Cross-Corpus Disparity of Parkinson's Voice Datasets Observed on Control Group Distribution

Nemuel D. Pah
Electrical Engineering Department
University of Surabaya
Surabaya, Indonesia
nemuelpah@staff.ubaya.ac.id
(ORCID: 0000-0002-0181-3199)

Veronica Indrawati
Electrical Engineering Department
University of Surabaya
Surabaya, Indonesia
veronica@staff.ubaya.ac.id
(ORCID: 0000-0003-1036-744X)

Dinesh K. Kumar
School of Engineering
RMIT University
Melbourne, Australia
dinesh@rmit.edu.au
(ORCID: 0000-0003-3602-4023)

2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) | 978-1-6654-5645-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIIC57133.2023.10066982

***Abstract*— Parkinson's disease (PD) is one of the most common neurodegenerative disorders. PD has been the fastest growth in prevalence, and it has become the leading cause of disability. The severity or progression of PD can be reduced if diagnosed at the early stages. It is therefore necessary to develop rapid and simple screening methods or tools to diagnose PD. Speech impairment is one of the early symptoms of PD which is commonly termed Parkinsonian hypokinetic dysarthria. Many researchers have developed a computerized method to identify of diagnosing PD based on voice features. However, the inaccuracy of the developed models was inconsistent especially when being tested on different datasets. The possible cause is the unwanted variability and biases between datasets. This study investigates the possible inconsistencies between Parkinson's voice datasets. The inconsistencies were investigated in the statistical distribution of voice parameters of the healthy-control (HC) group. This work observes the statistical distribution of sustained phoneme parameters extracted from the healthy-control (HC) group of five datasets using ANOVA and the Post-Hoc Turkey-Cramer test. The result suggests that the diversity in language and ethnicity were not contributing significantly to any biases between databases. The other result confirms that noises in the recording contribute to the biases in the extracted voice features, especially the harmonic features**

Keywords*— *Parkinson's Disease, voice features, sustained phoneme, statistical analysis.

I. INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder [1] after Alzheimer's disease. Among neurological disorders, PD has been the fastest growth in prevalence and it has become the leading cause of disability [2]. It is reported by the Global Burden Neurological Disorder of Disease (GBD) study reported in 2019 [3] that the prevalence of PD in 2016 is more than 6 million globally with a 145% increase in the last 26 years. The real number could be higher than six million since PD diagnostic procedures are complex and it is entirely depending on the patients' initiative to report. The prevalence of PD is expected to increase in the coming decade with the increasing portion of the aging population. The rapid growth of PD will increase the burden on the health system since the disability caused by PD increases the expenses to provide drugs, rehabilitation, and caregiving. The severity or progression of PD can be reduced if diagnosed at the early stages. It is therefore necessary to develop rapid and simple screening methods or tools to diagnose PD.

PD symptoms are commonly manifested in motor and non-motor impairments. The current diagnosis of PD is based on clinical assessment of motor and non-motor impairment. The presence of two or more motor symptoms of tremor, rigidity, bradykinesia, postural impairment, or non-motor symptoms (dysarthria, functional impairment, or cognitive impairment) is indicative of the disease. The assessment needs to be performed in a clinical environment by a trained clinician. These assessment procedures limit access to PD diagnosis and possible subjective biases.

Speech impairment is one of the early symptoms of PD [4] which is commonly termed Parkinsonian hypokinetic dysarthria. Approximately 70%–90% of patients with PD show some form of speech impairment [5]. Parkinsonian hypokinetic dysarthria is manifested in a variety of disturbances such as reduced voice intensity, increased voice nasality, increased acoustic noise, reduced speech prosody, imprecise articulation, significantly narrower pitch range, mono loudness, longer pauses, vocal tremor, harsh and breathy voice quality, and disfluency [6][7] he common clinical method for assessing Parkinsonian hypokinetic dysarthria is perceptual evaluation, which however is subjective [8]. Computerized voice analysis has been studied as a more accurate, objective, and quantifiable alternative that can be developed into a rapid and portable Parkinson's screening tool [9]. It also opens the possibility for telehealth and remote monitoring of the patients.

