Disease Delineation for Multiple Sclerosis, Friedreich Ataxia, and Healthy Controls Using Supervised Machine Learning on Speech Acoustics

Benjamin G. Schultz, *Member, IEEE*, Zaher Joukhadar, Usha Nattala, Maria del Mar Quiroga, Gustavo Noffs, Sandra Rojas, Hannah Reece, Anneke Van Der Walt, and Adam P. Vogel, *Member, IEEE*

***Abstract*— Neurodegenerative disease often affects speech. Speech acoustics can be used as objective clinical markers of pathology. Previous investigations of pathological speech have primarily compared controls with one specific condition and excluded comorbidities. We broaden the utility of speech markers by examining how multiple acoustic features can delineate diseases. We used supervised machine learning with gradient boosting (CatBoost) to delineate healthy speech from speech of people with multiple sclerosis or Friedreich ataxia. Participants performed a diadochokinetic task where they repeated alternating syllables. We subjected 74 spectral and temporal prosodic features from the speech recordings to machine learning. Results showed that Friedreich ataxia, multiple sclerosis and healthy controls were all identified with high accuracy (over 82%). Twenty-one acoustic features were strong markers of neurodegenerative diseases, falling under the categories of spectral qualia, spectral power, and speech rate. We demonstrated that speech markers can delineate neurodegenerative diseases and distinguish healthy speech from pathological speech with high accuracy. Findings emphasize the importance of examining speech outcomes when assessing indicators of neurodegenerative disease. We propose large-scale initiatives to broaden the scope for differentiating other neurological diseases and affective disorders.**

***Index Terms*— Neurodegenerative disease, speech acoustics, machine learning, multiple sclerosis, Friedreich ataxia, dysarthria, speech science.**

Manuscript received 23 January 2023; revised 20 September 2023; accepted 28 September 2023. Date of publication 4 October 2023; date of current version 1 November 2023. This work was undertaken in collaboration with the Melbourne Data Analytics Platform (MDAP) at The University of Melbourne. Data collection for the multiple sclerosis group was supported by a National Health and Medical Research Council (NHMRC) Project grant (#108546). Adam P. Vogel was supported by a NHMRC Fellowship (#1135683) and an Australian Research Council Future Fellowship (#220100253). *(Corresponding author: Benjamin G. Schultz.)*

Benjamin G. Schultz was with the Department of Audiology and Speech Pathology, The University of Melbourne, Carlton, VIC 3055, Australia. He is now with PSI Connect, Melbourne, VIC 3060, Australia, and also with Escient, Melbourne, VIC 3000, Australia (e-mail: ben.schultz@psiconnect.org; ben.schultz@escient.com.au).

Zaher Joukhadar, Usha Nattala, and Maria del Mar Quiroga are with the Melbourne Data Analytics Platform, The University of Melbourne, Carlton, VIC 3055, Australia (e-mail: zaher.joukhadar@unimelb.edu.au; usha.nattala@unimelb.edu.au; mar.quiroga@unimelb.edu.au).

Gustavo Noffs was with the Department of Audiology and Speech Pathology, The University of Melbourne, Carlton, VIC 3055, Australia. He is now with the Department of Medicine, Nursing and Health Sciences, Monash University, Clayton, VIC 3168, Australia (e-mail: gustavo.noffs@monash.edu).

Sandra Rojas was with the Department of Audiology and Speech Pathology, The University of Melbourne, Carlton, VIC 3055, Australia. She is now with the Escuela de Fonoaudiología, Facultad de Odontología y Ciencias de la Rehabilitación, Universidad San Sebastián, Santiago 8340593, Chile (e-mail: sandra.rojas@uss.cl).

Hannah Reece was with the Department of Audiology and Speech Pathology, The University of Melbourne, Carlton, VIC 3055, Australia (e-mail: hannahreece@ymail.com).

Anneke Van Der Walt is with the Bruce Lefroy Centre, Murdoch Children's Research Institute, The Royal Children's Hospital Melbourne, Parkville, VIC 3052, Australia, and also with the Central Clinical School, Monash University, Melbourne, VIC 3004, Australia (e-mail: anneke.vanderwalt@monash.edu).

Adam P. Vogel is with the Department of Audiology and Speech Pathology, The University of Melbourne, Carlton, VIC 3055, Australia, also with the Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany, and also with Redenlab, Melbourne, VIC 3000, Australia (e-mail: vogela@unimelb.edu.au).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3321874, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3321874

I. Introduction

NEURODEGENERATIVE disease can alter speech due to impaired motor control and execution. Acoustic features of speech can be used as objective clinical markers for diseases of the central nervous system (CNS). Previous studies examining acoustic changes in neurodegenerative disease have primarily focused on differences between healthy controls and various patient populations, such as, multiple sclerosis (MS) [1], [2], [3], [4], [5], Huntington's disease (HD) [6], [7], [8], Parkinson's disease (PD) [4], [5], [9], [10], and Friedreich ataxia (FA) [9], [11], [12]. These studies report that various acoustic features change as the disease progresses, and patients tend to exhibit slower and more variable speech rates, lower and more variable pitch, and reduced spectral clarity compared to patients. Although machine learning has been used to differentiate healthy controls from single, well-defined patient populations (e.g., Parkinson's disease, spasmodic dysphonia), real-world machine learning implementations will encounter multiple different diseases that may have overlapping acoustic profiles. The present study aims to broaden the utility of these speech markers by determining how different acoustic profiles of speech may accurately identify specific neurodegenerative diseases. We will move beyond discriminating between

healthy and pathological speech by examining differences between different neurodegenerative diseases with similar speech phenotypes (ataxia) simultaneously across multiple acoustic dimensions.

