

Comparative Motor Pre-clinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches

ERIKA ROVINI,¹ CARLO MAREMMANI,² ALESSANDRA MOSCHETTI,¹ DARIO ESPOSITO,¹ and FILIPPO CAVALLO ¹

¹The BioRobotics Institute, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio, 34, 56025 Pontedera, Italy; and ²U.O. Neurologia, Ospedale Delle Apuane (AUSL Toscana Nord Ovest), Viale Mattei, 21, 54100 Massa, Italy

(Received 19 March 2018; accepted 19 July 2018; published online 20 July 2018)

Associate Editor Andreas Anayiotos oversaw the review of this article.

Abstract—Millions of people worldwide are affected by Parkinson's disease (PD), which significantly worsens their quality of life. Currently, the diagnosis is based on assessment of motor symptoms, but interest toward non-motor symptoms is increasing, as well. Among them, idiopathic hyposmia (IH) is associated with an increased risk of developing PD in healthy adults. In this work, a wearable inertial device, named SensFoot V2, was used to acquire motor data from 30 healthy subjects, 30 people with IH, and 30 PD patients while performing tasks from the MDS-UPDRS III for lower limb assessment. The most significant and non-correlated extracted parameters were selected in a feature array that can identify differences between the three groups of people. A comparative classification analysis was performed by applying three supervised machine learning algorithms. The system resulted able to distinguish between healthy and patients (specificity and recall equal to 0.967), and the people with IH can be identified as a separate class within a three-group classification (accuracy equal to 0.78). Thus, the system could support the clinician in objective assessment of PD. Further, identification of IH together with changes in motor parameters could be a non-invasive two-step approach to investigate the early onset of PD.

Keywords—Decision support systems, Idiopathic hyposmia, Inertial wearable sensors, Motion analysis, Supervised learning.

INTRODUCTION

Clinical Background

Parkinson's disease (PD) is a common neurodegenerative pathology that currently affects millions of people worldwide with increasing rate.¹⁰ PD is caused

by a critical loss of dopamine in the forebrain and is characterized by cardinal motor symptoms (i.e., tremor, postural instability, muscular rigidity, bradykinesia)¹¹ and also by many non-motor manifestations (NMMs) (e.g., sleep disorders, bladder disturbances, olfactory symptoms), which are widely disabling for PD patients,³² hugely worsening their quality of life. Currently, PD diagnosis is primarily based on the assessment of motor symptoms according to diagnostic criteria.¹³ Clinical scales, such as the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS)¹⁵ and the Hoehn & Yahr (HY),¹⁷ are typically used to assess PD patients, assigning a score according to the disease severity. These scales are semi-quantitative, and the assigned scores are subjective and affected by variability. Specifically, in the MDS-UPDRS section III motion is evaluated and patients are asked to perform detailed tasks (e.g., finger and foot tapping, gait), which are visually observed and assessed by neurologists. However, small worsening in subjects' performance are difficult to detect, thus, quantitative assessment of motion capabilities for PD patients seems to be mandatory to obtain an objective evaluation of disease onset and progression. In this context, wearable sensors play a fundamental role thanks to their accuracy, unobtrusiveness, and ease of use, both in the hospital and for long-term monitoring,^{30,36} thus representing a solution for supporting clinicians' evaluations through the measurement of motor performance.

Meanwhile, an increasing interest is emerging also toward NMMs, because they are involved in the neuropathological changes in the brain and, as such, they can anticipate the motor manifestations of the disease by 5–7 years.²⁴ Recently, experts have recognized the

Address correspondence to Filippo Cavallo, The BioRobotics Institute, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio, 34, 56025 Pontedera, Italy. Electronic mail: filippo.cavallo@santannapisa.it

need for new diagnostic criteria⁷ which aim to investigate and treat the pathology since its beginning, when small neurological areas are damaged and neuroprotective therapies could slow down the PD progression.³¹ Among the NMMs, idiopathic hyposmia (IH) is a common symptom affecting > 95% of PD patients.¹⁶ Hyposmia is a reduced olfactory sensitivity accompanied by poor perception of flavors that is observed independently from disease duration. IH in healthy adults is associated with a 10–12% increase in risk of developing PD.^{23,28} Since IH shows high sensitivity, but low specificity for PD development, it is not sufficient to identify people with early PD. Imaging diagnostic techniques, able to investigate cerebral pathologies, are currently used to confirm the PD diagnosis, but they require specially trained dedicated staff and are invasive, such as the single proton emission computed tomography (SPECT) DaTSCAN, expensive, such as the nuclear magnetic resonance with diffusion tensor imaging (NMR-DTI), or not specific enough, such as the Transcranial Sonography. Particularly, SPECT is a cerebral scintigraphy performed using the DaTSCAN[®], a substance developed by General Electric Healthcare to reveal the presence of dopamine transmitters in the brain. Since PD is caused by critical deficiency of dopamine in the forebrain, scanners of PD patients, for instance, are able to show low level of dopamine.³ Differently, NMR-DTI is a nuclear magnetic resonance with diffusion tensor imaging that investigates the PD pathophysiology studying the white fiber integrity.⁵ In a recent study, NMR-DTI enabled to discover structural disruptions in certain brain networks of people with PD and cognitive impairment.¹²

