaluated using metrics such as ROC-AUC (Receiver Operating Characteristic - Area Under Curve) and accuracy. Training and evaluation was repeated over 5 random splits for the replication study, and over 10 random splits for the holdout study.

5 Results and Discussion

Table 2 Comparison of Replication and Holdout Study Results

Replication Study (70:30 split)	**Holdout Study (70:15:15 split)**
Achieved a ROC-AUC of 93% and an accuracy of 83%	Achieved a ROC-AUC of 76% and an accuracy of 70%, with standard deviations of 11% for both metrics

Discussion

- Comparison with Related Works: The study's findings align with previous research, such as the work by Iyer et al. (2023), highlighting the potential of voice analysis for PD detection. While the hybrid ViT model demonstrated significant accuracy, it underperformed compared to the original study. This may however be explained by differences in preprocessing.
- Practical Implications: The use of telephone-collected voice samples underscores the feasibility of remote and non-invasive diagnosis. This approach is particularly beneficial for early detection and monitoring in underserved areas where traditional diagnostic methods may not be accessible.
- Future Work: Further optimization and testing on more diverse datasets are recommended to enhance model robustness. Incorporating longitudinal data at different stages of PD could provide deeper insights into disease progression and improve diagnostic accuracy.

6 Conclusion

This study aimed to use a Vision Transformer (ViT) machine learning model to distinguish between individuals with Parkinson's disease and healthy controls using voice recordings. The model results show promise as a non-invasive, cost-effective diagnostic tool.

7 Recommendation

To improve our study, we suggest further testing on individuals with early-stage Parkinson's disease (PD). This focus will help in identifying subtle speech changes that occur early on, which is vital for timely intervention. It's also important to test a diverse group of people, including different ages, genders, and ethnic backgrounds, to ensure that our findings apply broadly and are not limited to a specific group.

We recommend using a variety of speech tasks, such as reading, spontaneous talking, and repeating sentences, to provide a comprehensive analysis of speech

patterns associated with PD. Additionally, collecting speech samples in multiple languages will help us understand how PD affects speech across different linguistic backgrounds. This approach, coupled with collecting longitudinal data to study disease progression and improve predictive models, is crucial for enhancing the robustness and applicability of our approach. Exploring advanced machine learning techniques, along with additional voice features and multimodal data, such as facial expressions and body movements, will provide a more complete picture of how PD affects individuals. This holistic approach will strengthen our ability to detect the disease early and accurately. Our overall goal is to create a detection method that is effective for early detection, low in cost, easy to use, and capable of being conducted remotely.

References

1. Virmani, T., Lotia, M., Glover, A., Pillai, L., Kemp, A.S., Iyer, A., Farmer, P., Syed, S., Larson-Prior, L.J., Prior, F.W.: Feasibility of telemedicine research visits in people with parkinson's disease residing in medically underserved areas. Journal of Clinical and Translational Science (2022)
2. Govindu, A., Palwe, S.: Early detection of parkinson's disease using machine learning. Procedia Computer Science (2023)
3. Suppa, A., Costantini, G., Asci, F., Di Leo, P., Al-Wardat, M.S., Di Lazzaro, G., Scalise, S., Pisani, A., Saggio, G.: Voice in parkinson's disease: A machine learning study. Frontiers in Neurology (2022)
4. Dinesh, A., He, J.: Using machine learning to diagnose parkinson's disease from voice recordings. IEEE Xplore (2018)
5. Iyer, A., Kemp, A., Rahmatallah, Y., Pillai, L., Glover, A., Prior, F., Larson-Prior, L., Virmani, T.: A machine learning method to process voice samples for identification of parkinson's disease. Scientific Reports (2023)