digital biomarkers

Digit Biomark 2019;3:72–82
DOI: 10.1159/000500354
Received: February 14, 2019
Accepted: April 12, 2019
Published online: June 12, 2019

Research Reports – Brief Report

Depression Screening from Voice Samples of Patients Affected by Parkinson's Disease

Yasin Ozkanca[a] Miraç Göksu Öztürk[b] Merve Nur Ekmekci[a]
David C. Atkins[c] Cenk Demiroglu[a] Reza Hosseini Ghomi[d]

[a]Electrical and Electronics Engineering, Ozyegin University, Istanbul, Turkey; [b]Computer Engineering, Bogazici University, Istanbul, Turkey; [c]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA; [d]Department of Neurology, University of Washington, Seattle, WA, USA

Keywords

Depression screening · Parkinson's disease · Feature selection · Deep neural networks · Voice technology · Audio features · Voice biomarkers

Abstract

Depression is a common mental health problem leading to significant disability worldwide. It is not only common but also commonly co-occurs with other mental and neurological illnesses. Parkinson's disease (PD) gives rise to symptoms directly impairing a person's ability to function. Early diagnosis and detection of depression can aid in treatment, but diagnosis typically requires an interview with a health provider or a structured diagnostic questionnaire. Thus, unobtrusive measures to monitor depression symptoms in daily life could have great utility in screening depression for clinical treatment. Vocal biomarkers of depression are a potentially effective method of assessing depression symptoms in daily life, which is the focus of the current research. We have a database of 921 unique PD patients and their self-assessment of whether they felt depressed or not. Voice recordings from these patients were used to extract paralinguistic features, which served as inputs to machine learning and deep learning techniques to predict depression. The results are presented here, and the limitations are discussed given the nature of the recordings which lack language content. Our models achieved accuracies as high as 0.77 in classifying depressed and nondepressed subjects accurately using their voice features and PD severity. We found depression and severity of PD had a correlation coefficient of 0.3936, providing a valuable feature when predicting depression from voice. Our results indicate a clear correlation between feeling depressed and PD severity. Voice may be an effective digital biomarker to screen for depression among PD patients.

Reza Hosseini Ghomi
Department of Neurology, University of Washington
1959 NE Pacific Street
Seattle, WA 98195 (USA)
E-Mail rezahg@uw.edu

Digit Biomark 2019;3:72–82
DOI: 10.1159/000500354

Introduction

Depression alone accounts for 10% of all disability due to physical and mental health problems globally. Moreover, it is the primary reason for suicide, and it is estimated to be responsible for 1.4% of all deaths around the world [1]. It is also predicted to be the leading cause of disease burden by 2030 [2]. However, better diagnosis of depression followed by successful treatment was shown to be effective in mitigating the symptoms and decreasing suicide rates [3]. One particular case of interest to us is the detection of depression in patients diagnosed with Parkinson's disease (PD), a disease with depression as common comorbidity [4–7].

According to Nazem et al. [8], up to one-third of PD patients exhibit symptoms of depression and experience suicidal ideation. Another systematic review found clinically significant depressive symptoms in 35% of patients with a PD diagnosis [9]. Another interesting finding is the frequency with which affective symptoms precede motor symptoms in PD – up to several years prior to PD [10]. It has been shown that treatment of depression in PD leads to significant improvement in quality of life and disability measures, which are evident after only 8 weeks of treatment and sustained at the 24-week follow-up [11]. Further, indicating the important role of depression in PD is the finding that depressive symptoms occur more often than motor symptoms, which are a strong predictor of initiation of dopaminergic medications. Similar studies have shown that initiation of dopaminergic medication is delayed after treatment of depression along with earlier initiation of dopaminergic treatment when depression symptoms are present [12].

Voice signals have previously been shown to carry significant information regarding the mental health of the speaker [13–15]. According to Vlasenko et al. [16], performance of depression detection was improved when the feature extraction process was conducted differently depending on gender. In a study by Helfer et al. [17], distortions in formant trajectories of speech were shown to be a reliable indication of depression. Additionally, Cummins et al. [18] claimed that occurrence of degradation in spectral variability can also be an implication of depression. Moreover, there are several studies [19–21] reporting that retardations in motor control due to depression may cause distortions in coordination and timing of speech production.

In this study, we present the results of an analysis including 921 unique PD patients who provided both samples of their voice and subjective depression symptoms, which demonstrated the accuracy of depression state prediction from voice data.

