

HIGH ACCURACY DISCRIMINATION OF PARKINSON'S DISEASE PARTICIPANTS FROM HEALTHY CONTROLS USING SMARTPHONES

Siddharth Arora^{1,2}, Vinayak Venkataraman³, Sean Donohue⁴, Kevin M. Biglan⁵,
Earl R. Dorsey^{5,6}, Max A. Little^{1,7}

¹Nonlinearity and Complexity Research Group, Aston University, Birmingham, United Kingdom

²Somerville College, University of Oxford, Oxford, United Kingdom

³Duke University School of Medicine, Durham, NC, United States

⁴Johns Hopkins University, Baltimore, MD, United States

⁵Department of Neurology, University of Rochester Medical Center, Rochester, NY, United States

⁶Center for Human Experimental Therapeutics, University of Rochester Medical Center, Rochester, NY, United States

⁷Media Lab, Massachusetts Institute of Technology, Boston, MA, United States

ABSTRACT

The aim of this study is to accurately distinguish Parkinson's disease (PD) participants from healthy controls using self-administered tests of gait and postural sway. Using consumer-grade smartphones with in-built accelerometers, we objectively measure and quantify key movement severity symptoms of Parkinson's disease. Specifically, we record tri-axial accelerations, and extract a range of different features based on the time and frequency-domain properties of the acceleration time series. The features quantify key characteristics of the acceleration time series, and enhance the underlying differences in the gait and postural sway accelerations between PD participants and controls. Using a random forest classifier, we demonstrate an average sensitivity of 98.5% and average specificity of 97.5% in discriminating PD participants from controls.

Index Terms— Gait, Postural sway, Smartphones, Parkinson's disease, Random forest, Tri-axial acceleration.

1. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative disorders, the prevalence of which is rapidly growing around the world. It is estimated that from 2010 to 2030, the number of individuals (65 years or older) with Parkinson disease will increase by 77% from 300,000 to 530,000 [1]. Furthermore, it is reported that for Western Europe's five and the world's ten, most populous nations, the number of people with PD over the age of 50 was between 4.1 and 4.3 million in 2005, and will double to about 8.7 to 9.3 million by 2020, whereby the burden of PD is expected to shift from Western nations to developing, Eastern nations [2].

There are no low-cost objective tests for the diagnosis of PD. Existing tests for the diagnosis of PD are based on subjective neurological examinations, which are performed in-clinic. This incurs considerable staff time and costs, and logistical costs for patients. There are no inexpensive, and yet reliable, tests that can be performed outside the clinic with minimal expert supervision, that can accurately track symptom progression. It is noteworthy that the current clinical consensus understanding of the progression of PD after diagnosis comes mostly from longitudinal symptom assessments obtained at monthly, and longer, intervals [3]. Comparatively little objective information about the fluctuation in symptoms on hourly, daily and weekly timescales is available. Given the growing number of people with Parkinson's (PWP) and the potential burden it will add to national healthcare services, there is tremendous need for ubiquitous, objective tests that can be used to support expert diagnosis and help improve the quality of life for PWP.

With advancements in wearable devices [4], several researchers have investigated the use of wearable accelerometers and other similar technologies to objectively measure and monitor key movement severity symptoms in PD and related disorders [5]-[9]. It is noteworthy that the typical motor characteristics of the PD include tremor, bradykinesia, rigidity, and impaired postural balance. Existing studies on monitoring PD motor symptoms typically use wearable accelerometers or sensors in conjunction with video recordings to obtain relevant data in a lab-based setting [8]-[9]. On the other hand, given that smartphones are relatively inexpensive and ubiquitous, have built-in tri-axial accelerometers, and can objectively measure the motor symptoms of PD [10]-[11], we investigate the efficacy of smartphones to be used as an inexpensive, reliable, and an accurate diagnostic support

tool for PD in a home and community setting. The method of this study is to calculate appropriate features from the acceleration time series, and thereby enhance the movement patterns relevant to PD participants and controls, and use these patterns to accurately discriminate PD participants from controls. The goal of this study is to reproduce the clinical assessment (PD/healthy) as accurately as possible, using only the acceleration time series for gait and postural sway tests.