Many researchers have developed a computerized method to identify of diagnosing PD based on voice features. The features were extracted from paragraph reading, diadochokinetic voice, and sustained phonation. The sustained phonation features were less biased by language, the cognitive, and the psychological condition of the subject. Moro-Velázquez [10] developed GMM-UBM and i-Vectors-GPLDA models to detect PD based on voice with an accuracy of 87%. Sakar et al [11] used Q-factor Wavelet Transform features to identify PD patients with 86% accuracy. Almeida et al [12] developed a machine learning classifier with 94.55% accuracy using a voice dataset that was recorded from subjects in Lithuanian. However, the relatively significant achievement in the above studies was not converted easily into a publicly available computer tool or application to diagnose PD. This is due to the inconsistent accuracies of the model especially when being tested on different datasets [13].

Pah et al [14] developed an SVM classification model to diagnose PD patients based on features extracted from sustained phonemes. The study was using a Viswanathan-RMIT dataset of Australian native who speaks English. The study was extended to include the PC-GITA dataset [15] recorded from Columbians who speak Spanish. The result shows a significant performance difference between the two

The This work was supported in part by the PDUPT Kemendikbud ristek 067/SPLit/LPPM01/KemendikbudRistek/Multi/FT/ III/2022 and in part by the University of Surabaya Internal Research Grant..

TABLE I
DESCRIPTION OF DATASETS

Dataset	Participants	Age range (mean ± STD) in years	Phonemes	Ethnicity/ language	Recording • format • equipment • environment	SNR (dB) range (mean ± STD)
Dataset 1: PC-GITA [15]	50 HC (25 M/25 F)	31 - 86 (60.9 ± 11.3)	/a/, /e/, /i/, /o/, /u/	Columbian/ Spanish	- WAV file; fs = 44.1 kHz; 16 bit resolution - dynamic omnidirectional microphone (Shure, SM 63L) - noise controlled-sound proof booth	3.3 – 54.5 (27.0 ± 11.9)
Dataset 2: Saarbruecken Voice Database (SVD)	24 HC (11 M/13 F)	60 - 75 (64.9 ± 4.3)	/a/, /i/, /o/	Germany/ Germany	- WAV files; fs = 50 kHz; 16 bit resolution - N/A - Controlled room [25]	23.3 – 37.4 (31.6 ± 3.9)
Dataset 3: Italian Parkinson's Voice and Speech Dataset [26]	22 HC (10 M/12 F)	60 – 77 (67.1 ± 5.2)	/a/, /e/, /i/, /o/, /u/	Italian/ Italian	- WAV files; fs = 16 kHz; - Common microphone; 15-25 cm from the speaker - Quiet, echo-free room	62.8 – 81.1.6 (70.8 ± 5.0)
Dataset 4: UCI - Parkinson's Disease Classification Dataset [11]	64 HC (23 M/41 F)	41 - 82 (61.1 ± 8.9)	/a/	Turkish/ Turkish	- WAV files; fs = 44.1 kHz; 16 bit resolution; Parameters in CSV format. - Common microphone - N/A (uncontrolled)	N/A
Dataset 5: Viswanathan's-RMIT [27]	22 HC (13 M/9 F)	45 - 75 (66.3 ± 6.2)	/a/, /o/, /m/	Australian/ English	- WAV files; fs = 48 kHz; 16 bit resolution - Omnidirectional head-worn microphone (Samson-SE50) - Noise-restricted room	13.0 – 38.4 (18.7 ± 5.4)

datasets. Karan [13] used Hilbert spectrum features to identify patients with PD. The method was trained with the PC-GITA dataset and achieved 82% - 90% accuracy when tested with the same dataset. However, the accuracy was dropped to the range of 50% – 80% when being tested on another dataset.

The other handicap to developing a more complex machine learning model for the task is the limited size of the available databases. To develop a deep-learning model, a large dataset that captures the wide spectrum of the PD is needed during the learning phase of the deep-learning model. One of the possible solutions is by accumulating the available datasets into a large training set. However, this approach may introduce another problem due to the unwanted variability and biases between the datasets. The variability may be due to language, noises, recording protocols, ethnicity, or other factors.

This study investigates the possible inconsistencies between Parkinson's voice datasets. The inconsistencies were investigated based on the statistical distribution of voice parameters extracted from the recordings of the healthy-control (HC) group of the datasets. The results of this study may be used as the basis to select datasets of research in the area and may indicate the aspect that contributes to the disparity.