Materials and Methods

Data Collection

The mPower Voice Dataset was collected by Sage Bionetworks [22]. Participants recorded their voice for 10 s making the single phoneme sound, /ˈɑ/, pronounced "ahhh." These recordings were obtained using a mobile application on the iPhone. There are over 64,000 recordings but not all match with a completed PD Questionnaire (PDQ-8) and Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS). Narrowing based on these criteria, we have 921 unique subjects who provided an answer to the third question of the PDQ-8 survey, which is: "Due to Parkinson's disease, how often during the last month have you felt depressed?" Some subjects answered the survey multiple times providing a total of 921 subjects and 45,869 voice recordings.

There was an overall bias in age and sex present with almost a 4:1 ratio of male:female participants (5,314 male and 1,461 female) in total, which is expected given the nature of PD.

Of the participants who had a professional PD diagnosis, only 715 were male and 369 female, reflecting a large number with missing or no diagnosis. The average age of all participants was 36.6 years, but when broken down by diagnosis presence, those with a PD diagnosis participating had an average age of 60 years, while those without a PD diagnosis had an average age of 32 years, again reflecting a bias in age.

To reduce identity confounding during the modeling process, we have only used one recording per subject. These recordings were chosen at random from each subject. We did not choose to use leave-one-subject-out cross validation in order to incorporate more of the data due to the concerns of within subject variation in data, which is well known in similar data sets [23]. Particularly with voice features, there tends to be large variation in features within subjects which violates the primary assumption of within subject consistency at the core of leave-one-subject-out cross validation. In our case, the likelihood of a mismatch between the train and test set distributions was high and so was the likelihood of underfitting the model with this method. Having a relatively small subject pool compared to the population and variance we see clinically in PD, the chances of removing a critical pattern by removing an entire subject's observations are high here. For these reasons, we did not use different combinations of recordings from each patient in our analysis. This is an important topic when addressing data analysis for human subjects, which is excellently described by Little et al. [23].

Possible answers to PDQ-8 question 3 were: "never," "occasionally," "sometimes," "often," and "always." We selected subjects who answered "never" as control group, while the remainder of the data was categorized as depressed, providing 603 nondepressed and 318 unique depressed recordings.

Parts 1 and 2 of the MDS-UPDRS survey have 13 questions each, but the mPower study altered the survey to use only 6 questions from part 1 and 10 questions from part 2 in order to increase likelihood of completion. Notably, changes in UPDRS create a potential change in the validity of the tool, but we decided to continue analysis with the data available. Part 1 pertains to non-motor experiences of daily living (e.g., emotional status) while part 2 pertains to motor experiences of daily living (e.g., difficulty in getting dressed). The questions have possible answers 0, 1, 2, 3, or 4 (with a total score range of 0–64), which indicate severity levels. We calculated the total scores provided by the subjects to identify their PD severity. To correlate both PD severity and depression levels, we needed subjects who answered both surveys, which is how we arrived at 921 unique subjects.

Feature Extraction and Preprocessing

We first removed silences from the recordings using a voice activation detection algorithm from the MATLAB Voicebox toolkit [24]. After cleaning, features were extracted. For feature extraction, openSMILE [25], an open-source application, was used. We extracted 2 sets of features. The first was derived from AVEC 2013 [26], consisting of 2,268 features – 32 being energy and spectral related, such as loudness, zero crossing rate, harmonicity, and skewness, and 6 being vocal acoustic related, such as jitter, shimmer, and F0, including their functionals and Mel-frequency cepstral coefficients (MFCCs) 1–16. Other descriptive features included voiced portions of the recordings. The second set of features was derived from GeMAPS [27], consisting of 62 features, chosen by an expert panel to be most important in voice analysis. These 2 feature sets do have overlap and were thus kept separate for analysis.

After feature extraction, we implemented a feature selection algorithm using the AVEC 2013 feature set, namely minimum redundancy maximum relevance (MRMR) [28]. The MRMR algorithm filtered the most relevant and least overlapping features to provide a reduced set of features providing the highest correlation to the desired variable.

Machine Learning Methods

To examine depression prediction from voice features, we tested several machine learning classification methods, including: (1) support vector machine (SVM), (2) random forest, and (3) fully connected, feed-forward deep neural network (DNN) models for both the AVEC 2013 and GeMAPS feature sets.

Prior to the application of each algorithm, the data were split into training and test sets, with a 76/24 split. For the deep learning algorithm, to address the bias in this data set having more nondepressed than depressed samples, the depressed samples were duplicated and added to the original set until the number of nondepressed samples is equal to the number of depressed samples [29, 30]. This duplication occurred only in the training set; there were no duplicates in the test set. Notably, oversampling was only used for the deep learning architecture and not for other models.