Clinicians use a combination of tools to diagnose neurodegenerative disease including genetic sequencing, neurological scans (e.g., magnetic resonance imaging), and neuropsychological and motor tests. Comorbidities across modalities makes differential diagnosis or genetic test selection (where possible) challenging and can exacerbate the length of time to correct diagnosis (e.g., [13], [14]). Clinical acoustic markers have several advantages over traditional tools. First, speech can be recorded remotely in a home environment without the need to visit a specialist or hospital. Given that some populations with neurodegenerative disease are considered at-risk, remote identification reduces the risk of contracting potentially life-threatening pathogens. Remote testing is also more accessible for populations with limited mobility and those living in rural areas. Second, acoustic markers are obtained using non-invasive techniques. Invasive procedures (e.g., blood tests, surgery) can cause discomfort and have a risk of infection. These procedures can also impose large financial burdens, particularly when multiple tests are required due to misdiagnosis. Acoustic markers have the potential to alleviate these burdens and streamline processes by providing accessible, low-cost, and low-risk tests that can guide clinician decisions in the early stages of diagnosis.

Acoustic features of speech can be used to construct profiles of different patient populations. The most common acoustic features used to identify pathologies include speech rate, the number of syllables per second, pauses, the duration between utterances or syllables, and frequency information related to the pitch of the voice (fundamental frequency; *f0*) and its formants [15], [16]. These features reflect underlying cortical, subcortical, or cerebellar pathology of clinical populations leading to multiple speech subsystem impairments [7].

Speech can be elicited through specific tasks or naturalistic settings. The diadochokinetic (DDK) task is a common task in which the speaker repeats a syllable string. (e.g., /PATAKA/) as quickly and clearly as possible for 10 seconds [17]. The DDK task is a controlled method of speech elicitation that allows high consistency between different speakers while remaining sensitive to speech performance [18]. Although other speech tasks (e.g., reading, semi-structured interviews) increase ecological validity, they may also increase cognitive load, which may induce speech changes based on individual differences like education level, language or reading impairments, or cognitive ability [19], [20]. Moreover, the linguistic content may encourage changes in prosodic features based on emphasis, stress patterns, and emotion which may differ based on personality, accent, or emotional state. To avoid these concerns, the present study examined speech from a DDK task that was performed by healthy controls (HC) and two patient populations (FA, MS) using uniform practices (see Methods). We calculated acoustic features that have been examined in previous studies comparing HCs and various patient populations (1–9) and include several new acoustic features related to speech quality [15] and speech timing [21], [22] that may improve the differentiation of these groups.

Previous implementations of machine learning on speech have compared healthy controls with only a single patient group (cf. [16]). For example, machine learning approaches using acoustic features have shown high accuracy (>90%) when differentiating healthy control groups from patient groups with Parkinson's disease [23], spasmodic dysphonia [24], and various other vocal conditions (e.g., oral cancer or vocal fold nodules) [16]. Although these approaches are useful as initial triage for identifying pathological voice disturbances that require further investigation [25], they do not provide nuanced classification of the underlying pathology or disease phenotypes potentially due to small sample sizes and, consequently, low accuracy [26]. This is especially relevant for deep learning models with hidden layers that reflect latent variables that are not defined and, therefore, do not aid in developing specific acoustic profiles that may characterize a disease [27]. We used an interpretable machine learning approach using gradient boosting that quantifies the contribution of each acoustic feature in distinguishing between healthy and pathological voices, and between multiple diseases [28].

II. Methods

A. Participants

Healthy controls were recruited through advertisements within Australia. Clinical groups were recruited through medical centers in Australia. We recruited people diagnosed with multiple sclerosis ($N = 112$) and Friedreich ataxia ($N = 73$) as well as healthy controls ($N = 229$). All patient participants were diagnosed by a physician and confirmed genetically for patients with Friedreich ataxia. Demographic information of participant groups is shown in Table I. Some participants were recorded on more than one occasion, leading to a larger final number of speech tokens per group: Multiple sclerosis ($N = 787$), Friedreich ataxia ($N = 158$), and healthy controls ($N = 483$). To ensure that results were not driven by profiles of individuals, data were averaged over participants, resulting in one data point per participant for training and test phases (see [29]).

B. Apparatus

A condenser headset microphone (AKG C520, AKG Acoustic, Vienna, Austria) positioned 8-10cm from the mouth at an angle of 45° recorded speech. A Roland Quad-capture external soundcard connected to a Dell laptop using captured speech through Audacity software [30] and Redenlab ®software at a sampling rate of 44.1kHz.

C. Procedure

Participants performed a DDK task where the syllables /PA/, /TA/, and /KA/ were repeated in an alternating fashion as many times as possible within one breath for a maximum of 10 seconds. Speech recordings were screened prior to feature analysis to manually remove speech artefacts and background noise.

D. Acoustic Feature Extraction

Acoustic features were extracted using custom-made MATLAB scripts that used standard signal processing functions from MATLAB [31], onset and offset detection algorithms [32], beat detection algorithms [22], music information

TABLE I
DEMOGRAPHIC INFORMATION FOR HEALTHY CONTROLS (HC), FRIEDREICH ATAXIA (FA), AND MULTIPLE SCLEROSIS (MS)

Variable	Statistic	Group		
		HC	FA	MS
Age (years)	Mean	56.44	38.78	47.04
	SD	14.61	13.04	11.42
	Min	21	7	22
	Max	92	68	71
Sex	Female	126	39	83
	Male	104	41	29
Disease duration (years)	Mean	NA	22.89	13.66
	SD	NA	12.67	9.14
Severity score*	Mean	NA	79.67	3.36
	SD	NA	35.88	2.15

*FA = Friedreich ataxia rating scale (between 0 and 117), MS = Expanded disability status scale (between 0 and 10). These are not comparable measures of severity and were administered after diagnosis of the disease.

retrieval [33], and speech analysis toolboxes [34], [35]. Acoustic features consisted of summary statistics (mean, standard deviation, coefficient of variation, minimum, maximum, range) of 74 variables that measure different aspects of speech qualia [15], resulting in an initial set of 444 features. These features include speech rate, utterance duration, pause duration, fundamental frequency, the first five formants, intensity, summed and peak energy across frequency bands, spectral decrease and spread, and a range of other spectral features used in the clinical acoustic marker literature (see Supplementary Materials for a full list and additional references).