Hypothesizing that IH can be assumed as a pre-clinical biomarker for PD, a two-step approach can represent a valuable solution to identify the early onset of PD²² by combining olfactory screening and non-invasive measurements of motion capabilities in IH subjects.

Related Works

In several recent studies, researchers applied machine learning techniques to classify PD patients and healthy subjects of control (HC) based on their motion capabilities in the lower limbs (Table 1). They used different types of technologies, including accelerometers (ACC), gyroscopes (GYR), electromyography (EMG), and force sensors. The majority of the studies focused on gait only, which represents a single task in MDS-UPDRS III, whereas only two works^{6,21} included alternative exercises, such as foot tapping. Further, three out of nine works analyzed a smaller dataset,^{4,18,21} which involved 15 or fewer PD patients.

Different machine learning approaches were implemented in these studies to distinguish between two groups (i.e., healthy subjects and PD patients) with prevalence of Support Vector Machine (SVM),^{2,6,18,21,33,34} Linear Discriminant Analysis (LDA),^{6,25,27} and Random Forest (RF).^{2,4,34} Among these classifiers, LDA seemed to achieve the overall worst results in terms of misclassification rate,²⁵ sensitivity, and specificity (reported in Table 1 as “sens.” and “spec.”, respectively).⁶ Thus, it will not be used in the proposed work.

Actually, only in one study³³ did researchers try to implement a multiclass algorithm, dividing PD patients into two groups—patients with and patients without significant gait impairments—but results for class recall and class precision were poor (i.e., on average < 70%).

To the best of our knowledge, no studies have been conducted that have tried to apply these algorithms to people at risk for developing PD (i.e., IH subjects). Currently, only imaging diagnostic techniques have demonstrated that idiopathic hyposmia is a preclinical marker of this pathology.¹⁶

This work aims to investigate differences in motor performance in the lower limbs of different classes of people (i.e., healthy subjects, people with IH, and PD patients), measured by using wearable inertial measurement units (IMU), comparing different supervised machine learning approaches. In particular the authors propose to:

- (i) analyze a comprehensive experimental protocol for a complete motor evaluation of lower limbs in PD patients using four exercises (i.e., leg agility, toe tapping, gait, and rotation) according to the tasks described in MDS-UPDRS III, while increasing the number of exercises compared to previous works;
- (ii) evaluate a wide set of extracted features related to biomechanical characteristics of the users that allow examination of spatial, temporal, and frequency parameters of the lower limb movements;
- (iii) investigate the performance of supervised machine learning approaches for motor assessment of lower limb performance in Parkinson’s disease by comparing three different classifiers (i.e., SVM, RF, and NB).

MATERIALS AND METHODS

Participants

Thirty HC (25 male, 5 female, mean age \pm standard deviation [SD] 65.2 ± 2.5), 30 IH subjects (21 male, 9

TABLE 1. Related works.