II. MATERIALS AND MENTHODS

To investigate the disparity of the Parkinson's voice datasets and the possible external factors that contribute to the disparity, this work observes the distribution of sustained phoneme parameters extracted from the healthy-control (HC) group of five datasets. It is assumed that if the HC group of the datasets were isolated from any symptoms of PD or other pathological conditions related to voice parameters and were not biased by any demographic and technical factors then the distribution of their voice parameters should be relatively similar. Any variation in the distribution of the parameter indicates the existence of biases in the dataset.

A. Datasets

Five Parkinson's or pathological voice datasets were used in this work as shown in Table I. The datasets recorded sustained phoneme/vowels, words, sentences, and repeating syllable/diadochokinetic recordings. However, for this work, only sustained phoneme recordings were considered. The datasets were the PC-GITA, Saarbruecken Voice Database (SVD), Italian Parkinson's Voice and Speech Dataset, UCI Parkinson's Disease Classification Dataset of Okan Sakar, and Viswanathan's dataset of RMIT University. Out of 869 healthy participants in the SVD dataset, only participants with an age range of 60-75 years old were selected to be included in this work. The number of healthy participants in each dataset were ranging from 22 to 50 participants. Based on ANOVA analysis of the age distribution of the five datasets, the datasets can be categorized into two groups. The participants in dataset 1 and dataset 4 were younger than the participants in dataset 2, 3, and 5. Datasets 1, 2, 3, and 5 were provided in the '.wav' format. UCI Parkinson's Disease Classification dataset was only available in the format of extracted parameters in a '.csv' file.

Datasets 2 and 3 recorded sustained phonemes from all five vowels. Dataset 2 and dataset 5 have only the recording from 3 phonemes, while dataset 4 only provides parameters from sustained phoneme /a/. For this reason, the comparison reported in this work was only based on the recording of the phoneme /a/. The ethnicity, language, and recording condition were detailed in Table I. Most datasets were recorded on a relatively noise-controlled condition with an average SNR of

TABLE II
LIST OF VOICE FEATURES EXTRACTED FROM THE RECORDINGS

No	Feature (unit)	Description
1	Jitter-abs (s)	Absolute time perturbation of glottal pulses
2	Jitter-rel (%)	Relative time perturbation of glottal pulses
3	Shimmer-rel (%)	Relative amplitude perturbation of glottal pulses
4	Shimmer-abs (dB)	Absolute amplitude perturbation of glottal pulses
5	HNR (dB)	Harmonics-to-noise ratio
6	NHR	Noise-to-harmonics ratio
7	F1-mean (Hz)	Mean of formants F1 frequency
8	F2-mean (Hz)	Mean of formants F2 frequency
9	F3-mean (Hz)	Mean of formants F3 frequency
10	F4-mean (Hz)	Mean of formants F4 frequency
11	Intensity-mean (dB)	Mean of voice intensity
12–24	MFCC (13 coefficients)	Mell Frequency Cepstral Coefficients

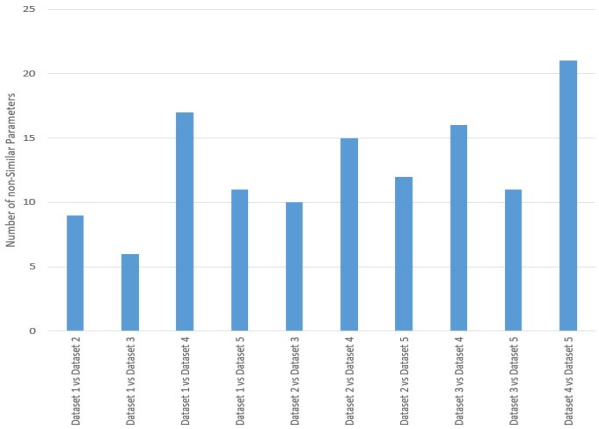

Fig. 1. The number of parameters that produce a p-value of less than 0.05 for each pair of the Post-Hoc Turkey-Cramer test.

18.7 dB – 70.8 dB. The SNR of the datasets was calculated as the ratio between the signal power and the noise power (segment of the recordings with only the background noise). Because the recordings in datasets 1, 2, and 3 were cropped to the voice segments only, the noise segment was taken from the silent segment of the corresponding sentences recording from the dataset. The SNR of dataset 4 was not available.

