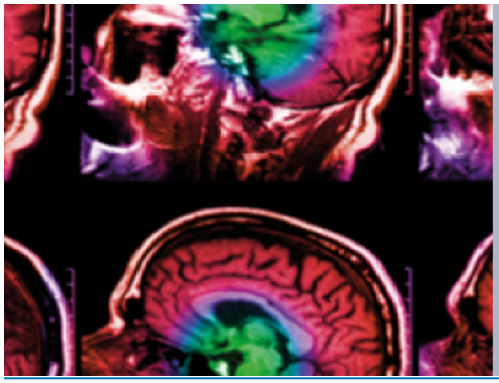
Biomedical Physics & Engineering Express

ENDORSED BY IPEM Institute of Physics and Engineering in Medicine

PAPER

Machine learning-based motor assessment of Parkinson's disease using postural sway, gait and lifestyle features on crowdsourced smartphone data

To cite this article: Hamza Abujrida *et al* 2020 *Biomed. Phys. Eng. Express* **6** 035005



Biomedical Physics & Engineering Express

PAPER

Machine learning-based motor assessment of Parkinson's disease using postural sway, gait and lifestyle features on crowdsourced smartphone data

RECEIVED
20 March 2019

REVISED
4 August 2019

ACCEPTED FOR PUBLICATION
8 August 2019

PUBLISHED
4 March 2020

Hamza Abujrida, Emmanuel Agu and Kaveh Pahlavan

Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609, United States of America

E-mail: hiabujrida@wpi.edu

Keywords: Parkinson's disease, gait analysis, smartphone sensing, machine learning, remote monitoring

Abstract

Objectives: Remote assessment of gait in patients' homes has become a valuable tool for monitoring the progression of Parkinson's disease (PD). However, these measurements are often not as accurate or reliable as clinical evaluations because it is challenging to objectively distinguish the unique gait characteristics of PD. We explore the inference of patients' stage of PD from their gait using machine learning analyses of data gathered from their smartphone sensors. Specifically, we investigate supervised machine learning (ML) models to classify the severity of the motor part of the UPDRS (MDS-UPDRS 2.10-2.13). Our goals are to facilitate remote monitoring of PD patients and to answer the following questions: (1) What is the patient PD stage based on their gait? (2) Which features are best for understanding and classifying PD gait severities? (3) Which ML classifier types best discriminate PD patients from healthy controls (HC)? and (4) Which ML classifier types can discriminate the severity of PD gait anomalies? *Methodology*: Our work uses smartphone sensor data gathered from 9520 patients in the mPower study, of whom 3101 participants uploaded gait recordings and 344 subjects and 471 controls uploaded at least 3 walking activities. We selected 152 PD patients who performed at least 3 recordings before and 3 recordings after taking medications and 304 HC who performed at least 3 walking recordings. From the accelerometer and gyroscope sensor data, we extracted statistical, time, wavelet and frequency domain features, and other lifestyle features were derived directly from participants' survey data. We conducted supervised classification experiments using 10-fold cross-validation and measured the model precision, accuracy, and area under the curve (AUC). *Results*: The best classification model, best feature, highest classification accuracy, and AUC were (1) random forest and entropy rate, 93% and 0.97, respectively, for walking balance (*MDS-UPDRS*-2.12); (2) *bagged trees* and MinMaxDiff, 95% and 0.92, respectively, for shaking/tremor (*MDS-UPDRS*-2.10); (3) *bagged trees* and entropy rate, 98% and 0.98, respectively, for freeze of gait; and (4) random forest and MinMaxDiff, 95% and 0.99, respectively, for distinguishing PD patients from HC. *Conclusion*: Machine learning classification was challenging due to the use of data that were subjectively labeled based on patients' answers to the MDS-UPDRS survey questions. However, with use of a significantly larger number of subjects than in prior work and clinically validated gait features, we were able to demonstrate that automatic patient classification based on smartphone sensor data can be used to objectively infer the severity of PD and the extent of specific gait anomalies.

1. Introduction and background

Parkinson's disease is a neurodegenerative disease with a worldwide prevalence estimated at 16 million people (Muangpaisan, Weerasak *et al* [1]). This number is expected to double by 2050. PD symptoms include tremor, rigidity, akinesia, and postural instability (Jankovic 2008 [2]), as shown in figure 1. In general, hospitalization is the largest component of PD health system costs (69% of total costs) (Shalika *et al* [3]).

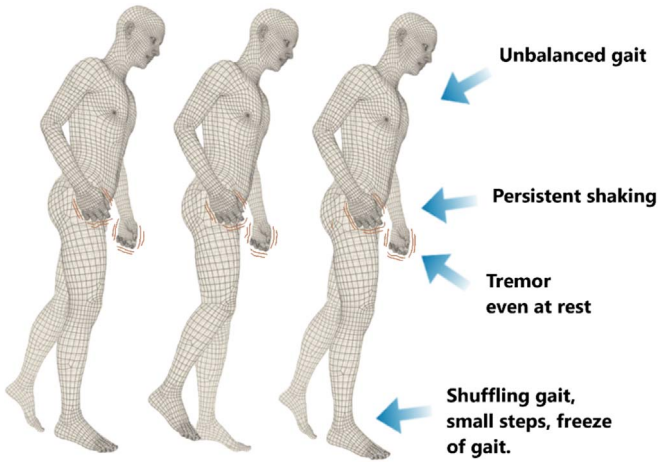


Figure 1. Symptoms of Parkinson's disease.

Because the condition of PD patients mostly improves at the early stages of the disease, it is important to detect PD as early as possible and monitor its progression in the early stages. Since the number of PD patients is increasing ([4] Kowal *et al* 2013), it is essential to reduce patient dependency on care by enabling them to perform PD assessment and monitor their progress at home ([5] Mera *et al* 2012).

Remote monitoring objectively informs the care provider of each patient's progress and how regularly they take their medication, information that can be used to make more accurate future treatment recommendations. However, home self-assessment has its own challenges and might not always generate reliable results. The need for remote assessment in addition to the boom in mobile health applications has led to active research on methods for remote monitoring of chronic diseases, including PD ([6] Kubota *et al* 2016).

As of January 2018, over 77% of adults in the United States own smartphones [7]. With this high adoption rate, smartphones also contain sensors that can be used in sensing the behavior and health of the smartphone user. Moreover, smartphones contain powerful processors for analyzing sensor data and supply a unique portable platform for mobile health applications. Researchers can conduct large-scale studies on millions of participants with the use of smartphone sensors to capture the fingerprint of the patient's unique behaviors. Smartphone sensing for healthcare can facilitate remote follow-up and enable doctors to fine-tune medication to meet each patient's needs. Several smartphone sensors have been demonstrated to be useful in assessing the symptoms of PD, including use of the microphone for voice analysis, the accelerometer/gyroscope for gait assessment and screen tapping for tremor/shaking evaluation ([8] Tsanas *et al* 2010, [9] Bayestehtashk *et al* 2015, [10] Patel *et al* 2009, [11] Horak and Mancini 2013, [12] LeMoyne *et al* 2013, [13] Kassavetis *et al* 2016). Mobile health applications could play a key role in reducing healthcare costs and the burden of PD on patients, especially those living in remote areas.

The most popular rating scale for PD is the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) ([14] Goetz *et al* 2008). On this scale, PD anomalies are rated on a scale of zero (normal) to four (severe PD). In this paper, we use the MDS-UPDRS as our PD measure and focus on two main goals to assess disease severity: (1) Identification of the best gait features that can be used to analyze MDS-UPDRS motor tests performed remotely, and (2) identification of which machine learning classification model best distinguishes the severity of PD anomalies for motor aspects of the MDS-UPDRS.

2. Our main contributions

- **A large sample size of patient data gathered in natural/home settings:** Although certain prior studies have investigated machine learning methods for classifying PD characteristics, most were based on a low number of participants or were conducted in clinical environments that might not be reproducible by patients who are self-monitoring in their homes. The studies were also inconclusive as to which specific features yielded the best classification results. Our study is based on a large number of participants (152 PD patients and 304 HC).
- **Comprehensive, clinically validated gait features:** Moreover, prior work did not explore selected gait features that have previously been used in clinical gait analysis and posturography, such as the sway area. Table 1 compares prior PD gait studies based on the number of subjects, aspect of PD analyzed, type of device/sensor used in capturing the data, and the location of the PD tests (home versus clinic). For completeness, we have included prior PD work that used other wearables such as smartwatches, infrared cameras, and force plates. Our study is

Table 1. Summary of the most recent PD GAIT studies.