The DNN architecture consists of 1 fully connected hidden layer, having 64 units with a rectified linear unit (ReLU)

$$ReLU(x) = max(0, x) \quad \text{Eq. 1}$$

activation and an output layer having one-dimensional output with a sigmoid activation

$$\text{Sigmoid}(x) = 1/(1 + e^{-x}) \quad \text{Eq. 2}$$

as shown in Figure 4. The weights of the model were trained using the Adam optimization technique [31]. In addition to the test set, 10% of the training set was allocated separately as a validation set. This validation set was never shown to the model to reduce the likelihood of overfitting. During the training, the model performance was observed separately on the validation set at regular intervals, and the best version was saved as the final model. The process of validation during training is essential to prevent the model from memorizing the training data and to obtain a better general performance. The same DNN architecture was applied to both AVEC and GeMAPS feature sets.

For SVM, C and γ values were determined by using grid search on development data. This required a third partition in addition to the test and training data. The development data portion allowed us to tune the SVM classifier in order to optimize the hyperparameters first, then use the training data to train the model before applying it to the test data to determine classification performance. Radial basis function kernel was used.

For the random forest model, the number of estimators were set to 500 for all applications.

Data Analysis

The experiments were conducted in 4 separate groups formed according to the data set chosen and the PD severity used in training. We initially used the PD severity values as a predictor of depression, setting a threshold value where those with higher severity scores than the threshold were classified as depressed and the rest as nondepressed. The best severity threshold value is selected based on the accuracy of the training set, which was 12, while the minimum and maximum values were 1 and 40, respectively.

We incorporated the PD severity information into the training of our detection models in 2 ways. First, the severity value of a person was concatenated to the audio features of that person's speech, and the models were trained with these new combined features. The results of this approach were not significantly different than the second approach and not therefore reported. In the second approach, we used either the PD severity or the voice feature model to make the depression prediction based on a threshold value. Let s_i be the PD severity of a patient with index i and let t_1 and t_2 be 2 threshold values. Whenever $s_i \leq t_1$ or $s_i \geq t_2$ for patient i, we made the prediction according to the PD severity, i.e., we used the PD severity-based predictor. Otherwise, we used the voice feature model. This idea was adopted after observing that PD severity was a good predictor of depression in and of itself as seen in Figure 3. The best t_1 and t_2 values varied depending on the ML model.

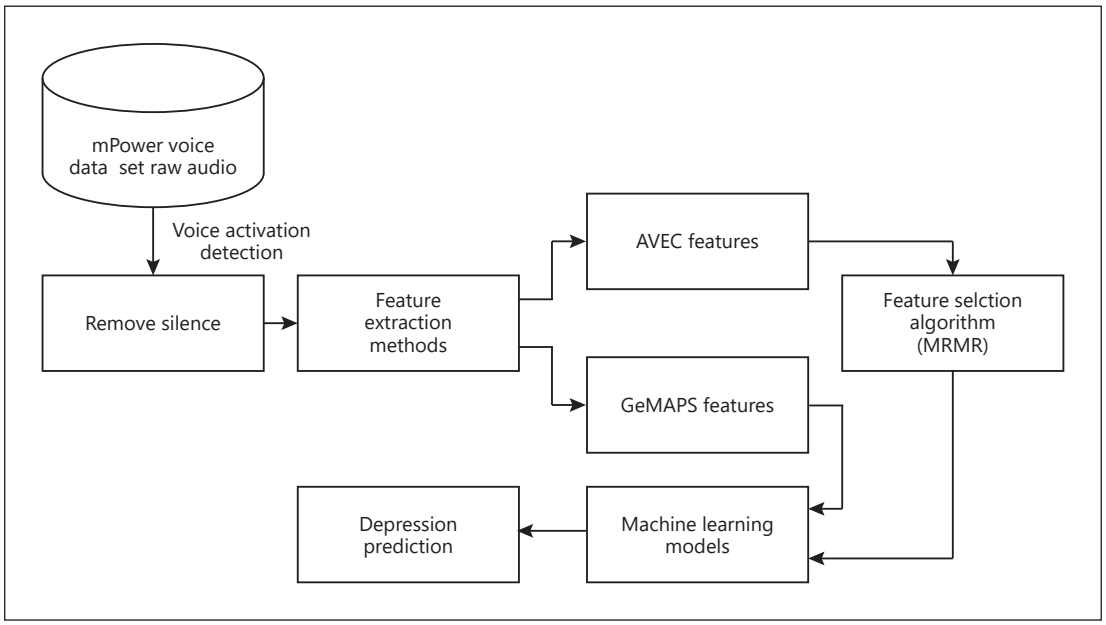

Fig. 1. A flow chart visualization of the system.