B. Feature Extraction

Twenty-four features, as shown in Table II, were extracted from sustained phoneme /a/ in datasets 1, 2, 3, and 5 using a code developed based on Praat [16], a publicly available speech analysis software. The corresponding 24 features were sorted from the 754 features provided in dataset 4. The 24 features were selected to represent the whole aspect of sustained phoneme production. Jitter, Shimmer, and the harmonics features were related to the magnitude [17], [18], time-frequency perturbation, and noise of vocal cord vibration. The formants (F1 to F4) [19] and the MFCCs [20] represent the condition of vocal tract modulation. The mean intensity captured the condition of the lung as the power source of sustained phoneme production.

C. Statistical Analysis

To investigate the disparity of the dataset, one-way ANOVA [21] was applied to each extracted parameter in the datasets to determine any statistical differences among the datasets. A significance level of $\alpha = 0.05$ was used to test the differences. A p-value of less than 0.05 indicates the rejection of the H0 hypothesis that all the dataset's distributions were similar. To identify which particular dataset creates the differences, the Post-Hoc Turkey-Cramer test [22] was performed between pairs of the datasets. A p-value of less than 0.05 was used to indicate the significant difference between the pairs. The statistical analyses were performed using MATLAB 2018b (MathWorks).

III. RESULTS

Table III presents the result of the ANOVA test and the corresponding Post-Hoc Turkey-Cramer test for each parameter. The '< 0.05' indicates a p-value of less than 0.05 and hence suggests that the mean and distribution of the datasets for the particular parameter were not similar. On the other hand, the dash '-' sign indicates a p-value greater or equal to 0.05. In this case, the mean/distribution of the tested datasets was considered similar.

The p-values of the ANOVA test were below 0.05 for all the parameters. The difference was mostly contributed by the difference between dataset 4 and dataset 5. The p-value between these two datasets in almost all parameters was below 0.05. The result shows that dataset 1 and dataset 3 form the most match. The distribution of datasets 1 and 3 was statistically similar for all the parameters, except the MFCCs.

Fig. 1 shows the number of parameters that produce a p-value of less than 0.05 for each pair of the Post-Hoc Turkey-Cramer test. The lowest three pairs are the combination of datasets 1, 2, and 3. This result suggests that the three datasets can be assumed to have a relatively similar distribution. On the other extreme, the highest four pairs are the comparison between dataset 4 to all the other datasets. It suggests that dataset 4 is the most alienated dataset among the investigated datasets. Dataset 5 is closer to dataset 1,2, and 3.

The distribution of Jitter and Shimmer of datasets 1, 2, and 3 were similar (p-value > 0.05). The harmonics parameters distribution confirms that dataset 5 was different from the other four datasets. Except for the p-value between dataset 4 and dataset 5, the distribution of formants of the datasets was relatively closer. There was mostly no similarity between the MFCC of the five datasets, especially in the lowest-order coefficients.

IV. DISCUSSION

Many earlier studies have suggested the possible use of using voice parameters extracted from sustained phonemes to identify people with Parkinson's disease. It provides the possibility of a non-invasive method of PD diagnosis which is more comfortable for patients, faster, as well as cost-effective [23]. Many classification models, including machine learning and deep learning models, have been developed to identify PD based on parameters extracted from the patient's voices. However, these models still suffer low classification accuracy especially due to their robustness and generalization performance across the wide demographic spectrum of the population with PD.

Most of the studies on this topic, develop their model using relatively small databases with a limited demographic range.