Author	year	PD patients /controls	Aspect of PD	Device/Sensor	Test location	Limitation
[23] Mazilu, Sinziana *et al*	2012	10/0	FoG.	Smartphone and wearable accelerometers	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[24] Tahir *et al*	2012	12/20	Recognize gait pattern of PD.	Infrared camera/ force plate.	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[15] Bot *et al*	2016	5718/ 1087	Data collection	Smartphone sensors	Remote	Did not combine gait as well as lifestyle features.
[25] Arora *et al*	2015	10/10	Voice, posture, gait, finger tapping, and response time	Smartphone sensors	Clinic and remote	Small number of participants. Did not combine gait as well as lifestyle features.
[26] Ellis *et al*	2015	12/12	Gait variability	Smartphone accelerometer, gyroscope, and heel-mounted footswitch sensors	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[27] Printy *et al*	2014	26/0	Bradykinesia.	Smartphone gyroscope, accelerometer, touch screen, microphone, and front- camera	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[28] Kim, Hanbyul *et al*	2015	15/0	FoG	Smartphone gyroscope, accelerometer	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[29] Sharma, Vinod *et al*	2014	0/5	Facial tremors, speech, dyskinesia, gait abnormalities	Smartphone/Smartwatch accelerometer, front camera, microphone	Remote	Small number of participants. Did not combine gait as well as lifestyle features.
[16] Zhan, Andong *et al*	2016	121/105	Voice, balance, gait, dexterity, and reaction time.	Smartphone/accelerometer, touch screen, mic.	Remote	Did not combine gait as well as lifestyle features.
[30] Lee, Chae Young *et al*	2016	57/87	Bradykinesia	Smartphone/screen, mechanical tapper	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[13] Kassavetis, Panagiotis *et al*	2016	14/0	Tremor, bradykinesia	Smartphone/accelerometer, touch screen,	Clinic	Small number of participants. Did not combine gait as well as lifestyle features.
[31] Memedi, Mevludin *et al*	2013	95/10	Tremor, bradykinesia	Touch-pad handheld computer	Clinic and Remote	Small number of participants. Did not combine gait as well as lifestyle features.

based on smartphone data gathered by PD patients in their homes and investigates a comprehensive list of gait features, including clinically validated sway features.

- **Combination of lifestyle and demographic features with gait features:** Our study combines user demographics (age, gender) and lifestyle features (smoking, recent history of exercise) with gait features. All features are listed in table 2. Many lifestyle features have been shown in the literature to be predictive of PD status. Previous studies focused on either a set of lifestyle features or a set of gait features. Our selection of comprehensive gait features in addition to selective lifestyle features is novel. We believe that the novelty of feature selection and specifically combining time, frequency statistical features, sway area, and lifestyle features lead to impressive ML prediction results for a large set of participants.
- **Machine learning classification of freeze of gait (FoG), walking imbalance and shaking/tremors in the same study:** Our study classifies multiple PD aspects, targeting the inference of 3 major modalities of PD: (1) Freeze of gait (FoG), (2) walking balance, and (3) shaking/tremors. FoG is a brief episodic forward progression of the feet despite the intention to walk. FoG is associated with falls and negatively impacts patient quality of life. Walking imbalance is the inability to achieve postural equilibrium and orientation and includes decreased arm swing, shuffling steps, increased postural sway, and slow turning and often leads to falls. Shaking/tremors is an involuntary quivering movement, characteristically occurring at rest. The rhythmic tremor of Parkinson's disease typically starts in one limb and can eventually affect both sides of the body. The tremor is annoying and attracts attention, and PD patients tend to keep their bad hand in a pocket or sit on hands during meetings. We were able to rank and highlight the most important machine learning feature(s), which offers insights that can be used in the classification of PD gait modalities and identification of the best ML algorithm for analyzing each gait modality and the one that best discriminates HC from PD patients. We compared various supervised machine learning algorithms, which are explained in the methodology, results, discussion, and conclusion sections.
- **Crowdsourced smartphone sensor data:** We use smartphone sensor data gathered with a crowdsourcing methodology in the mPower study. Successful classification of data crowdsourced from patients could facilitate the population-level screening of PD patients. However, PD data collected in the home environment might be more heavily confounded by noise and more challenging to analyze.

Selected PD studies allowed participants to perform periodic assessment activities in their homes over extended periods of time using mobile health apps ([15] Bot *et al* 2016, [16] Zhan *et al* 2016). Accelerometer and gyroscope sensors have been demonstrated as useful in assessing gait, tremor and walking balance ([17] Abujrida *et al*). Those Previous studies were able to quantify multiple gait modalities of PD, including walking imbalance and FoG. Home capture yielded an increased number of participants compared with similar studies that were performed in clinics. However, this method of collecting data has it is own challenges ([6] Kubota *et al* 2016). Sources of error include factors such as the variance of different devices/sensors used, the lack of expert proficiency in subjects' self-assessment of gait severities, variability of the environment in which the assessment was performed, and level of subject adherence to the smartphone app instructions, which dramatically affects what subjects record as an observation. These factors ultimately lead to inconsistency in analyzing and classifying each activity and increase the margin of prediction error.

Lifestyle features are often ignored when analyzing PD gait. To the best of our knowledge, lifestyle features have not been considered when assessing gait severities of PD using engineered sensor features. ([18] Van *et al*) found that PD rapidly increased over the age of 60 years, with only 4% of the cases under the age of 50 years. The rate for men (19.0 per 100 000, 95% CI: 16.1, 21.8) was 91% higher than that for women (9.9 per 100,000, 95% CI: 7.6, 12.2). Smoking reduced tremor, rigidity, bradykinesia, and gait disturbance, including frozen gait. These effects lasted for approximately 10–30 min after smoking a cigarette and relieved PD symptoms in the off-medication period ([19] Ishikawa *et al*). Fertl ([20] Fertl *et al*) found a significant reduction in physical activity during the course of the disease, but no complete abandonment of sports was observed. Swimming, hiking, and gymnastics were the favored sports. ([21] Reuter *et al*) concluded that motor disability in PD patients can be improved by intensive sports activities in the early to medium stages of PD.

3. Methodology

3.1. Dataset

Data were acquired from the mPower study ([15] Bot *et al* 2016), a clinical observational study on PD conducted entirely through an iPhone (Apple Inc., Cuppertino CA, USA) app interface, figure 2 shows the versions of the smartphone used in the study. The mPower study interrogated aspects of movement disorder through surveys and continuous sensor-based recordings from participants with and without Parkinson disease. The mPower study had a large enrollment (n = 9520) of participants who opted to share data broadly and contributed at least two measurements. The

Table 2. Features definitions and descriptions.

S.N	Feature	Feature definition	Description
		Time Domain Features and their use cases for Gait analysis	
1	Number of Steps	Local Peaks	The number of steps taken in a given time interval
2	Average Step Time	$\frac{time}{\#Steps}$	The average time elapsed for each step
3	Average Cadence	$\frac{\#steps}{time}$	The ratio of the total number of steps to the total time
4	Skewness[a]	$\frac{\frac{1}{n}\sum(x_i - \mu_x)^3}{\left[\frac{1}{n}\sum(x_i - \mu_x)^2\right]^{3/2}}$	Asymmetry of the signal distribution
5	Coefficient of Variation of Step Time	$\frac{\sqrt{\frac{1}{n}\sum(interval_i - \mu_{interval})^2}}{\mu_{interval}}$	The within-subject standard deviation of the stride interval divided by the mean stride interval
6	Average Step Length	$\frac{0.084}{average\ Step\ Time} + 1.89$	The average distance covered by each step
7	Gait Velocity	$\frac{\left(\frac{0.084}{averageStepTime} + 1.89\right)}{averageStepTime}$	The ratio of the total distance covered by the total time
8	Minimum and Maximum Difference[a]	$\max(x_i) - \min(x_i)$	Global maximum of one step minus global minimum of one step averaged over all steps of one subject
9	Root Mean Square[a]	$\sqrt{\frac{1}{n}\sum {x_i}^2}$	Root Mean Square or quadratic mean is a statistical measure
10	Entropy Rate[a]	$-\sum possibility_{unique\ freq} \times \log_2(possibility_{unique\ freq})$	The uncertainty measure of the signal, and the regularity of a signal when its anticipated that consecutive data points are related
11	Sway Area[b] X.Y, Y.Z, X.Z	$\pi(AB)$	Area of an ellipse that encloses the 95 percent confidence interval of all observed gyroscope points in the XY, YZ, and XZ planes. (A and B are the lengths of the semi-major and semi-minor axes of the ellipse)
		Frequency Domain Features and their use cases for Gait analysis	
12	Harmonic Ratio[a]	$\frac{\sum_{i=1,3,5,\ldots} V_i}{\sum_{j=2,4,6,\ldots} V_j}$	Harmonic Ratio quantifies the harmonic composition of the accelerations for a given stride via DFT
13	Average Power[a]	$\frac{total\ power\ of\ the\ signal}{bandwidth\ of\ the\ signal}$	The mean of the total power underneath the curve of the PSD estimate for a signal
14	The ratio of Spectral Peak[a] (with Welch, FFT, DCT)	$\frac{\max(power_{freq})}{mean(power_{freq})}$	The ratio of the energies of low and high-frequency bands
15	Signal Noise Ratio[a]	$\frac{power_{signal}}{power_{noise}}$	Power of the whole signal over the power of its computed noise
16	The energy in Band 0.5 to 3Hz[a]	$\int_3^{0.5} psd_f\, df$	The energy in a frequency band describes components of distinct frequencies in the signal, and the frequency range is recommended as 0.5 Hz to 3 Hz
17	Windowed Energy in Band 0.5 to 3Hz[a]	$\int_3^{0.5} windowed\ psd_f\, df$	The energy in the frequency band of 5-s windows with an overlap of 2.5 s; windows from the complete signal sequence are averaged
18	Peak Frequency[a]	$\max(power_f)$	The maximum spectral power
19	Spectral Centroid[a]	$\frac{\sum f \times {power_f}^2}{\sum {power_f}^2}$	The frequency that divides the spectral power distribution into two equal parts
20	Bandwidth[a]	$\frac{\sum(f - spectralCentroid)^2 \times {power_f}^2}{\sum {power_f}^2}$	The difference between the uppermost and lowermost frequencies/range of frequencies in the signal (Weighted Average)
		Wavelet Domain Features and their use cases for Gait analysis	
21	Wavelet Bandwidth[a]	$\frac{cA'*cA}{(cA'*cA + cD'*cD)}$	The relative energy contribution in a time-frequency band