Table 1. Classification results found using the AVEC 2013 feature set

Method	Dimension	Precision	Recall	*F*-score	Accuracy
SVM	80	0.46 (0.67)	0.31 (0.80)	0.37 (0.73)	0.62
SVM + severity	40 + 1	0.66 (0.81)	0.65 (0.81)	**0.66** (0.81)	0.76
RF	800	0.67 (0.67)	0.13 (0.96)	0.22 (0.79)	0.66
RF + severity	400 + 1	0.72 (0.78)	0.56 (0.88)	0.63 **(0.83)**	**0.77**
DNN	30	0.56 (0.67)	0.19 (0.91)	0.29 (0.77)	0.66
DNN + severity	30 + 1	0.69 (0.80)	0.62 (0.84)	0.65 (0.82)	0.76
Severity	1	0.70 (0.78)	0.55 (0.87)	0.62 (0.82)	0.76

Experiments are conducted by taking only one audio sample from each patient. Highest scores are shown in bold as depressed (nondepressed). Voice feature model accuracy alone as well as with Parkinson's disease severity are shown for each model. RF, random forest; SVM, support vector machine; DNN, deep neural network.

Results

Figure 1 illustrates the sequence of steps taken to analyze our data in the final system. The accuracies of the depression classification models for each set of features are reported in Tables 1 and 2. Based on this analysis, we found a significant correlation between PD severity and depression presence and severity. Depression classification achieved F-scores as high as 0.66 using the AVEC 2013 feature set, and nondepressed classification achieved F-scores of 0.83 with total accuracy of 0.77. These were achieved using SVM + severity and random forest + severity models, respectively. Using the GeMAPS features, depressed F-scores were as high as 0.62, nondepressed 0.83, and total accuracy 0.76 with random forest + severity, DNN + severity, and severity alone models, respectively Table 3 reflects the distribution of patients in the analysis with how many were in the training and test sets based on their