The acoustic features used in the present study and their physiological and perceptual correlates have previously been described in detail in comprehensive reviews [15], [36], [37]. The fundamental frequency (*f0*) is the lowest frequency of a periodic waveform and is perceived as the pitch of a voice [38]. Formants are the resonant frequencies in the vocal tract that contribute to the timbre and quality of a voice [39].

Other measures of speech and sound quality were also used [40], [41]. Spectral centroid measures the center of mass of the frequency spectrum. Spectral slope is the slope of the linear regression line over across the spectral amplitude values. Spectral flatness measures the uniformity of the frequency spectrum. Spectral decrease is the reduction in signal magnitude across higher frequency bands. Spectral spread measures the distribution of frequencies around the mean frequency in the spectrum. Spectral skew measures the asymmetry in the spectral distribution around its mean frequency. Spectral kurtosis measures the shape of the spectral distribution, indicating the presence of heavy tails or peaks. Spectral crest is the peak amplitude in a frequency spectrum, indicating its highest point. Spectral entropy is the degree of randomness in the distribution of spectral components. Spectral flux is the rate of local change of spectral magnitude and reflects shifts in energy distribution over time.

Correlates of perceived loudness included acoustic intensity, amplitude, the alpha ratio, and energy as measured by wavelet analysis. Intensity is the power of a sound wave per unit area, perceived as loudness [42]. The alpha ratio is the ratio of energy below 1kHz and between 1-4kHz [43]. Amplitude is the magnitude of the maximum displacement of a wave from its equilibrium position and is also perceived as loudness [44]. Acoustic energy (measured here by Morlet wavelets) is the quantification of sound energy across different frequency components using Morlet wavelet transforms [45]. Both the summed and peak energy within frequency bands were measured, as was the frequency at which the energy peaked. Five frequency bands consisting of sub-bands between 1Hz and 8,000Hz were examined, specifically 1Hz to 4,000Hz (i.e., the "broad frequency range"), 75Hz to 500Hz (i.e., the "*f0* frequency range"), 4kHz to 8kHz (i.e., the articulator and expiration spectrum or "high frequency range"), 75Hz to 4,000Hz (i.e., the common vocal frequency spectrum or "mid frequency range"), and 1Hz to 75Hz (i.e., the articulatory-unit and speech-unit range or "low frequency range"). The latter was measured to obtain energy metrics for articulation and speech rate [22], [46], [47].

Speech rate was also measured by determining the onsets and offsets of speech syllables based on amplitude, intensity, spectral flux, and the summed and peak energy in the five frequency bands described above. Onsets and offsets were obtained using the Schultz Musical Instrument Digital Interface Toolbox applied to the time series of these features [32], [48]. From the onsets and offsets, we determined speech rate (i.e., the time difference between consecutive onsets), speech duration (i.e., the time between the onset and offset of speech), and pause duration (i.e., the time between the offset of speech and the next onset). The stress rate of speech was also measured using a beat tracking algorithm that measures recurrent moments of increased energy [22], [49].

E. Machine Learning Procedure

We used CatBoost as our machine leaning classification algorithm. CatBoost is an open-source decision tree-based algorithm with gradient boosting and hardware optimization [50]. The main advantage of CatBoost over other algorithms is that it builds symmetric trees, employs weighted sampling, and performs ordered boosting. It also lowers the weights of variables that are less useful in identifying groups. These features decrease the need for hyperparameter tuning and reduces the chance of overfitting [50]. Cross-validation was performed using 67%-33% Train-Test splits with 100 resamples using stratification to achieve the same balance for each class [28].

F. Statistical Analysis

One-sample t-tests were conducted on Matthew's correlation coefficients and $f1$ scores to assess if model performance surpassed chance levels. Effect sizes were measured using Cohen's d. Performance differences between groups were analyzed using an analysis of variance with group as a fixed

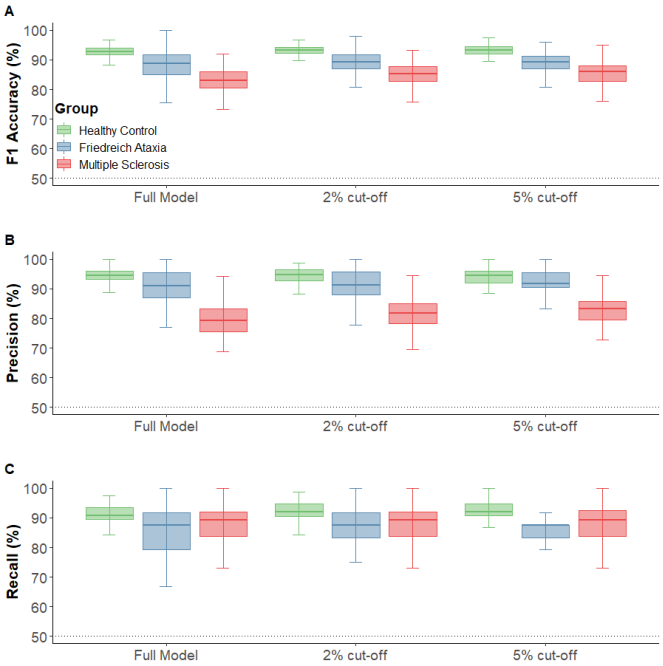

Fig. 1. Mean, dispersion, and range for classification accuracy (A = F1 Accuracy, B = Precision, C = Recall) for healthy controls, and groups with Friedreich ataxia and multiple sclerosis for the full model, and models using a subset of acoustic features using the maximum group-wise average SHAP value cut-offs of 2% and 5%.

factor and resample number as the random factor. Effect sizes for group differences were measured using generalized eta squared (η^2_G). Spearman rank-order correlations were used to assess relationships between the acoustic features and disease severity scores (see Supplementary Materials). All analyses were performed using R software [51].