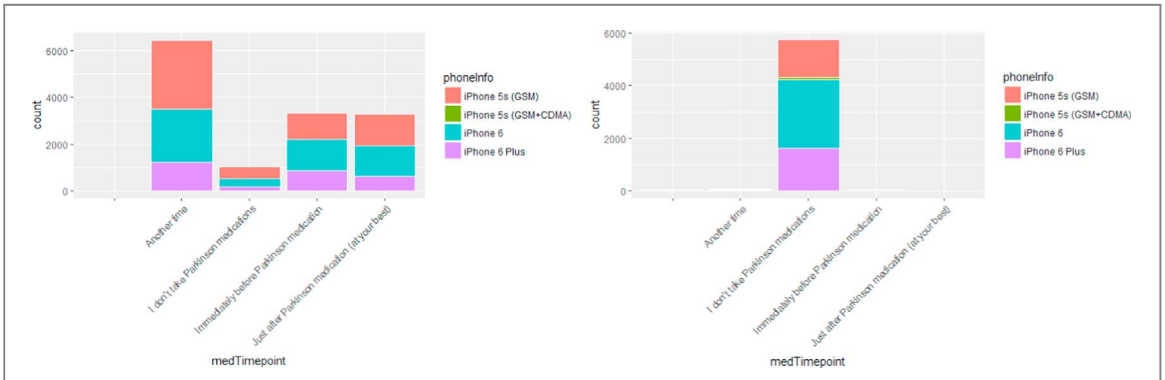
Table 2. (Continued.)

S.N	Feature	Feature definition	Description
		Time Domain Features and their use cases for Gait analysis	
22	Wavelet Entropy Rate[a]	$-\sum possibility_{unique\ freq} \times \log_2(possibility_{unique\ freq})$	Wavelet entropy represents signal disorder in the time-frequency domain
		Statistical Features and their use cases for Gait analysis	
23	Zeroth-Lag Cross-Correlation Coefficient[a]	$\frac{\sum(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum(x_i-\mu_x)^2\sum(y_i-\mu_y)^2}}$	The agreement or similarity between 2 directional acceleration signals
24	Kurtosis[a]	$\frac{\frac{1}{n}\sum(x_i-\mu_x)^4}{\left[\frac{1}{n}\sum(x_i-\mu_x)^2\right]^2}$	The extent to which the distribution of signal amplitudes lies predominantly on the left of the mean amplitude
25	Standard Deviation[a]	$\sqrt{\frac{1}{n}\sum(x_i-\mu_x)^2}$	Measure for signal spreading, defined as the square of standard deviation
		Lifestyle Features and their use cases for Gait analysis	
26	GELTQ.1a	The number of times the participant performed strenuous exercise for more than 15 min over the past week.	
27	GELTQ.1b	The number of times the participant performed moderate exercise for more than 15 min over the past week.	
28	GELTQ.1c	The number of times the participant performed minimal effort exercise for more than 15 min over the past week.	
29	Smoked	Ever smoked? (True/false question).	
30	Age	Participant's age (a number in years).	
31	Years. Smoking	Number of years participant has smoked (a number)	
32	Packs.per.day	Number of packets smoked per day.	
32	Gender	Female/male	

Table 2 shows the definition of features, where x_i refers to a data sequence from which the feature is to be calculated; μ_x refers to the average of all x_i; $interval_i$ refers to a sequence of stride intervals; $\mu_{interval}$ refers to the average of all $interval_i$; V_i refers to the amplitude of odd ordered harmonic frequency in the frequency domain, and V_j refers to that of even-ordered harmonic frequency; dwt refers to discrete wavelet transform (approximation coefficients vector cA and detail coefficients vector cD).

[a] Features that are calculated for both accelerometer and gyroscope.

[b] Features that are calculated for gyroscope only.



(a) Time of walking for PD Patients

(b) Time of walking for Healthy Controls

Figure 2. Smartphone version and Time of walking.

goal of the study was to help establish baseline variability of real-world activity measurement collected via mobile phones that might ultimately lead to quantification of the ebbs and flows of PD symptoms. The collected mPower activities included 35410 walking, 78887 tapping, and 8569 memory records. Subjects conducted the PD tests

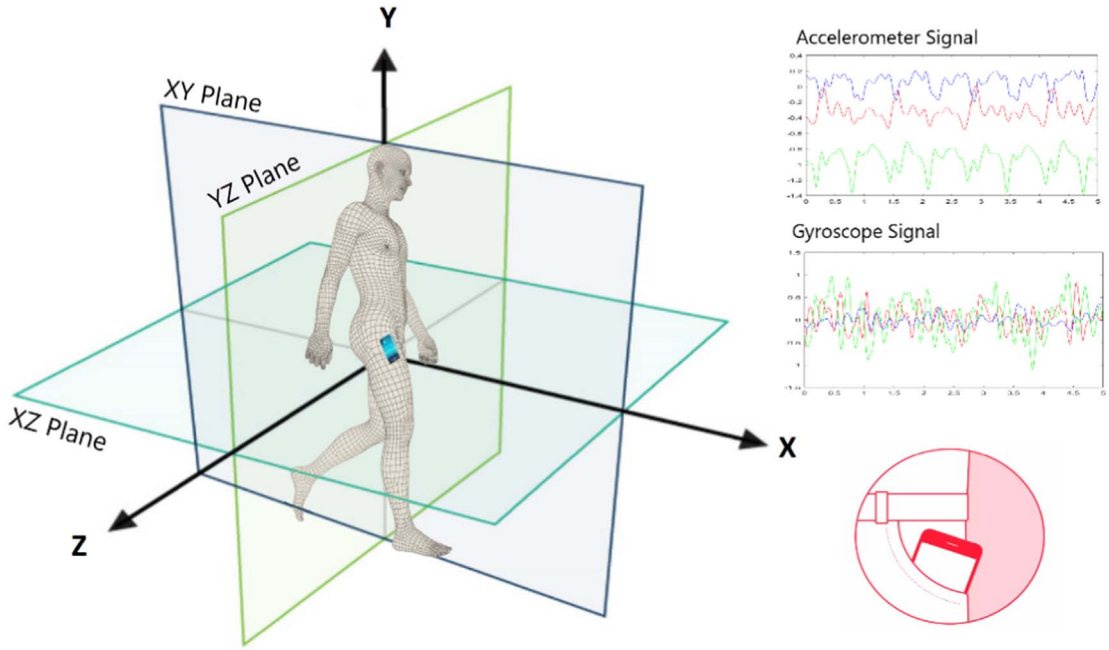


Figure 3. Participant Walking with Smartphone in pocket.

using an iPhone smartphone running the mPower data-gathering application. Participants self-reported PD severities and contributed activities several times during the day, before/after taking medication and at another time of the day, shown in figure 2. Not all participants complied with the application protocol, and therefore their number of recorded activates varied from a few to hundreds of recordings. Approximately 658 PD patients performed 24001 walks, and 2165 HC performed 10585 walking activities in the first 6 months of the study. Only 815 participants performed at least 3 walking activities.

Participants also filled out surveys including a subset of the UPDRS section 1 (non-motor experiences of daily living) and section 2 (motor experiences of daily living). Participants also completed a demographic survey, which included information on their general health history, PD history, and general lifestyle questions. To perform the walking activity, participants were asked to walk for 30 s in a straight line while placing the smartphone in their pants front pocket (figure 3), stand for 30 s, and subsequently turn and walk back for 30 s. The raw data gathered included the smartphone's gyroscope and accelerometer data sampled at 100 Hz, as well as pedometer values and the time of the activity. Our work only analyzes the outbound walking to enable uniform analysis across patients and healthy controls (HC) by avoiding FoG and walking imbalance events that usually occur when PD patients attempt to start walking after they turn between the outbound and inbound walks.

3.2. Selection of participants

In this paper, we included only PD patients that contributed 3 walks before and 3 after taking medication. Healthy controls were selected if they contributed at least 3 walks in total. We also excluded activity records in which key values of the demographic survey were missing or certain sensor readings were missing. Because the MDS-UPDRS survey data are highly important for this study, we filtered out participants whose survey data were not complete. In a few instances in which lifestyle questions had missing values, they were replaced (inputted) by the mean of the feature. The above subject selection rules yielded a working dataset with 152 PD patients (NPD = 152), and 304 healthy controls (NHC = 304). For participants who performed more than 3 walks in each category (before/after medication or at another time), we selected walks performed close to the date on which participants completed the demographic survey to increase the accuracy of labels for each activity. PD symptoms are known to calm after medication ([22] Haslinger B *et al*). Therefore, we wanted to capture the patient symptoms at peak occurrence, so we only analyzed the walks recorded before taking medication.

3.3. Signal preprocessing

The overall methodology that we followed is illustrated in figure 4. We began by pre-processing the sensor data and reorganized the sensor data into a readable format. We obtained the two main signals from the smartphone sensors, i.e., acceleration and rotation vectors from the accelerometer and gyroscope, respectively:

$$\alpha(i) = [\alpha x(i), \alpha y(i), \alpha z(i)]\, T(\text{in m s}^{-2})$$

$$\omega(i) = [\omega x(i), \omega y(i), \omega z(i)]\, T(\text{in deg s}^{-1})$$

where i denotes discrete time, α indicates accelerometer, and ω represents gyroscope.