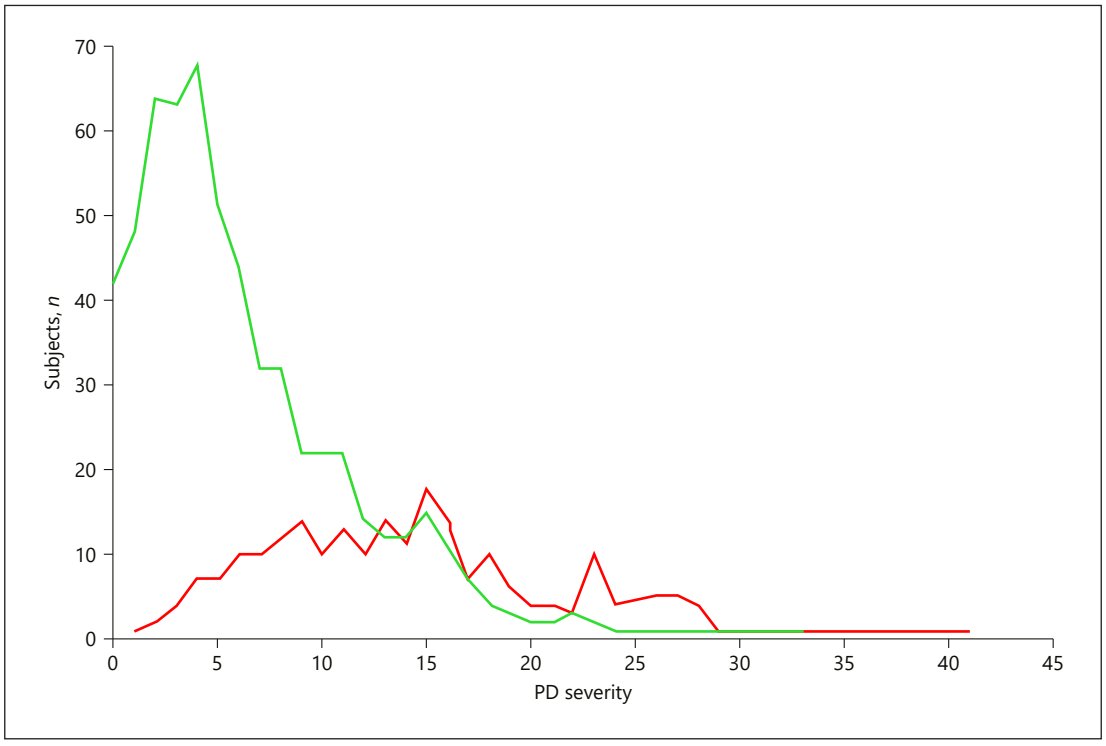

Fig. 2. Parkinson's disease (PD) severity is plotted versus number of subjects. Red line, depressed subjects; green line, nondepressed subjects. See data collection in the Materials and Methods section for PD severity source.

Table 2. Classification results found using the GeMAPS feature set

Method	Precision	Recall	*F*-score	Accuracy
SVM	0.45 (0.67)	0.37 (0.74)	0.41 (0.70)	0.60
SVM + severity	0.70 (0.77)	0.54 (0.86)	0.61 (0.81)	0.75
RF	0.45 (0.64)	0.06 (0.96)	0.11 (0.77)	0.62
RF + severity	0.79 (0.75)	0.47 (0.93)	0.59 **(0.83)**	**0.76**
DNN	0.45 (0.64)	0.11 (0.92)	0.18 (0.76)	0.62
DNN + severity	0.70 (0.78)	0.55 (0.87)	**0.62** (0.82)	**0.76**
Severity	0.70 (0.78)	0.55 (0.87)	**0.62** (0.82)	**0.76**

Experiments are conducted by taking only one audio sample from each patient. Highest scores are shown in bold as depressed (non-depressed). Voice feature model accuracy alone as well as with Parkinson's disease severity are shown for each model.

Table 3. Distribution of patients based on self-assessed depression frequency

Partition	Never	Occasionally	Sometimes	Often	Always	Total
Train	463	171	36	28	2	700
Test	140	56	18	7	0	221
Total	603	227	54	35	2	921

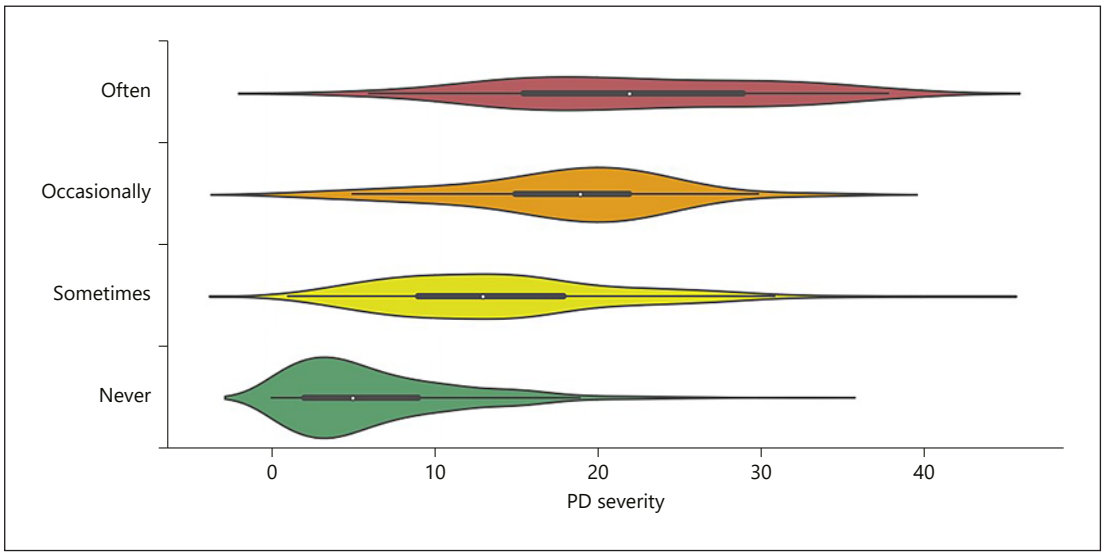

Fig. 3. Violin plot of Parkinson's disease (PD) severity level versus depression frequency. See data collection in the Materials and Methods section for PD and depression severity sources.