III. Results

A one-sample t-test revealed that overall model performance as assessed by Matthews's correlation coefficient ($M = 0.82$, $SD = 0.04$) was significantly above chance (33% accuracy), $t\ (99) = 114.56$, $p < 0.001$, Cohen's $d = 11.46$. Classification accuracy between groups was assessed using $f1$ scores that equally weight model specificity and sensitivity (see Supplementary Materials for full statistical analysis); these were also significantly better than chance for all groups ($ps < 0.001$) with large effect sizes for HC (Cohen's $d = 32.29$), MS (Cohen's $d = 11.82$), and FA (Cohen's $d = 10.70$). There was a significant main effect of group, $F\ (2, 198) = 238.20$, $p < 0.001$, $\eta^2_G = 0.49$. As shown in Figure 1, classification accuracy was higher for HC compared to FA ($p < 0.001$) and MS ($p < 0.001$), and higher for FA compared to MS ($p < 0.001$). Receiver operating curves (ROC) for the model with average performance based on Matthew's correlation coefficients are shown in Figure 2. The ROC area under the curve (AUC) values were 0.97 for HC, 0.98 for FA, and 0.96 for MS. These values indicate outstanding discrimination by the model [52].

A. Model Optimization

To measure the contribution of each acoustic feature for categorizing each group, the Shapley additive explanation

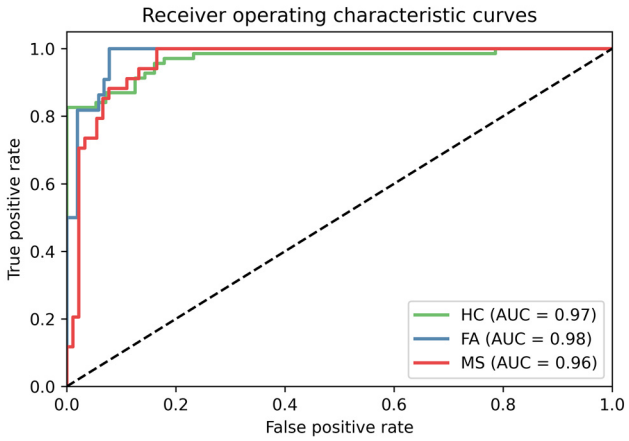

Fig. 2. Receiver operating curves for healthy controls (HC), Friedreich ataxia (FA), and multiple sclerosis (MS) obtained from the model with average performance.

(SHAP) values were examined. These show the probability of each outcome based on the information provided by each feature [53], [54]. To achieve a more parsimonious model, we performed the same machine learning procedures twice including features that produced SHAP values above criteria of 2% ($n = 87$) and 5% ($n = 21$) for at least one group (see Supplementary Materials for rankings). Overall model accuracy (Matthew's Correlation Coefficient) significantly increased relative to the full model ($M = 82.3\%$, $SEM = 0.4\%$) for the 2% cut-off ($p = 0.001$; $M = 83.7\%$, $SEM = 0.4\%$) and 5% cur-off ($p < 0.001$; $M = 83.6\%$, $SEM = 0.4\%$). Pairwise comparisons of accuracy between models for each group revealed significant increases in $F1$ accuracy between the full model and the 2% cut-off for all groups ($ps < .002$), and between the full model and 5% cut-off for the HC and MS group ($ps < 0.03$) but not the FA group ($p = 0.11$) (see Figure 1). These results suggest that high discrimination accuracy can be achieved with a reduced subset of 21 acoustic features. It should be noted, however, that larger subsets of variables may be required to achieve high discrimination accuracy if a broader scope of clinical groups are included.

B. Optimal Clinical Acoustic Markers

We describe the top 21 features overall, and top 10 for each group and overall (see Supplementary Materials for all features). As shown in Table II, the dominant acoustic features for accurate classification were spectral decrease, peak $f0$ energy, peak energy in the low, high, and broadband frequency ranges, low-frequency summed energy, utterance duration based on summed broadband energy (including low-, mid- and high-frequency sub-bands), spectral spread, and acoustic intensity. Figure 3 shows that healthy controls were characterized by a less steep and less variable spectral decrease, a smaller spectral spread and range of energy produced in low frequencies, greater energy in low and $f0$ frequency bands, and shorter utterance durations. The FA group was characterized by low intensity and energy in low, high, and broadband frequency bands, a higher and more variable spectral spread, and longer utterance durations. The MS group was characterized by a steeper and more variable spectral decrease, as well as

TABLE II
TOP 10 ACOUSTIC FEATURES FOR CATEGORIZING GROUPS BASED ON SHAP VALUES

Rank	Overall	Healthy Control (HC)	Friedreich Ataxia (FA)	Multiple sclerosis (MS)
1	Min. Spectral decrease	Min. Spectral decrease	Min. Spectral decrease	Min. Spectral decrease
2	Mean Spectral decrease	Mean Spectral decrease	Min. *f0* peak energy	Mean Spectral decrease
3	Range Spectral decrease	Range Spectral decrease	Min. High-frequency peak energy	Range Spectral decrease
4	Range Low-frequency summed energy	Range Low-frequency summed energy	Min. Utterance duration (low-freq. summed energy)	Range Low-frequency summed Energy
5	Min. High-frequency peak energy	Max. Spectral spread	Max. Spectral decrease	Min. High-frequency peak energy
6	SD Spectral decrease	SD Spectral decrease	Mean Utterance duration (broadband summed energy)	SD Spectral decrease
7	Max. Spectral spread	Min. Low-frequency peak energy	Mean Utterance duration (mid-freq. summed energy)	Min. Acoustic intensity (dB SPL)
8	Min. Low-frequency peak energy	Min. Broadband peak energy	Min. Low-frequency peak energy	Max. Spectral spread
9	Min. Acoustic intensity (dB SPL)	Range Spectral spread	Mean High-frequency summed energy	Min. *f0* peak energy
10	Min. *f0* peak energy	CoV Spectral decrease	Mean Utterance duration (high-freq. summed energy)	Min. Utterance duration (low-freq. summed energy)