TABLE III
THE RESULT OF ANOVA TEST AND POST-HOC TURKEY-CRAMER TEST

Parameter	p-value of ANOVA	p-value of Post-Hoc Test (Turkey-Cramer)									
	All Datasets	Dataset 1 vs Dataset 2	Dataset 1 vs Dataset 3	Dataset 1 vs Dataset 4	Dataset 1 vs Dataset 5	Dataset 2 vs Dataset 3	Dataset 2 vs Dataset 4	Dataset 2 vs Dataset 5	Dataset 3 vs Dataset 4	Dataset 3 vs Dataset 5	Dataset 4 vs Dataset 5
Jitter-abs (s)	< 0.05	-	-	< 0.05	-	-	< 0.05	< 0.05	< 0.05	-	< 0.05
Jitter-rel (%)	< 0.05	-	-	< 0.05	-	-	< 0.05	< 0.05	< 0.05	-	< 0.05
Shimmer-rel (%)	< 0.05	-	-	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Shimmer-abs (dB)	< 0.05	-	-	< 0.05	-	-	< 0.05	-	< 0.05	-	< 0.05
HNR (dB)	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05
NHR	< 0.05	-	-	-	< 0.05	-	-	< 0.05	-	< 0.05	< 0.05
F1-mean (Hz)	< 0.05	< 0.05	-	< 0.05	< 0.05	-	-	-	-	< 0.05	-
F2-mean (Hz)	< 0.05	< 0.05	-	-	< 0.05	-	< 0.05	< 0.05	-	< 0.05	< 0.05
F3-mean (Hz)	< 0.05	-	-	-	-	-	-	-	-	-	< 0.05
F4-mean (Hz)	< 0.05	-	-	-	-	-	< 0.05	-	< 0.05	-	< 0.05
Intensity-mean (dB)	< 0.05	-	-	< 0.05	-	< 0.05	< 0.05	-	< 0.05	-	< 0.05
MFCC(0)	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05
MFCC(1)	< 0.05	< 0.05	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
MFCC(2)	< 0.05	< 0.05	-	-	< 0.05	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05
MFCC(3)	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05	-	< 0.05	< 0.05	< 0.05
MFCC(4)	< 0.05	< 0.05	-	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05	-	< 0.05
MFCC(5)	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	-	-	< 0.05
MFCC(6)	< 0.05	-	< 0.05	< 0.05	-	< 0.05	< 0.05	-	-	< 0.05	< 0.05
MFCC(7)	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	-	-	< 0.05	< 0.05	< 0.05	< 0.05
MFCC(8)	< 0.05	-	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	-	-
MFCC(9)	< 0.05	-	-	< 0.05	-	-	-	-	-	-	< 0.05
MFCC(10)	< 0.05	-	-	-	-	-	-	< 0.05	-	-	< 0.05
MFCC(11)	< 0.05	-	-	-	-	-	-	-	< 0.05	-	-
MFCC(12)	< 0.05	-	-	< 0.05	-	-	-	-	< 0.05	-	< 0.05

The development of a robust model requires a training process that includes a large database that covers all the possible demographic and pathological ranges of PD patients. Such a database is rarely available due to the limited access to patients with PD, especially in developing countries.

One of the possible solutions to the problem of dataset limitation is to accumulate a large number of recordings from all the available datasets of PD voice datasets or other pathological voice datasets. However, combining a large number of datasets in a training process may introduce a new problem due to biases between the datasets. The unwanted biases may be related to the variation of recording techniques and equipment, language and ethnicity, noises in the recording environment, and other unknown factors.

In this study, the authors investigate the disparity between five commonly used Parkinson's voice datasets (the PC-GITA, Saarbruecken Voice Database (SVD), Italian Parkinson's Voice and Speech Dataset, UCI Parkinson's Disease Classification Dataset of Okan Sakar, and Viswanathan's dataset of RMIT University) based on the distribution of voice parameters extracted from the recordings of healthy-control (HC) group of the datasets.

The language and ethnic diversity of the subjects have been suspected as the possible source of biases between Parkinson's voice datasets [14]. The result of ANOVA and Post-Hoc Turkey-Cramer tests in this work suggest that the diversity in language and ethnicity were not contributing significantly to any biases between databases of the sustained phoneme, especially in the features related to vocal cord vibration (Jitter and Shimmer). Datasets 1, 2, and 3 were recorded from three different populations with different languages and ethnicity. The subjects recorded in PC-GITA datasets were Columbians that speak Spanish. The SVD database subjects were German who speak the German language, while the Italian Parkinson's and Voice dataset was recorded from Italian subjects. The statistical analysis shows that the p-value of Jitter and Shimmer features extracted from these three datasets was higher than 0.05 which indicates a condition of a non-significant difference. The formants distribution of these three datasets was also similar. It indicated that the was no significant difference in the pronunciation of /a/ across different languages.