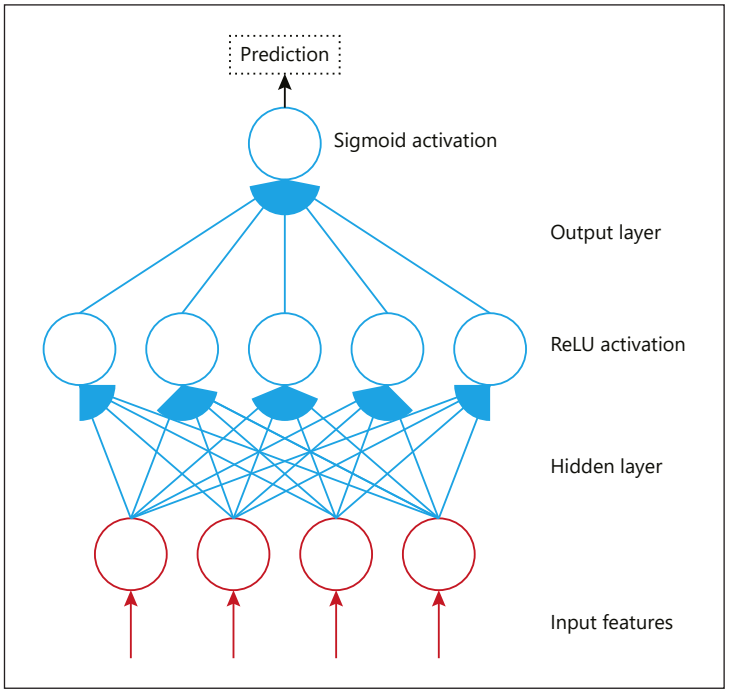

Fig. 4. Deep neural network architecture used in modeling.

depression severity. Notably, the performances reported are classification accuracies without cross validation (e.g., *k*-fold) and therefore do not represent a mean, only the single result of the model.

Although the applied techniques follow the same trend, performance-wise, in both AVEC and GeMAPS, the highest scores were obtained when the experiments were conducted on the AVEC feature set. This is expected as samples in the AVEC feature set have more features and, hence, are more likely to be separable. Regardless of the feature set used, Tables 1 and 2

Digit Biomark 2019;3:72–82
DOI: 10.1159/000500354

reflect the improvement in model accuracy when PD severity is used as a separate feature using a threshold value.

Figure 2 illustrates the number of patients with a certain PD severity score who are both depressed and nondepressed. The trend demonstrates as PD severity increases, the number of nondepressed individuals drops precipitously while the number of depressed individuals initially increases and then declines in the higher PD severity levels. The correlation coefficient between depression and PD severity is 0.3936. The mean severity level of subjects with depression is 15.43, whereas for the subjects without depression, the mean value is 6.25.

Figure 3 demonstrates the stepwise relationship where depression is broken down by frequency. Here, we see a gradual increase in depression as PD severity increases with a broadening of the population who reported feeling depressed "often" more evenly distributed across severity levels. Figure 4 illustrates the layers involved in the deep learning model architecture.

The top selected features by feature selection algorithm MRMR were mostly MFCCs. The top 10 selected features were MFFC 1, 5, and 11, and their functionals such as percentile 0.99, linear prediction gain, root quadratic mean, and absolute mean. These functionals are computed over 2-s windows. For more information regarding these features, please see the references for AVEC and GeMAPS.

Discussion

To our knowledge, this is the first analysis focused on the depression data within the PD data of the mPower study. Perhaps most notable in our findings is that we received the highest overall accuracy scores in both feature sets when the severity feature is included in the training. The best accuracy performance was obtained by the random forest model trained with the severity information on the AVEC feature set. This further validates previous findings of the close relationship between PD severity and depression. From our results, the complete separation of depression symptoms from PD symptoms is difficult and possibly impossible given that they may share some basic biological correlates.

Our accuracy here using PD severity along with voice features to detect depression in patients with PD is comparable to the accuracy of using the UPDRS and other rating scales as discussed in Goodarzi et al. [32]. However, using voice features presents an opportunity to automate depression detection and increase screening ability given the ease of obtaining a voice sample rather than completing a questionnaire. The purpose of our work is to explore and demonstrate feasibility of using digital biomarkers for screening purposes in order to improve treatment and outcomes. We chose to look at depression because of its subjective nature and current dependence on rating scales which do not correlate with a specific biological entity but rather the patient's experience. Using voice biomarkers allows capturing of the patient's symptoms and possible new therapeutic targets.

Given the significant correlation between PD severity and depression, it was expected using PD severity as a stand-alone feature would provide accurate results in classifying depression.