SD = Standard deviation, dB SPL = decibel sound pressure level, CoV = Coefficient of Variation

utterance durations and spectral spread values that fell between the control and FA groups (see link in Figure 3 note for figures of all acoustic features). Other acoustic features that were useful in delineating groups include metrics of speech timing (pause duration, speech rate, and stress rate [22]), spectral features (crest, slope, centroid, flatness, and entropy [55]), formants 1-5, and the alpha ratio [56].

IV. DISCUSSION

Our machine learning approach distinguished between healthy controls, people with Multiple Sclerosis, and people with Friedreich Ataxia with high accuracy using acoustic properties of speech alone. These results indicate that multiclass supervised machine learning has the potential to discriminate between diseases, a step beyond mere healthy-pathological dichotomies. Through the accumulation of big data that merges speech data from various patient populations, we may be able to use machine learning to assist in the detection of specific diseases using acoustic markers.

There are numerous advantages for using acoustic markers to detect neurodegenerative disease including the decreased risk and burden of travelling to a hospital to undergo a range of tests, some of which are invasive. Speech, on the other hand, can be recorded within a familiar and comfortable setting, using common household devices (e.g., smartphones). Smartphones have demonstrated relative robustness for obtaining acoustic clinical markers and, therefore, increase accessibility to these automated detection methods [57]. Although the present study recorded speech within laboratory settings, it is also possible to record speech data remotely [58]. Practitioners could use this information to refine test selection for differential diagnosis. This would be particularly useful for people living in rural communities with increased travel burdens or during situations where the risk of infection is heightened (e.g., pandemics). Speech markers can be used as a remote tool to initially detect signs of neurodegenerative disease, expand our understanding of the clinical characteristics of these diseases to improve our ability to develop targeted interventions, and to monitor disease progression or treatment response.

We identified several acoustic features that strongly contributed to distinguishing between groups. Spectral decrease, the average of all slopes between the peak amplitude at the fundamental frequency and the peak amplitude of the formants (i.e., harmonics), was the most useful variable in distinguishing our three groups. This finding is in line with previous results that suggest vocal fold dysfunction is associated with greater energy in the lower frequency range relative to higher frequencies (e.g., the soft phonation index [59]). Other spectral features associated with the distribution of vocal energy also contributed to classification accuracy, including summed and peak energy within low-frequency bands (1-75 Hz), peak energy within *f0* (75-500Hz), high (4000-8000 Hz), and broadband (1-8000Hz) frequency ranges, and the spectral spread of peak frequencies. Therefore, the distribution of acoustic energy across the spectrum that reflects voice qualia culminates as a strong set of acoustic clinical markers for distinguishing neurodegenerative diseases.

Speech timing measures were also strong contributors to classification accuracy, specifically, the duration of syllables based on summed energy in low and broadband frequency ranges, and the rate of stressed syllable onsets based on peak energy across broadband frequencies [22]. These results corroborate previous findings that demonstrate slowed speech rate and decreased phonation time for a range of neurodegenerative diseases including Parkinson's disease [60], Huntington's disease [8], [61], multiple sclerosis [4], [5], and other diseases [3], [16], [61] and ataxias [62], [63], [64]. Speech rate and phonation time reflect both pneumo-articulatory capacity and oral-motor function, and could serve as clinical acoustic markers for monitoring the progression of neurodegenerative

Fig. 3. Normalized (z-scores) values of the top 21 acoustic features for identifying members of the healthy control (HC), Friedreich ataxia (FA), and multiple sclerosis (MS), groups. Note. See here for interactive figures for all acoustic features.

disease, distinguishing between diseases, and determining disease severity.

To the knowledge of the authors, this is the first paper to use machine learning to simultaneously differentiate three groups of disease classes (i.e., healthy controls, and people Multiple Sclerosis or Friedreich Ataxia) using speech data. This novel application of machine learning and acoustic analysis paves the way for new pre-diagnostic methods that could leverage big data to discriminate between a range of neurodegenerative diseases and/or other conditions. Through initiatives that obtain and share speech data from various clinical populations, our innovative approach could be applied to any population that is able to produce speech. The implications of this approach are substantial and provide new opportunities for healthcare, particularly for remote and rural areas where access to health providers might be limited.

A. Limitations and Considerations

We used the most common acoustic features (or similar proxies) based on an *a priori* analysis of the neurodegenerative disease literature and timbral features used in music information retrieval. There are, however, other acoustic features that may increase the accuracy and sensitivity of the machine learning algorithm that were not considered here. For example, voice onset time, the time between the burst of a stop consonant and the onset of the vowel, is an acoustic feature that differs significantly between controls and people with Parkinson's disease [65]. We opted not to use this measure because our data contained a high degree of coarticulation, and there is little agreement for the best way to extract the burst and vowel onset times and which acoustic features should be considered (see [66]). Similarly, we did not include measures from other voice assessment tasks (e.g., sustained vowel) [67] that can more reliably measure certain features (e.g., jitter and shimmer) but preclude the measurement of speech timing. We chose to constrain the number of variables and tasks to avoid overfitting. Future studies could use feature selection and pruning methods (e.g., [68]) to find the best feature set and remove unreliable variables prior to analysis.