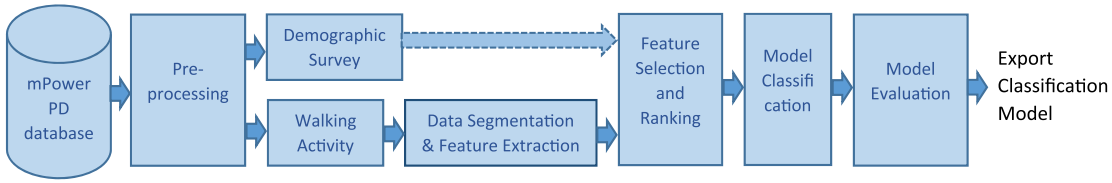


Figure 4. Flow Diagram for data collection, feature extraction, and classification.

PD tremor and balance classification have traditionally been captured by sway metrics derived from raw accelerometer values. However, we believe that features derived from the smartphone gyroscope data should supply additional information because it records angular velocity. To facilitate feature extraction, sensors signals are first divided into 5-s non-overlapping segments. The signals were subsequently smoothed by computing the moving average (n = 5) and removing sudden changes. The moving average calculation replaces each value in the sequence with the average of several points around it and is given by the following formula:

$$MA_{\alpha x} = \frac{1}{n}\sum_{n-1}^{i=0}(\alpha x_{(-i)})$$

The values of the three accelerometer and gyroscope axes are used to calculate the signal magnitudes, after which the signal's 5-s mean is subtracted to eliminate gravity or any constant factors such as breathing. The resulting formula is given below:

$$MagNG_{\alpha} = \left(\sum_{n}^{i=1}(\|\alpha_{(i)}\|) - \overline{Mag_{\alpha}}\right)$$

$$MagNG_{\omega} = \left(\sum_{n}^{i=1}(\|\omega_{(i)}\|) - \overline{Mag_{\omega}}\right)$$

where $MagNG\alpha$ and $MagNG\omega$ are the vector magnitudes of the acceleration and the rotation rates, respectively; and $\overline{Mag_{\alpha}}$ and $\overline{Mag_{\omega}}$ are the means of the acceleration and rotation rates, respectively. To compute the average step time metric, we initially found the peaks of the accelerometer axis with the largest magnitude. Only those peaks that are greater than a minimum peak height (MPH) are considered as a step. MPH is calculated by the following:

$$MPH_{\alpha} = \overline{Mag_{\alpha}} + \sigma_{\alpha}$$

where σ_{α} is the standard deviation of α.

Figure 5 shows a sample of the accelerometer and gyroscope signals on the three axes after preprocessing, smoothing and removing the gravity component. Figure 5(b) shows the peaks detected, which are used to estimate the steps for this walking segment. Our methodology does not require passing of the signal through filters.

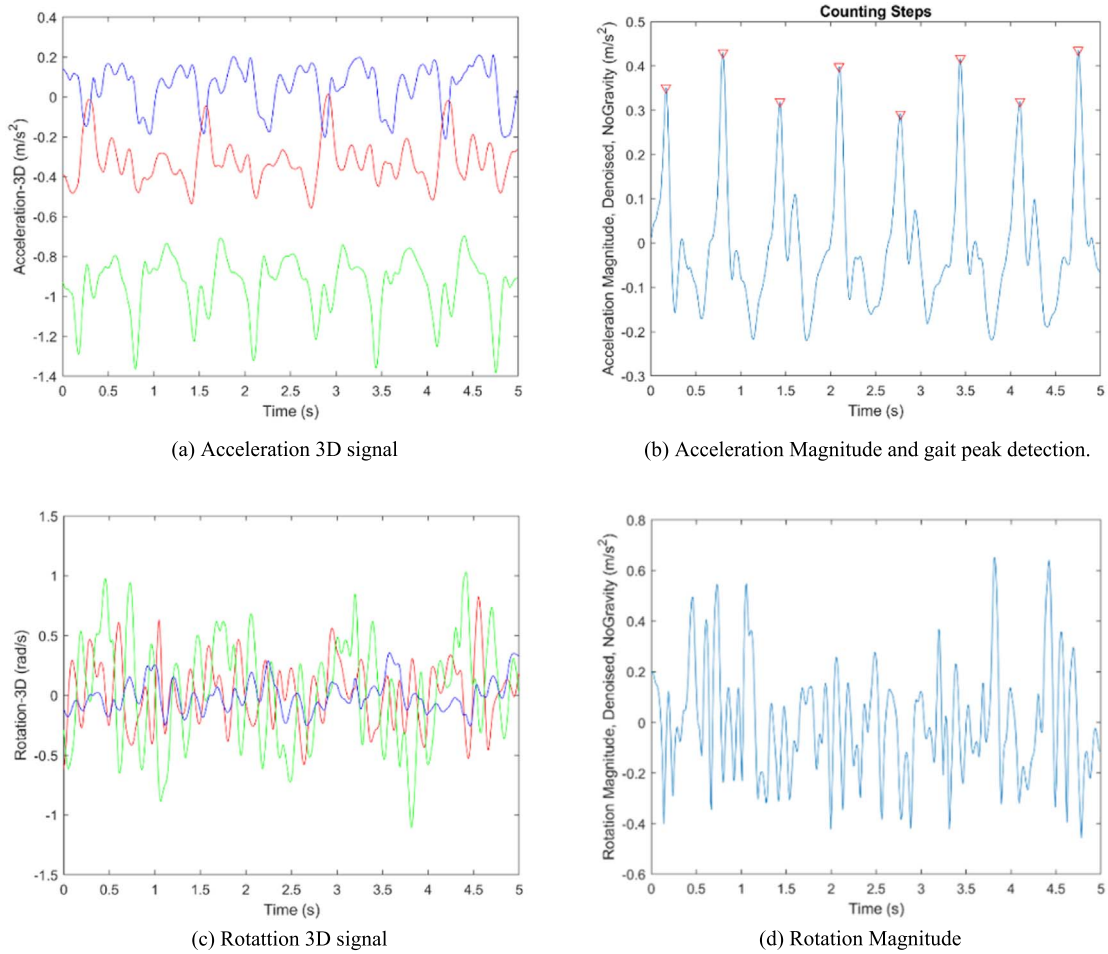
3.4. Feature extraction

From the two calculated magnitudes MagNGα and MagNGω (referred to as x_i in subsequent sections), we calculated the time, frequency, statistical and wavelet domain features shown in table 2. After pre-processing the data, gait features were extracted using Matlab (Mathworks version 2018b) from accelerometer and gyroscope data gathered during the walking activity. Sway area features were calculated for the gyroscope data. We subsequently combined the extracted features into larger data frames, and multiple datasets were derived and used in PD classification.

Time and statistical features were calculated directly from x_i. Frequency domain features were calculated after computing the fast Fourier transform (FFT) and power spectral density (PSD). Frequency domain features were subsequently extracted for each walking segment record. Wavelet domain features were calculated after calculating the discrete wavelet transform (DWT) of the *MagNGx* signal.

3.5. Sway area and posturography features for walking balance classification

Although certain prior studies were able to quantify abnormal sway levels in people with untreated PD, the majority of these prior studies were based on a small number of participants and measured sway by means of force plate or sensors connected to the bodies of the PD patients. To the best of our knowledge, no large studies (based on hundreds of patients) have extracted the sway area as a feature for PD patients using a smartphone gyroscope while walking at this scale of participants. Our calculation of sway area using the gyroscope is also novel because it has traditionally been calculated using accelerometer data. Our method for calculating sway area involves plotting values from two of the gyroscope axes, as shown in figure 6. The sway area is defined as the area of an ellipse that encloses the 95 percent confidence interval of all observed points. Using this methodology, we calculated the sway area for the XY, XZ and YZ planes, as shown in figure 6. This methodology is similar to that used to calculate the sway area using force plate readings. However, to the best of our knowledge, our use of the gyroscope to synthesize sway areas for PD patients has not been reported before. The novelty of using a smartphone for gait measurements is that the assessment can be performed while the patient is walking, which removes the need to explicitly perform the quiet standing test on force plates. This standing test measures the ground reaction forces generated by a body standing on or moving on the plates.



(a) Acceleration 3D signal

(b) Acceleration Magnitude and gait peak detection.

(c) Rotattion 3D signal

(d) Rotation Magnitude

Figure 5. Acceleration and Rotation signals and gait peak detection.

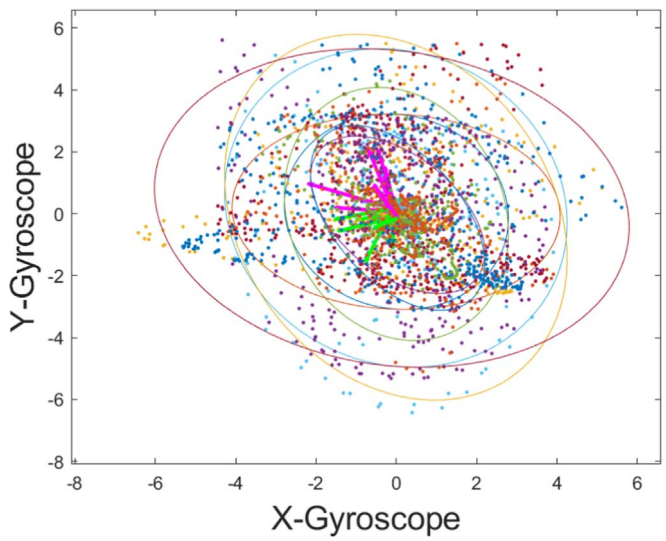
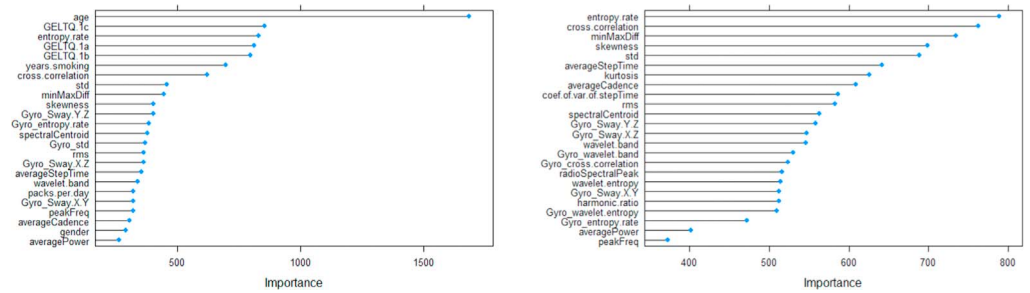


Figure 6. Sway Areas for multiple 5-s segments of the gyroscope XY plane.