The correlation between the number of depressed and nondepressed subjects to the severity of PD is clinically expected when PD severity is low or the disease is early in its course, patients have mild symptoms, are likely not significantly affected functionally, and are less likely to be depressed. As the disease progresses, patients with higher severity experience a more significant decline in daily functioning with an expected increased likelihood for depression.

The trend to increasing depression with increasing PD severity is also expected. The specific trend seen in Figures 2 and 3 where the most often depressed subjects represent a broad range of PD severity may be because as either depression or PD severity reaches extremes, it is more difficult to distinguish symptoms. As PD severity increases, there is also an increasing risk for cognitive decline and dementia, which may cloud symptom reporting. These reasons along with the expected lower number of patients with higher PD and depression severities participating in studies due to poor functioning contribute to the difficulty in gathering accurate data remotely in severely ill individuals.

Limitations of this study include the need to use a reduced data set given the problem that identity confounding is not easily addressed without additional data or restricting the number of data points provided by any given subject. Unfortunately, in this case, we do not believe leave-one-out subject-wise cross validation does not address identity confounding [23], although it is commonly used for this purpose with health care-related data. We believe identity confounding is best addressed by larger data sets representing more diversity in data requiring significant improvement in how current clinical data collection is achieved. Given the content of the voice signal available here is limited based on the brief speech task which does not include linguistic features, we were limited to determining absence or presence of depression. In the long term, our goal is to develop a comprehensive voice biomarker for depression such that not only presence but also severity of symptoms can be determined. Here, we primarily demonstrate sufficiency of using brief audio features to detect the presence of depression.

Conclusion

This data set elucidates the common co-occurrence of depression among people who are affected by PD with a somewhat positive linear correlation between PD severity and depression severity. Although our recordings do not contain linguistic data, we have demonstrated that the relationship between voice and depression in PD is strongly correlated. We also demonstrated that feature optimization can yield more accurate results by reducing redundancy and noise, which is important for both computational complexity and accuracy. We demonstrated successful prediction of depression state using voice features and PD severity at relatively high rates given the limited voice content. This study is unique in terms of the size of the database, given that many of the previous depression studies use much smaller cohorts. We believe voice may be used as an accurate, accessible, and efficient marker of mood in PD, which helps to screen and treat depression. Our hope is to gather more data across larger populations with more voice and speech tasks to build a more comprehensive digital voice biomarker for PD and depression.

Acknowledgment

Dr. Hosseini Ghomi wishes to thank Dr. John Neumaier for his research support.

Data were contributed by users of the Parkinson mPower mobile application as part of the mPower study developed by Sage Bionetworks and described in Synapse [22].

Statement of Ethics

Ethical oversight of the study was provided by the Western Institutional Review Board (WIRB #20141369). Subjects who participated were required to complete an online consent process.

Disclosure Statement

Dr. Hosseini Ghomi is a stock holder of NeuroLex Laboratories.

Funding Sources

Dr. Hosseini Ghomi's work was supported by NIH R25 MH104159 during the completion of this work. This funding provided protected time in Dr. Hosseini Ghomi's schedule. Dr. Hosseini Ghomi's work in preparing this manuscript was supported during his VA Advanced Fellowship Program in Parkinson's Disease which provided protected research time.

Author Contributions

R.H.G. was the principal investigator, wrote and edited the manuscript, and guided the overall project. Y.O., M.G.O., and M.N.E. carried out the data cleaning, analysis, and significant portions of the manuscript writing. C.D. and D.C.A. were the primary mentors for this research and provided editorial guidance. All authors read and approved the final manuscript.