The other finding from this study is that the widely used UCI Parkinson's Disease Classification Dataset (dataset 4) [11]was having a different distribution compared to the other four datasets in almost all the investigated voice features. The features in the dataset were extracted from the recordings of Turkeys' subjects speaking Tukey's language using a microphone operated at a 44.1 kHz sampling rate. Because the dataset was not provided in the form of raw voice recordings, the source of this unique distribution could not be pinpointed in this study. Any study in the future that uses this dataset in combination with other datasets must be done with the awareness of this unique distribution.

Noises introduced during the recording phase have been suggested as a possible reason for biases in Parkinson's voice dataset. The result of this study confirms that noises in the recording contribute to the biases in the extracted voice features especially the harmonic features (HNR and NHR). Dataset 5 (the Viswanathan's-RMIT dataset) was recorded at the lowest SNR of 18.7 ± 5.4 dB. The HNR and NHR distribution of this dataset were different (with p-value < 0.05) from the other four datasets.

The distributions of MFCC coefficients of the five datasets were not similar. Most of the p-values of the MFCC comparison were less than 0.05, especially in the low-order

coefficients. This is because the datasets were recorded with different sampling rates as shown in Table I. Two recordings with different sampling frequencies will have Mel-filters with different frequency ranges (in Hz or Mels) if the number of coefficients was kept the same [24]. Therefore, it is necessary to consider the same sampling frequency when using MFCC coefficients as the extracted features.

V. Conclusion

To develop a robust and accurate machine learning model to identify subjects with Parkinson's disease based on voice features, it is necessary to have a large dataset of voice recordings. Accumulating voice recordings from several available datasets could face a problem due to biases between datasets. This study investigates the possible biases between five commonly used Parkinson's voice datasets. The inconsistencies were observed in the statistical distribution of the sustained phoneme features of healthy control (HC).

The results suggest that, for sustained phoneme features, language and ethnic variation do not produce a significant bias between datasets. On the other hand, noises in the recording's environment have a significant contribution to producing bias between datasets. If the dataset extract MFCC features, the sampling frequency of the recordings must be normalized to the same frequency.

The other significant finding in this work is the unique distribution of the UCI Parkinson's Disease Classification Dataset to the other investigated dataset in this work. Accumulating UCI Parkinson's Disease Classification Dataset into a larger dataset has to be done with this uniqueness awareness.

References

[1] Z. Ou et al., "Global Trends in the Incidence, Prevalence, and Years Lived With Disability of Parkinson's Disease in 204 Countries/Territories From 1990 to 2019," Front. Public Heal., vol. 9, no. December, 2021, doi: 10.3389/fpubh.2021.776847.

[2] "Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017," Lancet, vol. 392, no. 10159, pp. 1789–1858, 2018, doi: 10.1016/S0140-6736(18)32279-7.

[3] W. M. Carroll, "Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," Lancet Neurol., vol. 18, no. 5, pp. 459–480, 2019, doi: 10.1016/S1474-4422(18)30499-X.

[4] B. Harel, M. Cannizzaro, and P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease : A longitudinal case study," Brain Cogn., vol. 56, pp. 24–29, 2004, doi: 10.1016/j.bandc.2004.05.002.

[5] J. Rusz et al., "Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder," Sleep Med, vol. 19, pp. 141–147, 2016, doi: 10.1016/j.sleep.2015.07.030.

[6] J. Rusz, T. Tykalova, L. O. Ramig, and E. Tripoliti, "Guidelines for Speech Recording and Acoustic Analyses in Dysarthrias of Movement Disorders," Mov. Disord., vol. 36, no. 4, pp. 803–814, 2021, doi: 10.1002/mds.28465.

[7] T. Tuncer and S. Dogan, "A novel octopus based Parkinson's disease and gender recognition method using vowels," Appl. Acoust., vol. 155, pp. 75–83, 2019, doi: 10.1016/j.apacoust.2019.05.019.

[8] S. Lahmiri and A. Shmuel, "Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine," Biomed. Signal Process. Control, vol. 49, pp. 427–433, 2019, doi: 10.1016/j.bspc.2018.08.029.

[9] N. D. Pah, M. A. Motin, and D. K. Kumar, "Phonemes based detection of parkinson's disease for telehealth applications," Sci. Rep., vol. 12, no. 1, pp. 1–9, 2022, doi: 10.1038/s41598-022-13865-z.