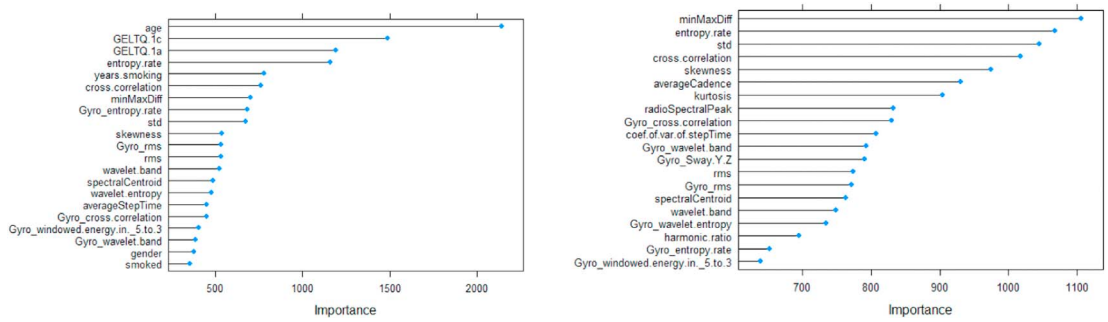
3.6. Model evaluation

To ensure efficient use of all the data, we conducted supervised classification experiments using 10-fold cross-validation and measured the model precision, accuracy, and area under the curve (AUC). Cross-validation performed by partitioning data into 10 disjoint folds at the population level. For each fold, we Train the model using the out-of-fold observations.



(a) Features by Importance for Walking Balance including lifestyle features (b) Features Importance for Walking Balance

Figure 7. (a) Features by Importance for Walking Balance including lifestyle features. (b) Features Importance for Walking Balance.



(a) Features by Importance for Shaking/Tremor including Lifestyle (b) Features Importance for Shaking/Tremor

Figure 8. (a) Features by Importance for Shaking/Tremor including Lifestyle. (b) Features Importance for Shaking/Tremor.

Then we assess the model performance using the in-fold data. The average performance metrics are calculated over all folds. Cross-validation requires multiple fits but gives a good estimate for the predictive accuracy of the final model trained and tested with all the data.

4. Results

In our prior work ([17] Abujrida *et al*), we used data from 50 participants in the mPower study to classify PD severity and were able to discriminate PD patients from healthy controls (HC) based only on gait features. In this work, we extended our prior work by increasing the number of participants analyzed and the derived gyroscope features, including the sway area feature computed on gyroscope data. We also included lifestyle features for all participants.

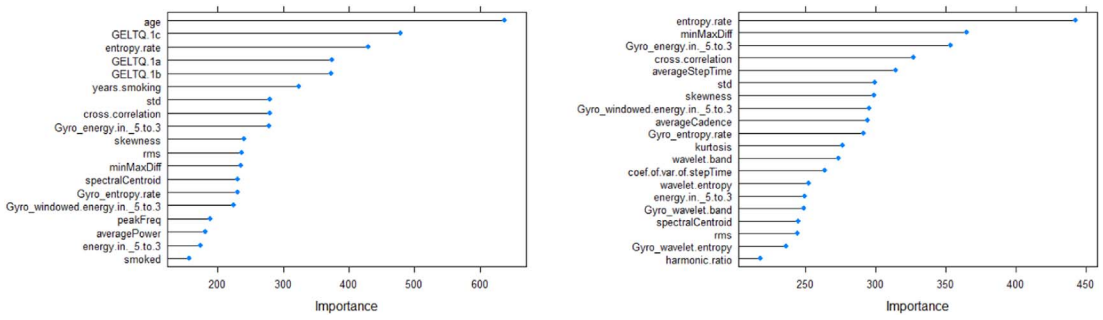
4.1. Feature selection

We established which features have statistically significant correlations with the MDS-UPDRS surveys and quantified the level of walking anomalies while patients walked for 30 s in a straight line. The walk data were labeled using participant self-assessments of their walk, which were used as labels for ML models. In constructing a decision tree, the importance of each feature is calculated by the decrease in the prediction error (mean squared error) and the increase of information gain, when the decision tree is split by the feature variable. Figures 7–9 below show the degree of importance of the selected features with and without consideration of lifestyle features.

From figure 7, the entropy rate is the most important feature for differentiating the walking balance severity. This finding agrees with the results of our prior work. As a demographic feature, age supersedes any calculated gait feature, which agrees with the findings in prior work ([18] Van Den Eeden) and ([32] Mayeux, Richard *et al*). However, the effect observed in this study is not the effect of aging on walking balance but is mostly a result of PD complications, as we address further in the discussion section.

Shaking and tremor can be inferred from gait features ([16] Zhan, Andong *et al*). However, most prior studies were based on a limited number of participants, as explained in table 1. Using data from participants in the mPower dataset, we identify the most important features that discriminate the level of shaking/tremor in figure 8. We noticed that the lifestyle features are strongly important in classifying shaking/tremor. We also noticed that multiple gyroscope features are highly important in predicting shaking/tremor.

Freeze of gait (FoG) has been studied extensively, and detection of FoG using smart sensors has been shown to be possible ([23] Mazilu, Sinziana *et al*) ([28] Kim, Hanbyul *et al*). Previous studies were based on a



(a) Features by Importance for FoG including Lifestyle (b) Features Importance for FoG

Figure 9. (a) Features by Importance for FoG including Lifestyle. (b) Features Importance for FoG.

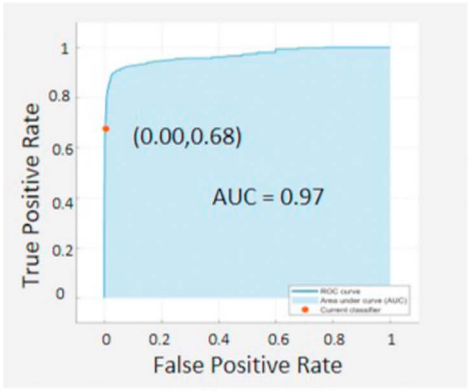
Table 3. Comparison of ML algorithms.

Walking balance		Precision	Accuracy	AUC
Accelerometer, Gyroscope	Random Forest	**92%**	**93%**	0.97
Posturography and Lifestyle Features	Bagged Trees	88%	90%	0.95
	Cubic SVM	72%	81%	0.92
	Weighted KNN	63%	82%	0.86
	Logistic Regression	71%	72%	0.78
	Fine Tree	75%	83%	0.88
	Quadratic Discriminant	71%	71%	0.75
Shaking Tremor		**Precision**	**Accuracy**	**AUC**
Accelerometer, Gyroscope	Random Forest	85%	83%	0.93
Posturography and Lifestyle Features	Bagged Trees	**95%**	**95%**	**0.92**
	Cubic SVM	63%	68.8%	0.86
	Weighted KNN	62%	68%	0.77
	Boosted Trees	71%	68%	0.83
	Fine Tree	60%	72%	0.87
	Linear Discriminant	48%	61%	0.74
Freeze of Gait (FoG)		**Precision**	**Accuracy**	**AUC**
Accelerometer, Gyroscope	Random Forest	92%	96%	0.90
Posturography and Lifestyle Features	Bagged Trees	**96%**	**98%**	**0.98**
	Fine Gaussian SVM	92%	93%	0.96
	Weighted KNN	91%	92%	0.95
	Boosted Trees	90%	91%	0.93
	Fine Tree	93%	94%	0.95
	Linear Discriminant	89%	87%	0.71

small number of participants and did not study the effect of lifestyle features and sway area as a gait feature. We were able to discriminate the severity of FoG reported by PD patients. We noticed that both the accelerometer and gyroscope features are highly predictive of FoG, as shown in figure 9.

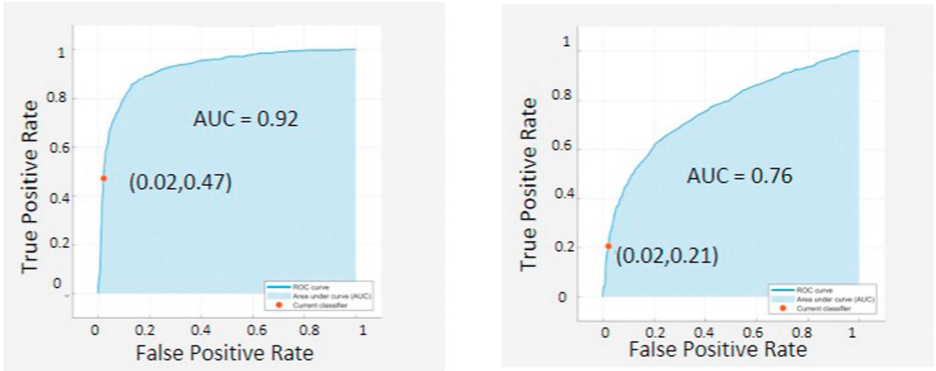
Using only the selected features shown in figures 7–9, we investigated multiple ML algorithms families, including decision trees (DT), discriminant analysis (DA), support vector machines (SVM), k-nearest neighbors (KNN), ensemble classifiers (EC) including random forest (RF) and logistic regression (LR) for classifying the walking balance, tremor/shaking and freezing of gait of 152 PD and 304 HC subjects. For walking balance, we found that entropy rate and cross-correlation were the best features for classifying the walking balance severity, with p-values of 0.2792094 and 2.481161e-07, respectively. Note that a lower p-value does not necessarily guarantee better ML performance. We discuss the effect of p-value on classification in additional detail in the discussion section.