2. METHODS

2.1. Data

We conducted a one-month controlled study with twenty, age- and gender-matched participants, comprising 10 PD participants and 10 controls. The Johns Hopkins Institutional Review Board reviewed and approved this study. Individuals with PD diagnosed clinically by a movement disorder specialist and control participants were recruited from an academic movement disorder clinic (Johns Hopkins). The baseline characteristics of all twenty participants are provided in Table 1.

All participants were provided identical LG Optimus S smartphones, capable of recording tri-axial acceleration. For the gait test, participants were instructed to walk 20 steps forward, turn around, and return back to the starting position. For the posture test, participants were instructed to stand upright unaided for 30 seconds. All participants were instructed to conduct these tests four times daily: just before taking their first (morning) dose of levodopa (or in one case, rasagiline), one hour later, in mid-afternoon, and before going to bed.

Table 1: Baseline characteristics of study participants.

<i>Characteristic</i>	<i>PD participants (N=10)</i>	<i>Control participants (N=10)</i>
Age (SD)	65.1 years (9.8)	57.7 years (14.3)
Percent taking levodopa	90%	0%
Percent with high school education	100%	100%
Baseline motor Unified Parkinson’s Disease Rating Score (SD)	19.6 (6.7)	NA
Baseline Parkinson’s Disease Questionnaire 39 score (SD)	18.5 (16.9)	NA

SD: standard deviation; NA: not applicable.

2.2. Feature Extraction

We extracted a range of time and frequency-domain features from the acceleration time series, partly inspired by the list of features extracted in [7]-[9]. We then use a classifier to map the features onto a binary diagnostic output variable (PD/healthy). Table 2 provides an exhaustive list of some of the features extracted in this study, along with a brief description of each.

Table 2: Brief description of features extracted from the tri-axial accelerometer time series recorded during postural sway and gait tests.

<i>Feature</i>	<i>Brief Description</i>
μ	Mean
σ	Standard deviation
Q_1	25 th percentile
Q_3	75 th percentile
IQR	Inter-quartile range ($Q_3 - Q_1$)
M	Median
M_o	Mode
R	Data range (maximum – minimum)
s	Skewness
k	Kurtosis
x_{mse}	Mean squared energy
S	Entropy
$r_{x,y}$	Cross-correlation between the acceleration in x and y -axis
$I(x; y)$	Mutual information between the acceleration in x and y -axis
$S(x, y)$	Cross-entropy between the acceleration in x and y -axis
1DFA	Extent of randomness in body motion
$mean$	Instantaneous changes in energy due to body motion
2TKEO	
ϕ_1	Autoregression coefficient at time lag 1
zcr	Zero-crossing rate
F_d	Dominant frequency component
r	Radial distance
θ	Polar angle
φ	Azimuth angle

¹DFA stands for Detrended Fluctuation Analysis.
²TKEO stands for Teager-Kaiser Energy Operator.

It is noteworthy that since the data was recorded using consumer-grade smartphone accelerometers, we were not able to compute some of the primary gait and postural sway metrics used in clinical ‘gait lab’ studies, for example, stride length, trunk flexion and minimum foot clearance, which can potentially be computed using more elaborate wearable instrumentation [7]. Encouragingly, despite this limitation, we were still able to differentiate PD participants from healthy controls with high accuracy, and thus establish the

feasibility of smartphones for quantifying key motor symptoms. As the acceleration time series were recorded at irregular time intervals, we applied the Lomb-Scargle periodogram to extract frequency-based features [12]. We extracted thirty different features from the acceleration time series recorded along the x , y and z -axis, and the derived absolute acceleration value (aav), $aav = \sqrt{x^2 + y^2 + z^2}$.

In addition to extracting features based on the time and frequency-domain properties of the acceleration time series, we extracted features based on spherical transformation of the tri-axial acceleration time series (radial distance, polar angle, and azimuth angle) [13], and demographics (age and gender).

2.3. Classification Technique and Classification Benchmarks

To provide a clinically meaningful output from the acceleration signals, we used a random forest classifier, which has shown excellent performance in very similar discrimination tasks, such as detecting Parkinson’s disease from digital voice signals [14]. Using random forest, we map the features extracted from the raw accelerometry signals into a determination of whether the participant had PD or was healthy. For details on random forests, see [15].