The inclusion of non-speech performance measurements could also increase discrimination accuracy, for instance, cognitive [69] and motor performance [70] measures. The primary aim of this experiment was to examine accuracy using speech features alone because speech data can easily be obtained in the absence of a clinician through websites and smartphone applications [71]. Other cognitive and motor tests often require scoring by a clinician or dedicated tools to measure gait and tremor, although some remote tests are

available [72]. We show that neurodegenerative diseases can be delineated with high accuracy from speech data alone, but future applications could also consider other non-verbal features, for example, irregular gait patterns using smartphone accelerometers or irregular typing patterns. Whether these movement features or others would increase the accuracy of machine learning algorithms for neurodegenerative disease remains unknown.

B. Future Directions

The current study differentiated neurodegenerative diseases with high accuracy, but the approach did not aim to determine the severity or stage of the disease [35], [73], [74]. Future studies could employ an approach in which the severity of the disease is predicted or estimated following identification. A two-phased approach might be necessary because measures of disease severity tend to be idiosyncratic to the specific disease. Therefore, it remains a challenge to provide a measure of severity that can be applied to a range of diseases and conditions while capturing the relevant clinical markers.

V. Conclusion

We provide strong evidence that neurodegenerative diseases can be differentiated through acoustic clinical markers and machine learning, even when the speech phenotype is subtle or similar across groups. This model can be expanded and improved through the inclusion of additional diseases and phenotypes. Big data initiatives that bring together researchers and speech data from multiple laboratories are necessary to increase the scope of diseases that can be identified by acoustic clinical markers and machine learning. Moreover, a combination of remote testing tools for physical and cognitive assessment could be included in addition to speech to improve identification accuracy. These technologies promise to provide tools that can aid practitioners in reaching a diagnosis and relieve the physical and financial burden of patients.

Conflict of Interest

APV is the CSO of Redenlab, a speech clinical marker company.

References

[1] A. V. Feijó, M. A. Parente, M. Behlau, S. Haussen, M. C. De Veccino, and B. C. de Faria Martignago, "Acoustic analysis of voice in multiple sclerosis patients," *J. Voice*, vol. 18, no. 3, pp. 341–347, Sep. 2004, doi: 10.1016/j.jvoice.2003.05.004.

[2] G. Noffs et al., "Acoustic speech analytics are predictive of cerebellar dysfunction in multiple sclerosis," *Cerebellum*, vol. 19, pp. 691–700, Jun. 2020.

[3] G. Noffs et al., "Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis," *Eur. J. Neurol.*, vol. 28, no. 1, pp. 259–268, Jan. 2021.

[4] K. Tjaden, J. Lam, and G. Wilding, "Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions," *J. Speech, Lang., Hearing Res.*, vol. 56, no. 5, pp. 1485–1502, Oct. 2013, doi: 10.1044/1092-4388(2013/12-0259).

[5] K. Tjaden and V. Martel-Sauvageau, "Consonant acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear and loud speaking conditions," *Amer. J. Speech-Language Pathol.*, vol. 26, no. 2S, pp. 569–582, Jun. 2017, doi: 10.1044/2017_AJSLP-16-0090.

[6] J. Rusz et al., "Objective acoustic quantification of phonatory dysfunction in Huntington's disease," *PLoS ONE*, vol. 8, no. 6, Jun. 2013, Art. no. e65881, doi: 10.1371/journal.pone.0065881.

[7] L. R. Kaploun, J. H. Saxman, P. Wasserman, and K. Marder, "Acoustic analysis of voice and speech characteristics in presymptomatic gene carriers of Huntington's disease: Biomarkers for preclinical sign onset?" *J. Med. Speech-Lang. Pathol.*, vol. 19, pp. 49–65, Jan. 2011.

[8] S. Skodda, U. Schlegel, R. Hoffmann, and C. Saft, "Impaired motor speech performance in Huntington's disease," *J. Neural Transmiss.*, vol. 121, no. 4, pp. 399–407, Apr. 2014, doi: 10.1007/s00702-013-1115-9.

[9] K. L. Lansford and J. M. Liss, "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification," *J. Speech, Lang., Hearing Res.*, vol. 57, no. 1, pp. 57–67, Feb. 2014, doi: 10.1044/1092-4388(2013/12-0262).

[10] M. Magee, D. Copland, and A. P. Vogel, "Motor speech and non-motor language endophenotypes of Parkinson's disease," *Expert Rev. Neurotherapeutics*, vol. 19, no. 12, pp. 1191–1200, Dec. 2019.

[11] A. P. Vogel et al., "Voice in friedreich ataxia," *J. Voice*, vol. 31, no. 2, pp. 243.e9–243.e19, Mar. 2017, doi: 10.1016/j.jvoice.2016.04.015.

[12] K. M. Rosen et al., "Spectral measures of the effects of Friedreich's ataxia on speech," *Int. J. Speech-Language Pathol.*, vol. 13, no. 4, pp. 329–334, Aug. 2011, doi: 10.3109/17549507.2011.529940.

[13] A. M. Pascu, P. Ifteni, A. Teodorescu, V. Burtea, and C. U. Correll, "Delayed identification and diagnosis of Huntington's disease due to psychiatric symptoms," *Int. J. Mental Health Syst.*, vol. 9, no. 1, p. 33, Dec. 2015, doi: 10.1186/s13033-015-0026-6.

[14] J. Warner, L. Barron, D. St Clair, and D. Brock, "Reliability of clinical diagnosis of Huntington's disease," *J. Neurol., Neurosurg. Psychiatry*, vol. 57, no. 10, p. 1277, Oct. 1994, doi: 10.1136/jnnp.57.10.1277.

[15] B. G. Schultz and A. P. Vogel, "A tutorial review on clinical acoustic markers in speech science," *J. Speech, Lang., Hearing Res.*, vol. 65, no. 9, pp. 3239–3263, Sep. 2022, doi: 10.1044/2022_JSLHR-21-00647.