Random forest is the best classifier for distinguishing between walking balance severities, with an accuracy of 93%, precision of 92% and AUC of 0.97. Table 3 compares the performance, accuracy, and AUC of each classifier type. For shaking/tremor, we found that MinMaxRate and EntropyRate were the best features for classifying severities. Bagged trees is the best classifier for distinguishing shaking/tremor severity with an accuracy of 95%, precision of 95% and AUC of 0.92. For FoG, features including entropy rate, MinMaxRate, and gyroscope energy successfully discriminated FoG severity. Bagged trees is the best classifier for distinguishing FoG severity with an accuracy of 98%, precision of 96% and AUC of 0.98.



(a) AUC for Walking Balance with lifestyle features. (b) AUC for Walking Balance without lifestyle features.

Figure 10. (a) AUC for Walking Balance with lifestyle features. (b) AUC for Walking Balance without lifestyle features.



(a) AUC for Shaking/Tremor with lifestyle features. (b) AUC for Shaking/Tremor without lifestyle features.

Figure 11. (a) AUC for Shaking/Tremor with lifestyle features. (b) AUC for Shaking/Tremor without lifestyle features.

4.1.1. Investigating PD classification with only gait features (excluding lifestyle features)

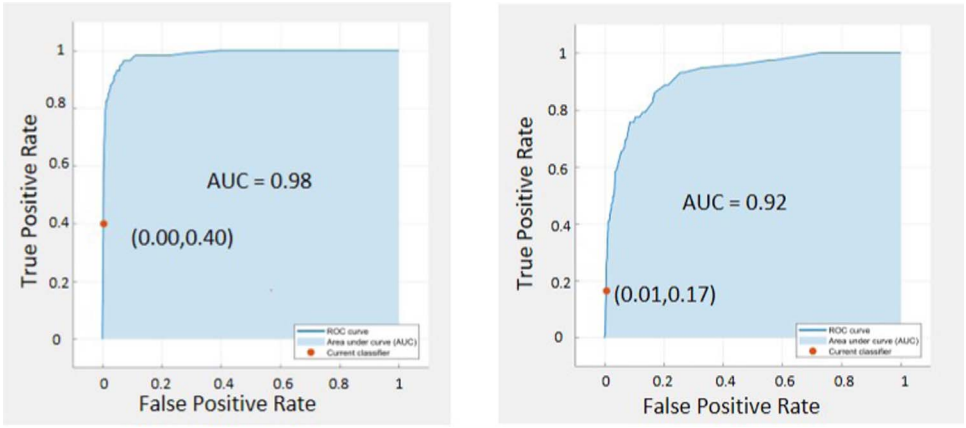
To isolate the effect of lifestyle and demographic/age features on model performance, we ran the ML classification algorithms using gait features while excluding lifestyle features, and the results of our classifiers were significantly degraded. For the walking balance, the best classifier results were random forest (Acc 77%, AUC 0.71). Figures 10(a), (b) shows the effect on the total performance of the model and how the result deteriorated. For shaking/tremor, the accuracy of the bagged trees models degraded to Acc 72% and AUC 0.76, as shown in figures 11(a), (b). For FoG, the accuracy of bagged trees slightly degraded to Acc 91% and AUC 0.92 but was still accurate. It can be noted from figures 12(a), (b) how AUC slightly deteriorated when lifestyle features were excluded.

4.2. PD Patients versus HC

One of the goals of this work is to be able to discriminate PD patients from healthy controls (HC) based on gait features. Using our selection of PD patients and HC, we performed ML analysis using the subject response to the question 'professionally diagnosed' on their enrolment questionnaire. This question was answered one time when participants filled out a demographic survey to report whether they had ever been diagnosed with PD.

Six ML algorithms were used to classify participant gait features, namely, bagged trees, fine Gaussian SVM, subspace KNN, boosted trees, fine tree, and linear discriminant. Entropy rate and MinMaxDiff were the top features to successfully discriminate HC from PD. Random forest was the best for discriminating between PD and HC, with an accuracy of 95%, precision of 94% and AUC of 0.99.

We presented in table 4 above comparison of the performance, accuracy, and AUC for the classifiers. In this comparison, lifestyle features helped to significantly improve the result of classification. Specifically, the false positive rate is significantly higher (worse AUC curve) if lifestyle features are not included. Removing the lifestyle features led to degradation of the classification results. Entropy rate and MinMaxDiff remained the top features. However, the accuracy of the random forest model deteriorated to Acc



(a) AUC for FoG with lifestyle features. (b) AUC for FoG without lifestyle features.

Figure 12. (a) AUC for FoG with lifestyle features. (b) AUC for FoG without lifestyle features.

Table 4. Comparison of ML algorithms for PD Patients versus HC Classification.

Classifier details		Precision	Accuracy	AUC
Accelerometer, Gyroscope	**Random Forest**	**94%**	**95%**	**0.99**
Posturography and Lifestyle Features	Bagged Trees	92%	93%	0.95
	Fine Gaussian SVM	88%	88%	0.96
	Subspace KNN	91%	90.4%	0.92
	Boosted Trees	84%	90%	0.97
	Fine Tree	90%	91.2%	0.96
	Linear Discriminant	83%	85.5%	0.91

82% and AUC 0.88. Figures 13(a), (b) below shows the top 20 features selected for discriminating between PD patients and HC with and without lifestyle features. Note that lifestyle features are less important in differentiating PD patients from HC.

The AUC also decreased significantly when lifestyle features were excluded. Figures 14(a), (b) below shows the AUC for HC/PD patients experiments with and without lifestyle features.

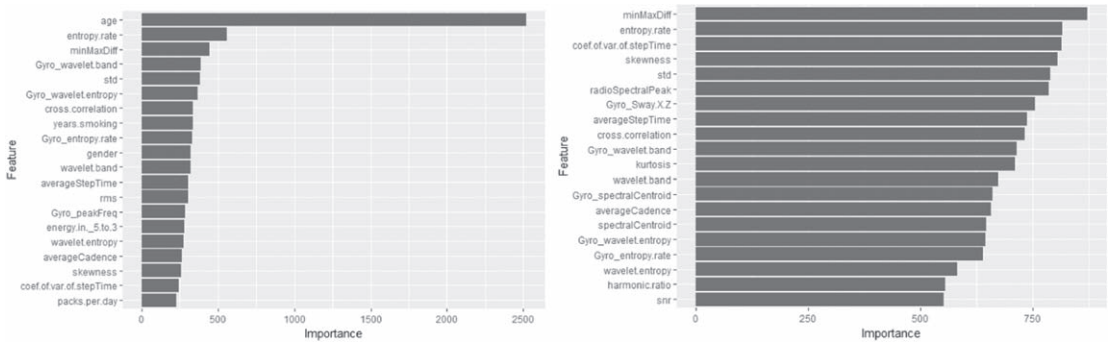
4.3. Top features

The ensemble model uses decision trees with high variance and low bias as base learners. At each node of a decision tree, the split feature is found based on information gain (I.G.) or the more computationally low-cost Gini impurity reduction method. The information gain due to a feature summed across all the levels of decision trees determines its feature importance. Random forest and bagged trees are composed of multiple decision trees, and thus the importance of a feature is the normalized sum of I.G. delivered by that feature across all trees. The output of these separate trees is aggregated and returned as the final ensemble result.

The correlation coefficient is commonly used to evaluate the degree of linear association between two variables. However, it can be shown that a correlation coefficient close to one might also be obtained for a clear curved relationship, depending on the nature of the ML algorithm used, and selection of features based on correlation can be misleading. We found that the selection of features based on the ML ensemble led to a set of features with high predictive power when used with nonlinear algorithms. Certain of our top features do not have a linear relationship with the response variable. For example, age does not correlate linearly with the label, as shown in figure 15(a). Prior studies have found that PD incidence rates for both men and women increased rapidly after the age of 60 years ([18] Van *et al*). Based on our normalized age feature, elderly people, in general, have higher severities, but as figure 15(a) shows, severity does not necessarily increase with age.

In contrast, top gait features (entropy rate and minMaxDiff) correlate linearly with gait severities. Figure 15(b) shows that minMaxDiff always increased as gait severities worsened, which occurs due to differences in step swing that are captured with accelerometer peaks. The mean of entropy rate (figure 15(c)) decreases with the increase in gait severities due to the irregularity of the walking signal associated with PD patients, which is captured by the accelerometer. Please note that the gait features in figure 15 are normalized on the participant level, whereas age is normalized on the population level.