To investigate if the accuracy of the random forest classifier could have been obtained by chance, we employed the following two “naïve benchmarks” that have, by design, no discriminative accuracy:

1) *Random classifier* – this method is akin to diagnosing a subject as having PD based on flipping an unbiased coin. Specifically, this classifier assumes $P_{PD}:P_{Control} = 0.5:0.5$, where P_{PD} is the probability of a subject being diagnosed as having PD, while $P_{Control}$ is the probability of a subject being identified as a control. For example, a subject is classified as having PD if the outcome of a fair coin toss is heads; else, the subject is identified as a control.

2) *Conditional random classifier* – this method is similar to the random classifier, with one key difference: the probability of an outcome is conditional on the available data. This classifier is akin to using a *biased* coin, such that $P_{PD}:P_{Control} = N_{PD}:N_{Control}$, where N_{PD} and $N_{Control}$ are the number of gait and postural sway test instances available for PD participants and controls, respectively. This benchmark takes into account the difference in size of PD and control samples in the acceleration training data set.

2.4. Cross-validation

We used 10-fold cross-validation (CV) with 100 repetitions [16]. Using this CV method, the data is split into a training set (consisting of 90% of the data, used to train the classifier) and a testing set (consisting of the remaining 10% of the data, used for classifier evaluation). The 10-fold CV method involves training the methods ten times using the training set and evaluating using the test set, ensuring

that each acceleration signal occurs exactly once in the test set. This process was repeated 100 times, and each time the data was randomly shuffled prior to training and testing. Note that the association between the labels and data was kept intact during the process of random shuffling. This evaluation method attempts to assess the “generalizability” of the classification for other similar, previously unseen datasets, and so helps guard against overfitting.

3. RESULTS

To quantify the accuracy of the classifier in discriminating PD participants from controls, we employed three commonly used *performance measures*:

- 1) Sensitivity (*true positive rate*) – proportion of PD participants correctly identified.
- 2) Specificity (*true negative rate*) – proportion of controls correctly identified.
- 3) Balanced accuracy – average of sensitivity and specificity.

It is noteworthy that we classify PD participants from controls using only the accelerometry data and basic demographics such as age and gender. The classification performance results are reported in Table 3.

Table 3: Performance measures for the methods used for discriminating PD participants from controls.

<i>Method/ Measure</i>	<i>Sensitivity</i> [†]	<i>Specificity</i> [*]	<i>Balanced Accuracy</i> [‡]
Random Forest	98.5 ± 1.3	97.6 ± 1.7	98.0 ± 1.1
Random Classifier	50.0 ± 5.7	50.2 ± 5.7	50.1 ± 3.9
Conditional Random Classifier	67.7 ± 5.1	32.6 ± 5.3	49.9 ± 3.6

Note: Results are reported in percentage (%) in the form average ± standard deviation.
[†]Sensitivity = TP/(TP+FN), where TP refers to true positives, while FN stands for false negatives.

^{*}Specificity = TN/(TN+FP), where TN refers to true negatives, while FP stands for false positives.

[‡]Balanced Accuracy = (Sensitivity + Specificity)/2.

The two-sided Kolmogorov–Smirnov test was employed for group comparison [17]-[18]. Using this test, the corresponding distributions of different performance measures obtained using random forests and the naïve benchmarks were found to be significantly different ($p < 0.001$). Based on these results, we can reject the null

hypothesis that random forests have no discriminative accuracy, and statistically verify that the results are meaningful and not obtained just by chance.

4. SUMMARY AND CONCLUDING REMARKS

Using tri-axial accelerometry data for self-administered tests of gait and postural sway recorded via consumer-grade smartphones, we were able to distinguish PD participants from healthy controls with very high accuracy (Table 3). Thus, smartphones appear to be a feasible means for performing rapid, self-administered, objective tests of critical movement symptoms of PD outside the clinic.

Most clinical data on PD is low-frequency and recorded during periodic assessments in academic research centers. Given the feasibility of using smartphones to collect high-frequency data, future studies could analyze this day-to-day data to uncover substantial variations in key symptoms (i.e. gait, upper limb, balance, voice, and dexterity impairments) with PD.