[10] L. Moro-Velázquez, J. A. Gómez-García, J. I. Godino-Llorente, J. Villalba, J. R. Orozco-Arroyave, and N. Dehak, "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's Disease," Appl. Soft Comput., vol. 62, pp. 649–666, 2018, doi: 10.1016/j.asoc.2017.11.001.

[11] C. O. Sakar et al., "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," Appl. Soft Comput. J., vol. 74, 2019, doi: 10.1016/j.asoc.2018.10.022.

[12] J. S. Almeida et al., "Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques," Pattern Recognit. Lett., vol. 125, pp. 55–62, 2019, doi: 10.1016/j.patrec.2019.04.005.

[13] B. Karan, S. S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, "Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech," Biomed. Signal Process. Control, vol. 61, p. 102050, 2020, doi: 10.1016/j.bspc.2020.102050.

[14] N. D. Pah, M. A. Motin, P. Kempster, and D. K. Kumar, "Detecting effect of levodopa in Parkinson's disease patients using sustained phonemes," IEEE J. Transl. Eng. Heal. Med., vol. Article in, 2021.

[15] J. R. Orozco-Arroyave, J. D. Arias-Ledono, J. F. Vargas-Bonilla, and M. C. Gonzalez-Rativa, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in International Conference on Language Resources and EvaluationAt: Reykjavik, Iceland, 2014, no. May 2014.

[16] B. P. Boersma and V. Van Heuven, "Speak and unSpeak with P RAAT," Glot Int., vol. 5, no. 9–10, pp. 341–347, 2001.

[17] J. P. Teixeira and A. Gonçalves, "Accuracy of Jitter and Shimmer Measurements," Procedia Technol., vol. 16, pp. 1190–1199, 2014, doi: 10.1016/j.protcy.2014.10.134.

[18] J. P. Teixeira, C. Oliveira, and C. Lopes, "Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters," Procedia Technol., vol. 9, pp. 1112–1122, 2013, doi: 10.1016/j.protcy.2013.12.124.

[19] J. Lee, S. Shaiman, and G. Weismer, "Relationship between tongue positions and formant frequencies in female speakers," J. Acoust. Soc. Am., vol. 139, no. 1, pp. 426–440, 2016, doi: 10.1121/1.4939894.

[20] F. T. Al-Dhief et al., "Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine," IEEE Access, vol. 9, pp. 77293–77306, 2021, doi: 10.1109/ACCESS.2021.3082565.

[21] S. Yakub and M. Arowolo, "A Feature Selection Based on One Way Anova For Microarray Data Classification A Feature Selection Based on One-Way-Anova for Microarray Data Classification," no. September, 2019.

[22] S. J. Richter and M. H. Mccann, "Using the Tukey-Kramer omnibus test in the Hayter-Fisher procedure," Br. J. Math. Stat. Psychol., vol. 65, no. 3, pp. 499–510, 2012, doi: 10.1111/j.2044-8317.2012.02041.x.

[23] V. Illner, P. Sovka, and J. Rusz, "Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease," Biomed. Signal Process. Control, vol. 58, p. 101831, 2020, doi: 10.1016/j.bspc.2019.101831.

[24] O. Lowna Baroi, M. S. Abrar Kabir, A. Niaz, M. Jahidul Islam, and M. Jakaria Rahimi, "Effects of Filter Numbers and Sampling Frequencies on the Performance of MFCC and PLP based Bangla Isolated Word Recognition System," Int. J. Image, Graph. Signal Process., vol. 11, no. 11, pp. 36–42, 2019, doi: 10.5815/ijigsp.2019.11.05.

[25] P. Harar, J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, R. Burget, and Z. Smekal, "Voice Pathology Detection Using Deep Learning: A Preliminary Study," in 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI), 2017. doi: 10.1109/IWOBI.2017.7985525.

[26] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system," IEEE Access, vol. 5, pp. 22199–22208, 2017, doi: 10.1109/ACCESS.2017.2762475.

[27] R. Viswanathan et al., "Complexity Measures of Voice Recordings as a Discriminative Tool for Parkinson's Disease," Biosens., vol. 10, no. 1, 2019, doi: 10.3390/bios10010001.