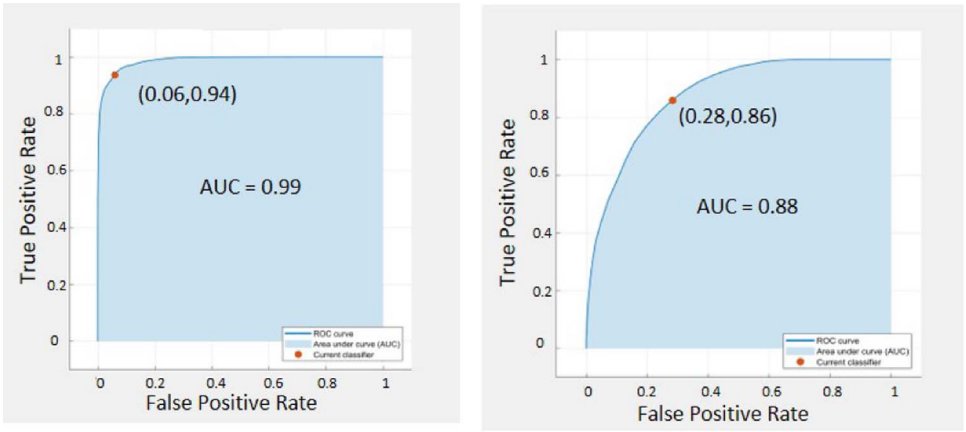
In our analyses, we found that the gyroscope sway areas contributed significantly to the classification of walking balance, as shown in figure 7(a), but it had



(a) Feature Importance for PD Patients vs HC (including lifestyle).

(b) Feature Importance for PD Patients vs HC (no lifestyle)

Figure 13. (a) Feature Importance for PD Patients versus HC (including lifestyle). (b) Feature Importance for PD Patients versus HC (no lifestyle).



(a) AUC for HC/PD Patient Classification with lifestyle features.

(b) AUC for HC/PD Patient Classification

Figure 14. AUC for PD patient versus HC classification with and without lifestyle features.

only minor contributions to shaking/tremor classification and no contribution to FoG classification. The accelerometer sensor and features were more useful in classifying shaking/tremor and FoG. Gyroscope-based analyses of sway can supply a powerful tool for early clinical trials and for monitoring the treatment efficacy for balance disorders in PD patients. Gyroscope sway area calculation can also be used in the online assessment of MDS-UPDRS walking balance (MDS-UPDRS 2.12).

4.4. Improving Accuracy with Ensemble methods

In this work, we attempted different variations of the ensemble random forest to tune its parameters. Random forest is composed of multiple estimators (decision trees) and aggregates their output to return the final ensemble result. If we have a classification problem with a data set in the form of $(X_1, Y_1),\ldots\ldots(X_n, Y_n)$, where X is a d-dimensional predictor variable and Y is a univariate response, to predict Y with J classes, $Y_i \in \{0, 1,\ldots, J-1\}$, the target function of interest is $P\,[Y = j|X = x](\,j = 0, 1,\ldots, J-1)$.

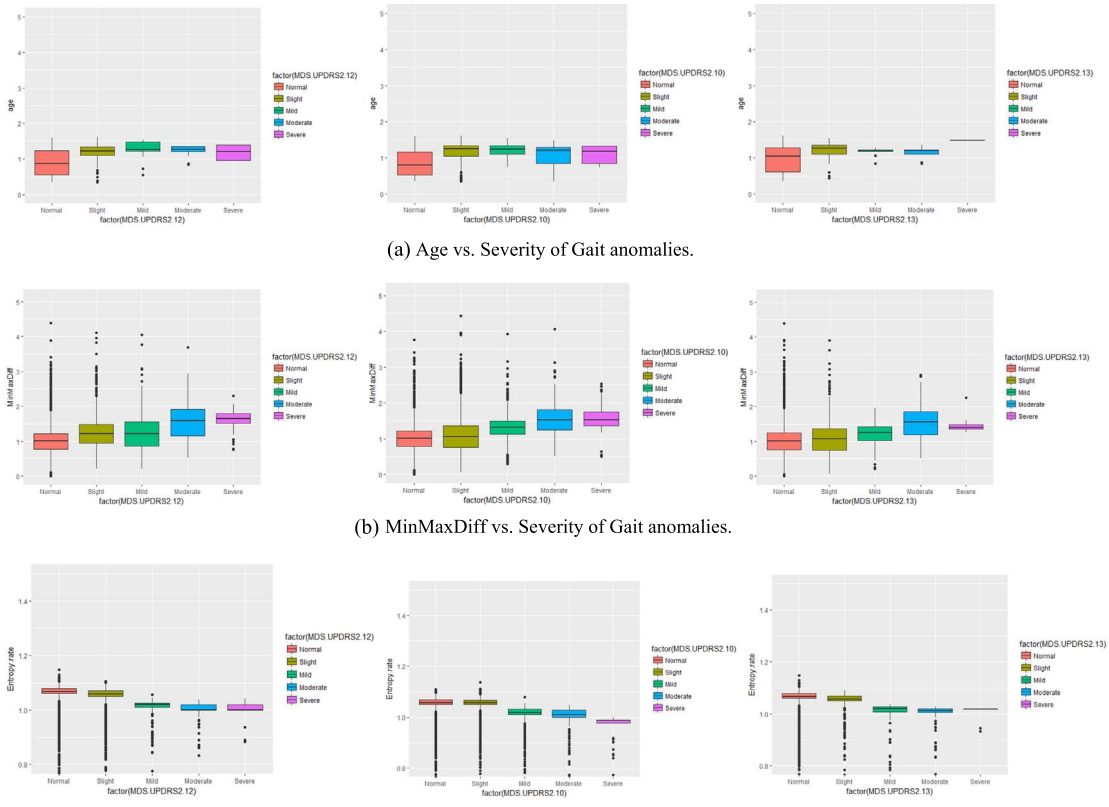
Random forest works by drawing $a_n$observations $(X'_1, Y'_1),\ldots\ldots(X'_n, Y'_n)$ at random with replacement from the original data set. These drawn observations are the only observations considered in growing M different randomized trees to obtain different estimates $g_1(\cdot)$, $g_2(\cdot)$, $g_3(\cdot)\ldots g_M(\cdot)$. In R random forest implementation, this number of (trees in the forest) is represented by (ntree). The resulting estimator functions can be written as follows:

$$g_n(\cdot) = h_n((X'_1, Y'_1),\ldots\ldots(X'_n, Y'_n))(\cdot)$$

where the function h_n (.)defines the estimator as a function of the dataset.

At each cell of each tree, a split is performed based on a number of variables mtry chosen randomly among the overall number of variables (p). The construction of individual trees is stopped when each cell contains less than nodesize points.

For any query point $x \in X_i$, each tree predicts Yi by growing the tree and making the final estimation that only depends on the a_n preselected data points. Because the overall decision is obtained via a majority vote among the classification trees, we can construct



(a) Age vs. Severity of Gait anomalies.

(b) MinMaxDiff vs. Severity of Gait anomalies.

(c) Entropy rate vs. Severity of Gait anomalies.

Figure 15. Top features versus severity of gait anomalies.

an ensemble-based function estimate gens(·) by taking linear combinations of the individual estimates:

$$g_{ens}(\cdot) = \sum_{k=1}^{M} (c_k g_k)$$

For ensemble bagging and (Breiman's [33]) original random forests, the linear combination coefficients $c_k = 1/M$ are averaging weights, which also result in variance reduction.

Tuning the forest parameters might result in a computational burden, particularly for large datasets with hundreds and thousands of observations and variables. Due to the manageable size of this study dataset, we tuned the following forest parameters with an affordable computational cost:

Number of trees to grow (ntree, _Acc)

Number of variables randomly sampled at each split (mtry_Acc)

Maximum number of terminal nodes (mx_Acc)

Minimum size of terminal nodes (nodeSize_Acc).

Figure 16 shows different parameters of the random forest ensemble method and their effects on the model performance. When addressing classification problems, it is usually recommended to set `nodesize` to 1, and `mtry` to $\sqrt{(p)}$ ([34] Liaw *et al*).

Extensive discussion exists in the literature relative to the influence of `mtry` on the overall performance of the model. ([35] Cutler *et al*) show that different values of `mtry` did not affect the classification rates of their model and that other performance metrics (sensitivity, specificity, kappa, and ROC AUC) were stable under different values of `mtry`. However, ([36] Strobl *et al*) show that mtry had a strong influence on predictor variable importance estimates. Additionally, ([37] Genuer *et al*) claim that the default value of mtry is too small. Therefore, their approach is to make mtry as large as possible (limited by available computing resources). We do not fully agree with the last finding, and we noticed that the overall accuracy improved significantly by increasing mtry. However, the relationship is not linear because the accuracy was maximized when mtry is 60% and 70% of p. By default, the maximum number of leaf nodes is set to the maximum possible. We experimented with limiting this parameter mx, which led to a negative impact on the overall accuracy.