References

1 WHO. Causes of death in 2008. Geneva: WHO; 2008 [accessed 2019 Apr 9]. Available from: https://www.who.int/gho/mortality_burden_disease/causes_death_2008/en/.
2 WHO. The global burden of disease: 2004 update. Geneva: WHO; 2004 [accessed 2019 Apr 9]. Available from: https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/.
3 Rihmer Z. Can better recognition and treatment of depression reduce suicide rates? A brief review. Eur Psychiatry. 2001 Nov;16(7):406–9.
4 Aarsland D, Larsen JP, Lim NG, Janvin C, Karlsen K, Tandberg E, et al. Range of neuropsychiatric disturbances in patients with Parkinson's disease. J Neurol Neurosurg Psychiatry. 1999 Oct;67(4):492–6.
5 Lauterbach EC. The neuropsychiatry of Parkinson's disease and related disorders. Psychiatr Clin North Am. 2004 Dec;27(4):801–25.
6 Schrag A. Psychiatric aspects of Parkinson's disease—an update. J Neurol. 2004 Jul;251(7):795–804.
7 Weintraub D, Stern MB. Psychiatric complications in Parkinson disease. Am J Geriatr Psychiatry. 2005 Oct;13(10):844–51.
8 Nazem S, Siderowf AD, Duda JE, Brown GK, Ten Have T, Stern MB, et al. Suicidal and death ideation in Parkinson's disease. Mov Disord. 2008 Aug;23(11):1573–9.
9 Reijnders JS, Ehrt U, Weber WE, Aarsland D, Leentjens AF. A systematic review of prevalence studies of depression in Parkinson's disease. Mov Disord. 2008 Jan;23(2):183–9; quiz 313.
10 Ishihara L, Brayne C. A systematic review of depression and mental illness preceding Parkinson's disease. Acta Neurol Scand. 2006 Apr;113(4):211–20.
11 Menza M, Dobkin RD, Marin H, Mark MH, Gara M, Buyske S, et al. The impact of treatment of depression on quality of life, disability and relapse in patients with Parkinson's disease. Mov Disord. 2009 Jul;24(9):1325–32.
12 Ravina B, Camicioli R, Como PG, Marsh L, Jankovic J, Weintraub D, et al. The impact of depressive symptoms in early Parkinson disease. Neurology. 2007 Jul;69(4):342–7.
13 France DJ, Shiavi RG, Silverman S, Silverman M, Wilkes DM. Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Trans Biomed Eng. 2000 Jul;47(7):829–37.
14 Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguist. 2007 Jan;20(1):50–64.
15 Stasak B, Epps J, Goecke R. Elicitation design for acoustic depression classification: An investigation of articulation effort, linguistic complexity, and word affect. Interspeech; 2017 Aug 20–24, Stockholm, p. 834–8. Available from: https://doi.org/10.21437/Interspeech.2017-1223.
16 Vlasenko B, Sagha H, Cummins N, Schuller B. Implementing gender-dependent vowel-level analysis for boosting speech-based depression recognition. Interspeech; 2017 Aug 20–24, Stockholm, p. 3266–70. Available from: https://doi.org/10.21437/Interspeech.2017-887.

17 Helfer BS, Quatieri TF, Williamson JR, Mehta DD, Horwitz R, Yu B. Classification of depression state based on articulatory precision. Interspeech; 2013 Aug 25–29, Lyon.
18 Cummins N. An investigation of depressed speech detection: features and normalization. Interspeech; 2011 Aug 27–31, Florence, p. 6–9.
19 Caligiuri MP, Ellwanger J. Motor and cognitive aspects of motor retardation in depression. J Affect Disord. 2000 Jan-Mar;57(1-3):83–93.
20 Syed ZS, Sidorov K, Marshall D. "Depression Severity Prediction Based on Biomarkers of Psychomotor Retardation," in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, New York, NY, USA, 2017, pp. 37–43.
21 Williamson JR, Quatieri TF, Helfer BS, Ciccarelli G, Mehta DD. "Vocal and Facial Biomarkers of Depression Based on Motor Incoordination and Timing," in Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, New York, NY, USA, 2014, pp. 65–72.
22 Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data. 2016 Mar;3:160011.
23 Little MA, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr DC, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. Gigascience. 2017 May;6(5):1–6.
24 Brookes M. Voicebox: Speech Processing Toolbox for Matlab. London: 1997.
25 Eyben F, Wöllmer M, Schuller B. openSMILE: the Munich versatile and fast open-source audio feature extractor. Proceedings of the 9th ACM International Conference on Multimedia; 2010. p. 1459.
26 Valstar M, et al. "AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. AVEC '13 Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge. 2013; Barcelona. p. 3–10. Available from: https://doi.org/10.1145/2512530.2512533.
27 Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. IEEE Trans Affect Comput. 2016 Apr;7(2):190–202.
28 Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. 2003; Stanford, CA. https://doi.org/10.1109/CSB.2003.1227396.
29 He H, Garcia EA. Learning from Imbalanced Data. IEEE Trans Knowl Data Eng. 2009 Sep;21(9):1263–84.
30 Longadge R, Dongre S. Class Imbalance Problem in Data Mining Review. ArXiv. 2013;1305.1707.
31 Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. ArXiv. 2014;1412.6980.
32 Goodarzi Z, Mrklas KJ, Roberts DJ, Jette N, Pringsheim T, Holroyd-Leduc J. Detecting depression in Parkinson disease: A systematic review and meta-analysis. Neurology. 2016 Jul;87(4):426–37.