In this study, we focused on discriminating PD participants from controls solely using accelerometer data. It would be worth investigating the efficacy of incorporating additional data (for example, reaction times, finger tapping, and voice) for the analysis. Future studies could investigate monitoring and predicting the severity of PD (as quantified using the Unified Parkinson's Disease Rating Scale (UPDRS) or Parkinson's disease questionnaire (PDQ-39)), in order to reproduce the clinicians' assessment using summary measures derived from smartphone data. Future probability density estimates of UPDRS or PDQ-39 generated for different time scales could help clinicians make informed decisions regarding drug dosage and timing for each individual patient.

5. ACKNOWLEDGEMENTS

The authors wish to thank Andong Zhan, for his contribution towards the development of smartphone software used in this study, and Reham Bedawy, for her help in identifying relevant literature for this work. We also extend our sincere gratitude to all the individuals who participated in this study and made this research possible.

6. REFERENCES

[1] E.R. Dorsey, B.P. George, B. Leff, and A.W. Willis. "The coming crisis: Obtaining care for the growing burden of neurodegenerative conditions," *Neurology*, 1989–1996, 2013.

[2] E.R. Dorsey, R. Constantinescu, J.P. Thompson, K.M. Biglan, R.G. Holloway, K. Kieburtz, F.J. Marshall, B.M. Ravina, G. Schifitto, A. Siderowf, and C.M. Tanner. "Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030," *Neurology*, 384–386, 2007.

[3] R. Kurlan and D. Murphy. "Parkinson's disease data and organizing center," *Movement Disorder*, 904–905, 2007.

[4] P. Bonato. "Advances in wearable technology and applications in physical medicine and rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, 2(1), 2005.

[5] J. Ghika, A.W. Wiegner, J.J. Fang, L. Davies, R.R. Young, and J.H. Growdon, "Portable system for quantifying motor abnormalities in Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, 276–283, 1993.

[6] S. Spieker, C. Jentgens, A. Boose, and J. Dichgans, "Reliability, specificity and sensitivity of long-term tremor recordings," *Electroencephalography and Clinical Neurophysiology*, 326–331, 1995.

[7] R.K. Begg, M. Palaniswami, and B. Owen. "Support vector machines for automated gait classification," *IEEE Transactions on Biomedical Engineering*, 828–838, 2005.

[8] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Transactions On Information Technology In Biomedicine*, 864–873, 2009.

[9] E. Sejdić, K.A. Lowry, J. Bellanca, M.S. Redfern, and J.S. Brach, "A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, in press.

[10] R.A. Joundi, J.S. Brittain, N. Jenkinson, A.L. Green, and T. Aziz. "Rapid tremor frequency assessment with the iPhone accelerometer," *Parkinsonism Related Disorders*, 288–290, 2011.

[11] M. Brunato, R. Battiti, D. Pruitt, and E. Sartori. "Supervised and unsupervised machine learning for the detection, monitoring and management of Parkinson's disease from passive mobile phone data. In: Predicting Parkinson's Disease Progression with Smartphone Data," Kaggle Competition. Available at: <https://kaggle2.blob.core.windows.net/prospector-files/1117/958625cf-3514-4e64-b0e7-13ebd3cf9791/kaggle.pdf>. Accessed October 2013.

[12] J.D. Scargle. "Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data," *Astrophysical Journal*, 835–853, 1982.

[13] D. Zwillinger (Ed.). *Spherical Coordinates in Space*. §4.9.3 in CRC Standard mathematical tables and formulae. Boca Raton, FL: CRC Press, 297–298, 1995.

[14] 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 Transactions on Biomedical Engineering*, 1264–1271, 2012.

[15] L. Breiman. "Random Forests," *Machine learning*, 5–32, 2001.

[16] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2nd edition, 2009.

[17] A.N. Kolmogorov. "On the empirical determination of a distribution function," *Breakthroughs in Statistics, Springer Series in Statistics*, 106–113, 1992.

[18] N.V. Smirnov. "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin of Moscow University*, 3–16, 1939.