[16] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A survey on machine learning approaches for automatic detection of voice disorders," *J. Voice*, vol. 33, no. 6, pp. 947.e11–947.e33, Nov. 2019, doi: 10.1016/j.jvoice.2018.07.014.

[17] T. D. Prins, "Motor and auditory abilities in different groups of children with articulatory deviations," *J. Speech Hearing Res.*, vol. 5, no. 2, pp. 161–168, Jun. 1962.

[18] B. M. Ben-David and M. Icht, "The effect of practice and visual feedback on oral-diadochokinetic rates for younger and older adults," *Lang. Speech*, vol. 61, no. 1, pp. 113–134, Jun. 2017, doi: 10.1177/0023830917708808.

[19] T. W. Powell, "A comparison of english reading passages for elicitation of speech samples from clinical populations," *Clin. Linguistics Phonetics*, vol. 20, nos. 2–3, pp. 91–97, Jan. 2006.

[20] T. Sorensen, A. Toutios, L. Goldstein, and S. S. Narayanan, "Characterizing vocal tract dynamics across speakers using real-time MRI," in *Proc. Interspeech*, Sep. 2016, pp. 465–469.

[21] J. S. Bhat, "Oral diadokokinetic rate—An insight into speech motor control," *Int. J. Adv. Speech Hearing Res.*, vol. 1, no. 1, pp. 10–14, 2012.

[22] B. G. Schultz, I. O'Brien, N. Phillips, D. H. McFarland, D. Titone, and C. Palmer, "Speech rates converge in scripted turn-taking conversations," *Appl. Psycholinguistics*, vol. 37, no. 5, pp. 1201–1220, Sep. 2016, doi: 10.1017/S0142716415000545.

[23] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[24] G. Schlotthauer, M. E. Torres, and C. Jackson-Menaldi, "Automatic diagnosis of pathological voices," *WSEAS Trans. Signal Proc.*, vol. 2, pp. 1260–1267, Sep. 2006.

[25] B. Ghoraani and S. Krishnan, "A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–11, Dec. 2009.

[26] S. Jothilakshmi, "Automatic system to detect the type of voice pathology," *Appl. Soft Comput.*, vol. 21, pp. 244–249, Aug. 2014.

[27] M. A. Myszczynska et al., "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Rev. Neurol.*, vol. 16, no. 8, pp. 440–456, Aug. 2020.

[28] B. G. Schultz, Z. Joukhadar, U. Nattala, M. del Mar Quiroga, F. Bolk, and A. P. Vogel, "Best practices for supervised machine learning when examining biomarkers in clinical populations," in *Big Data in Psychiatry #x0026; Neurology*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 1–34.

[29] E. C. Neto et al., "Detecting the impact of subject characteristics on machine learning-based diagnostic applications," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–6, Oct. 2019, doi: 10.1038/s41746-019-0178-x.

[30] *Audacity*, Audacity Development Team, Oak Park, MI, USA, 2018.

[31] *MATLAB*, MathWorks, Natick, MA, USA, 2019.

[32] B. G. Schultz, "The Schultz MIDI benchmarking toolbox for MIDI interfaces, percussion pads, and sound cards," *Behav. Res. Methods*, vol. 51, no. 1, pp. 204–234, Feb. 2019.

[33] O. Lartillot, P. Toiviainen, and T. Eerola, "A MATLAB toolbox for music information retrieval," in *Data Analysis, Machine Learning and Applications*. Cham, Switzerland: Springer, 2008, pp. 261–268.

[34] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu. (2009). *Voicesauce*. Program. [Online]. Available: http://www.seas.ucla.edu/spapl/voicesauce/. UCLA

[35] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. Roy. Soc. Interface*, vol. 8, no. 59, pp. 842–855, Jun. 2011.

[36] J. Kreiman, B. R. Gerratt, G. B. Kempster, A. Erman, and G. S. Berke, "Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research," *J. Speech, Lang., Hearing Res.*, vol. 36, no. 1, pp. 21–40, Feb. 1993.

[37] J. H. L. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 74–99, Nov. 2015.

[38] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the Institute of Phonetic Sciences*. Princeton, NJ, USA: Citeseer, 1993, pp. 97–110.

[39] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *J. Acoust. Soc. Amer.*, vol. 90, no. 5, pp. 2394–2410, Nov. 1991.

[40] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," *CUIDADO Ist Project Rep.*, vol. 54, pp. 1–25, Apr. 2004.

[41] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *J. Acoust. Soc. Amer.*, vol. 130, no. 5, pp. 2902–2916, Nov. 2011.

[42] J. Lefebvre, "Physical basis of acoustics," in *Acoustics*, P. Filippi, D. Habault, J. Lefebvre, and A. Bergassoli, Eds. London, U.K.: Academic, 1999, pp. 1–39.

[43] A. P. Vogel, J. Fletcher, P. J. Snyder, A. Fredrickson, and P. Maruff, "Reliability, stability, and sensitivity to change and impairment in acoustic measures of timing and frequency," *J. Voice*, vol. 25, no. 2, pp. 137–149, Mar. 2011.

[44] G. Prendergast, S. R. Johnson, and G. G. R. Green, "Extracting amplitude modulations from speech in the time domain," *Speech Commun.*, vol. 53, no. 6, pp. 903–913, Jul. 2011.

[45] C. J. Chang, *Time Frequency Analysis and Wavelet Transform Tutorial*. Taipei, Taiwan: National Taiwan Univ., 2010.

[46] B. G. Schultz and S. A. Kotz, "Finding the beat in German poetry: The role of meter, rhyme, and lexical content," in *Proc. 14th Int. Conf. Music Perception Cognition*, 2016, p. 374.