It is clear that the forest variance decreases as M grows. Thus, more accurate predictions are likely to be obtained by choosing a large number of trees (ntree). The computational cost of increasing a forest increases linearly with M, and thus a good choice results from a trade-off between computational complexity and accuracy. Finally, the default value of the parameter

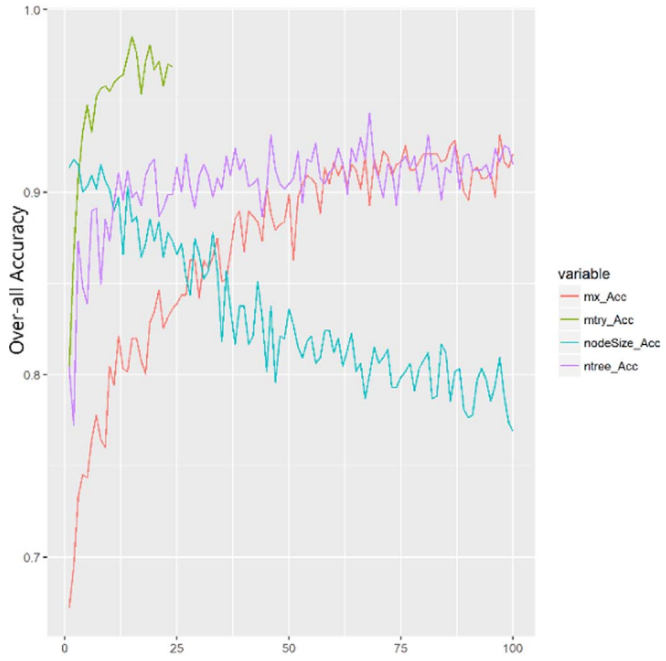


Figure 16. Random forest parameter tuning.

nodesize is 1 for classification and 5 for regression. These values are often reported as good choices ([38] Andrés *et al*), and even though this selection is not supported by solid theory, our results agree with those findings.

5. Discussion

We calculated and compared the feature importance computed using bagged trees and random forest Algorithms. While p-value helped in some instances to select the features most correlated with the response variables ([39] Schaafsma, Joanna D *et al*), many data scientists have raised concerns related to the validity of using the p-value as a significant measure for ML. For instance, Nuzzo *et al* posit that p-values, which are the 'gold standard' of statistical validity, are not as reliable as many scientists assume ([40] Nuzzo *et al*). Other alternatives to the p-value have also been proposed ([41] Lu *et al*). We found that the feature importance calculation based on machine learning is a better measure of feature significance. A p-value of $\leqslant 0.05$ did not always indicate the significance of a feature and did not lead to improvement of results if included in the ML classification.

6. Conclusion

Remote measurement of gait has become an important tool for monitoring the progression of PD. Although measurements reduce hospital visits and offer convenience to both PD patients and the healthcare provider, the validity of these measurements compared with assessments in the clinic continue to be a challenge. In this study, we addressed the unique gait characteristics of PD and inferred the stage of each PD gait modality through machine learning classification of smartphone sensor data collected by a mobile health application. This study contains three main contributions: (1) Combination of time, frequency, and statistical features with sway area and lifestyle features to remotely infer the level of PD walking modalities for a large set of participants; (2) identification of the most important features that offer deeper ailment understanding and classification of PD gait modalities; and (3) determination of the best ML algorithm for analyzing each gait modality and the one that best discriminates PD patients from HC. Although the classification results were affected by the subjective nature of PD labels assigned by patients based on their responses to the MDS-UPDRS questions, we were able to demonstrate with a relatively large number of participants that remote and automatic PD patient classification based on sensor activity data can supply objective assessments of PD-related gait patterns and severity of gait anomalies, which ultimately has the potential to improve remote healthcare for PD patients.

7. Future work

Planned future work includes exploration of various signal segmentation strategies, including various segment lengths, Bayesian segmentation, and overlapping segments. Neural networks, which have recently emerged as viable candidates for many smartphone sensing problems including activity recognition and gait assessment, will also be explored. Additional sway features such as sway velocity and acceleration will be

examined. Finally, we plan to test various ensemble classification methods such as gradient boosted machines and AdaBoost algorithms.

ORCID iDs

Hamza Abujrida https://orcid.org/0000-0002-7643-4775

References

[1] Muangpaisan W *et al* 2011 A systematic review of the worldwide prevalence and incidence of Parkinson's disease *Journal of the Medical Association of Thailand* **94** 749 http://jmatonline.com/index.php/jmat/article/view/57

[2] Jankovic J 2008 Parkinson's disease: clinical features and diagnosis *Journal of Neurology, Neurosurgery & Psychiatry* **79** 368–76

[3] Mudiyanselage Shalika B *et al* 2017 Cost of living with Parkinson's disease over 12 months in Australia: a prospective cohort study *Parkinson's Disease* **2017** 13

[4] Kowal S L *et al* 2013 The current and projected economic burden of Parkinson's disease in the United States *Movement Disorders* **28** 311–8

[5] Mera T O *et al* 2012 Feasibility of home-based automated Parkinson's disease motor assessment *J. Neurosci. Methods* **203** 152–6

[6] Kubota K J, Jason A C and Max A L 2016 Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures *Movement Disorders* **31** 1314–26

[7] Mobile Fact Sheet 2018 Pew Research Center Washington, DC 20036 http://pewinternet.org/fact-sheet/mobile/ February 5

[8] Tsanas A *et al* 2010 Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests *IEEE Trans. Biomed. Eng.* **57** 884–93

[9] Bayestehtashk A *et al* 2015 Fully automated assessment of the severity of Parkinson's disease from speech *Comput. Speech Lang.* **29** 172–85

[10] Patel S *et al* 2009 Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors *IEEE Trans. Inf. Technol.Biomed.* **13** 864–73

[11] Horak F B and Mancini M 2013 Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors *Mov. Disorders* **28** 1544–51

[12] LeMoyne R, Mastroianni T and Grundfest W 2013 Wireless accelerometer configuration for monitoring Parkinson's disease hand tremor *Adv. Parkinson's Dis.* **2** 62–7

[13] Kassavetis P *et al* 2016 Developing a tool for remote digital assessment of Parkinson's disease *Mov. Disorders Clin. Pract.* **3** 59–64

[14] Goetz C G *et al* 2008 Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results *Movement disorders: Official Journal of the Movement Disorder Society* **23** 2129–70

[15] Bot B M *et al* 2016 The mPower study, Parkinson disease mobile data collected using Research Kit *Scientific Data* **3** 160011

[16] Zhan A *et al* 2016 High frequency remote monitoring of Parkinson's disease via smartphone: platform overview and medication response detection *arXiv* - 12 pages arXiv preprint arXiv:1601.00960

[17] Abujrida H, Agu E and Pahlavan K 2017 Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data *Healthcare Innovations and Point of Care Technologies (HI-POCT) 2017 IEEE. IEEE*

[18] Eeden V D *et al* 2003 Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity *American journal of epidemiology* **157** 1015–22

[19] Ishikawa A and Miyatake T 1993 Effects of smoking in patients with early-onset Parkinson's disease *J. Neurol. Sci.* **117** 28–32

[20] Fertl E, Doppelbauer A and Auff E 1993 Physical activity and sports in patients suffering from Parkinson's disease in comparison with healthy seniors *Journal of Neural Transmission-Parkinson's Disease and Dementia Section* **5** 157–61

[21] Reuter I *et al* 1999 Therapeutic value of exercise training in Parkinson's disease *Medicine and Science in Sports and Exercise* **31** 1544–9

[22] Haslinger B *et al* 2001 Event-related functional magnetic resonance imaging in Parkinson's disease before and after levodopa *Brain* **124** 558–607

[23] Mazilu S *et al* 2012 Online detection of freezing of gait with smartphones and machine learning techniques *Pervasive Computing Technologies for Healthcare (PervasiveHealth) 2012 6th Int. Conf. on. IEEE*

[24] Tahir N M and Manap H H 2012 Parkinson disease gait classification based on machine learning approach *Journal of Applied Sciences* **12** 180–5

[25] Arora S *et al* 2015 Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study *Parkinsonism & Related Disorders* **21** 650–3

[26] Ellis R J *et al* 2015 A validated smartphone-based assessment of gait and gait variability in Parkinson's disease *PLoS One* **10** e0141694

[27] Printy B P *et al* 2014 Smartphone application for classification of motor impairment severity in Parkinson's disease *Engineering in Medicine and Biology Society (EMBC) 2014 36th Annual Int. Conf. of the IEEE. IEEE*

[28] Kim H *et al* 2015 Unconstrained detection of freezing of Gait in Parkinson's disease patients using smartphone *Engineering in Medicine and Biology Society (EMBC) 2015 37th Annual Int. Conf. of the IEEE. IEEE*

[29] Sharma V *et al* 2014 SPARK: personalized parkinson disease interventions through synergy between a smartphone and a smartwatch *International Conference of Design, User Experience, and Usability* (Cham: Springer)

[30] Lee C Y *et al* 2016 A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease *PLoS One* **11** e0158852

[31] Memedi M *et al* 2013 Automatic and objective assessment of alternating tapping performance in Parkinson's disease *Sensors* **13** 16965–84

[32] Mayeux R *et al* 1992 A population-based investigation of Parkinson's disease with and without dementia: relationship to age and gender *Arch. Neurol.* **49** 492–7

[33] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

[34] Liaw A and Wiener M 2002 Classification and regression by random *Forest. R News* **2** 18–22

[35] Cutler D R *et al* 2007 Random forests for classification in ecology *Ecology* **88** 2783–92

[36] Strobl C *et al* 2008 Conditional variable importance for random forests *BMC Bioinf.* **9** 307

[37] Genuer R, Poggi J-M and Tuleau-Malot C 2010 Variable selection using random forests *Pattern Recognit. Lett.* **31** 2225–36

[38] Díaz-Uriarte R and Alvarez de Andrés S 2006 Gene selection and classification of microarray data usingrandom forest *BMC Bioinform* **7** 1–13

[39] Schaafsma J D *et al* 2003 Gait dynamics in Parkinson's disease: relationship to Parkinsonian features, falls and response to levodopa *J. Neurol. Sci.* **212** 47–53

[40] Nuzzo R 2014 Scientific method: statistical errors *Nature News* **506** 150

[41] Lu M and Ishwaran H 2017 A machine learning alternative to P-values *arxiv* 17 arXiv preprint arXiv:1701.04944