[47] M. S. Ricketts, B. G. Schultz, and D. G. Watson, "Speech rate convergence in spontaneous conversation," *Appl. Psycholinguistics*, to be published.

[48] S. Dixon, "Onset detection revisited," in *Proc. 9th Int. Conf. Digit. Audio Effects*, 2006, pp. 133–137.

[49] A. Tierney, A. D. Patel, and M. Breen, "Acoustic foundations of the speech-to-song illusion," *J. Experim. Psychol., Gen.*, vol. 147, no. 6, pp. 888–904, Jun. 2018.

[50] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: An interdisciplinary review," *J. Big Data*, vol. 7, no. 1, pp. 1–45, Dec. 2020.

[51] *R: A Language and Environment for Statistical Computing*, RR Core Team, Vienna, Austria, 2013.

[52] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thoracic Oncol.*, vol. 5, no. 9, pp. 1315–1316, Sep. 2010.

[53] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using improved Shapley additive explanation," in *Proc. 10th ACM Int. Conf. Bioinf., Comput. Biol. Health Informat.*, Sep. 2019, p. 546.

[54] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[55] A. C. C. Gama, Z. Camargo, M. A. R. Santos, and L. C. Rusilo, "Discriminant capacity of acoustic, perceptual, and vocal self: The effects of vocal demands," *J. Voice*, vol. 29, no. 2, pp. 260.e45–260.e50, Mar. 2015, doi: 10.1016/j.jvoice.2014.06.012.

[56] A. P. Vogel, J. Fletcher, and P. Maruff, "Acoustic analysis of the effects of sustained wakefulness on speech," *J. Acoust. Soc. Amer.*, vol. 128, no. 6, pp. 3747–3756, Dec. 2010, doi: 10.1121/1.3506349.

[57] S. Jannetts, F. Schaeffler, J. Beck, and S. Cowen, "Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types," *Int. J. Lang. Commun. Disorders*, vol. 54, no. 2, pp. 292–305, Mar. 2019.

[58] A. Tsanas, M. A. Little, and L. O. Ramig, "Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network," *IEEE Access*, vol. 9, pp. 11024–11036, 2021.

[59] T. Bhuta, L. Patrick, and J. D. Garnett, "Perceptual evaluation of voice quality and its correlation with acoustic measurements," *J. Voice*, vol. 18, no. 3, pp. 299–304, Sep. 2004, doi: 10.1016/j.jvoice.2003.12.004.

[60] S. Skodda and U. Schlegel, "Speech rate and rhythm in Parkinson's disease," *Movement Disorders*, vol. 23, no. 7, pp. 985–992, May 2008.

[61] A. P. Vogel, C. Shirbin, A. J. Churchyard, and J. C. Stout, "Speech acoustic markers of early stage and prodromal Huntington's disease: A marker of disease onset?" *Neuropsychologia*, vol. 50, no. 14, pp. 3273–3278, Dec. 2012.

[62] A. P. Vogel et al., "Features of speech and swallowing dysfunction in pre-ataxic spinocerebellar ataxia type 2," *Neurology*, vol. 95, no. 2, pp. e194–e205, Jul. 2020.

[63] A. P. Vogel et al., "Coordination and timing deficits in speech and swallowing in autosomal recessive spastic ataxia of Charlevoix–Saguenay (ARSACS)," *J. Neurol.*, vol. 265, no. 9, pp. 2060–2070, 2018.

[64] A. P. Vogel et al., "Speech and swallowing abnormalities in adults with POLG associated ataxia (POLG–A)," *Mitochondrion*, vol. 37, pp. 1–7, Nov. 2017.

[65] D. Montaña, Y. Campos-Roca, and C. J. Pérez, "A diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 154, pp. 89–97, Feb. 2018, doi: 10.1016/j.cmpb.2017.11.010.

[66] R. P. A. Ozsancak, "Voice onset time in aphasia, apraxia of speech and dysarthria: A review," *Clin. Linguistics Phonetics*, vol. 14, no. 2, pp. 131–150, Jan. 2000, doi: 10.1080/026992000298878.

[67] L. Lee, J. C. Stemple, L. Glaze, and L. N. Kelchner, "Quick screen for voice and supplementary documents for identifying pediatric voice disorders," *Lang., Speech, Hearing Services Schools*, vol. 35, no. 4, pp. 308–319, Oct. 2004.

[68] Z. Zhang, "Variable selection with stepwise and best subset approaches," *Ann. Transl. Med.*, vol. 4, no. 7, pp. 1–6, 2016, doi: 10.21037/atm.2016.03.35.

[69] M. Jalakas et al., "A quick test of cognitive speed can predict development of dementia in Parkinson's disease," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Oct. 2019.

[70] M. Montero-Odasso et al., "Motor phenotype in neurodegenerative disorders: Gait and balance platform study design protocol for the Ontario Neurodegenerative Research Initiative (ONDRI)," *J. Alzheimer's Disease*, vol. 59, no. 2, pp. 707–721, Jul. 2017.

[71] S. Arora, C. Lo, M. Hu, and A. Tsanas, "Smartphone speech testing for symptom assessment in rapid eye movement sleep behavior disorder and Parkinson's disease," *IEEE Access*, vol. 9, pp. 44813–44824, 2021.

[72] B. McLaren et al., "Feasibility and initial validation of 'HD-mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease," *J. Neurol.*, vol. 268, no. 2, pp. 590–601, Feb. 2021.

[73] M. Asgari and I. Shafran, "Predicting severity of Parkinson's disease from speech," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Sep. 2010, pp. 5201–5204, doi: 10.1109/IEMBS.2010.5626104.

[74] D. Hemmerling and M. Wojcik-Pedziwiatr, "Prediction and estimation of Parkinson's disease severity based on voice signal," *J. Voice*, vol. 36, no. 3, pp. 439.e9–439.e20, May 2022.