Ref	Technology	Experimental design	Subjects	Features extracted	Classifiers	Classifier performance (PD/HC)
6	ACC, GYRO on foot (100 Hz)	10 m Walking: circling the foot (20 s); heel toe tapping (20 s)	14 early and 13 mid PD, 16 HC	Step duration, rise/fall gradient of swing phase, SD of min, max-min difference, variance, integral, dominant freq, energy ratio, energy in band 0.5-3 Hz and in band 3-8 Hz	Boosting with decision stump, LDA, SVM	LDA: 88% sens., 86% spec. early PD/HC, 100% sens., 100% spec. for therapy monitoring (mid PD/HC, early PD/mid PD)
21	ACC, EMG on heels (296.3 Hz)	10 m Walk; heel toe tapping (30 s); foot rotation (30 s)	5 PD; 5 HC	Statistical features: variance, skewness, kurtosis; RMS energy; frequency features: mean, median and dominant frequency, total power	SVM (linear, leave-one-out cross validation)	90% sens., 90% spec.
4	Smartphone with ACC	Walk 20 steps forward, turn around and return back (1 month controlled study)	10 PD, 10 HC	Mean, SD, 25th and 75th percentile, IQR, median, mode, data range, skewness, kurtosis, mean squared energy, entropy, cross correlation ACCx-ACCy, mutual information ACCx-ACCy, cross-entropy ACCx-ACCy, extent in randomness in body motion, instantaneous changes in energy, autoregression coefficient at lag 1, zero-crossing rate, dominant freq, radial distance, polar and azimuth angle	RF, Random classifier, conditional random classifier	98.5% sens., 97.6% spec., 98.0% acc. for RF; 50.0% sens., 50.2% spec., 50.1% acc. for Random Classifier; 67.7% sens., 32.6% spec., 49.9% acc. for Conditional Random Classifier
18	Force sensors beneath each foot (100 Hz)	Normal level ground walking, 5 min	15 PD, 16 HC (Physionet DB)	Stride time, swing time, stance time, double support time	Least squares SVM, HMM (leave-one-out cross validation)	Acc.: 90.32% both for SVM and HMM
33	ACC, GYRO on foot	Walking task along a predetermined path along a hallway	21 gait disturbance PD (PDgd), 24 no gait disturbance PD, 24 HC	67 features including: ROM, max angles of dorsiflexion and plantar flexion, SD of plantar flexion, roll, pitch and yaw angles, cadence	PCA; SVM (RBF kernel, tenfold cross validation)	Two-class: 93.9% sens., 95.8% spec., 4.2% false positive rate, 97.7% precision. Multi-class (class recall/class precision): PDgd 52.4/84.6%, PD 66.7/64.0%, HC 91.7/71%
34	Force platforms under the feet (100 Hz)	8 m walkway, preferred speed, 10 times	23 PD, 26 HC	Stride length, cadence, stance time, swing time, double support time, step length, step time, stride time, step width	RF, SVM, kernel Fisher discriminant (KFD), kNN, NB	Acc.: 0.93 RF, 0.89 SVM, 0.90 KFD 0.84 NB, 0.84 kNN. AUC: 0.96 RF, 0.96 SVM, 0.95 KFD, 0.89 NB, 0.91 kNN
2	Force sensors beneath each foot (100 Hz)	Normal level ground walking, 2 min	29 PD, 18 HC (Physionet DB)	CV swing time, CV stride time, mean and SD of center of pressure for x and y-coordinates, mean and SD of peak force at heel strike, mean SD of peak force at toe strike, mean kurtosis, mean skewness, mean peak power of GRF signal	PCA, SVM (different kernels), kNN, RF, DT	Acc.: 89.4% RF, 87.2% DT, 85.1% kNN, 91.6% SVM linear, 91.5% SVM gaussian, 89.4% SVM quadratic, 95.7% SVM cubic, 93.6% SVM cubic + PCA. AUC: 0.89 RF, 0.88 DT, 0.85 kNN, 0.94 SVM linear, 0.97 SVM gaussian, 0.95 SVM quadratic, 0.98 SVM cubic, 0.97 SVM cubic + PCA

TABLE 1. continued.

Ref	Technology	Experimental design	Subjects	Features extracted	Classifiers	Classifier performance (PD/HC)
25	ACC on lower back (100 Hz)	iTUG test 7 m	20 PD, 20 HC	Time of: TUG, sit-to-walk, gait, turning, walk-to-sit. RMS of: sit-to-walk, gait, turning, walk-to-sit. Normalized jerk score of the ACC of: sit-to-walk, gait, turning, walk-to-sit. Step time, SD of step time, CV of step time. Phase of step, SD of phase, CV of phase. Phase coordination index (PCI) during gait. HRs of the trunk ACC	LDA, Quadratic discriminant analysis (QDA), Mahalanobis (MC) classifiers	Misclassification rate: 22.5% for LDA, 27.5% for QDA, 37.5% for MC
27	Force sensors below toe (100 Hz)	Gait	93 PD (HY 2/3), 73 HC (Physionet DB)	Step length, stride time, stance time, swing time, heel, below toe and toe forces and normalized heel, below toe and toe forces	LDA fivefold cross validation	Acc. 91.58%; ROC: 0.72 sens., 0.81 spec.; AUC 96%

female, mean age \pm SD 66.0 \pm 3.2), and 30 PD patients (25 male, 5 female, mean age \pm SD 67.9 \pm 8.8) were involved in this study. The PD patients were mild to mid moderate (mean MDS-UPDRS III \pm SD score: 14.7 \pm 8.6; mean HY \pm SD score: 1.9 \pm 0.8), with prevalence of the onset of the disease almost equally distributed (9 unilateral right, 10 unilateral left, and 11 bilateral). The three groups were age-matched. Subjects with IH were recruited through the IPMP-MS Project (2013–2016, coordinated by ASLIMS Department of Neurology, Carrara Hospital). A screening was provided for IH using the IOIT olfactory test,²³ and the relationship between IH and the development of PD was examined in a wide sample of subjects. For PD patients, the exclusion criteria consisted of impairments or diseases other than PD that could affect the performance of daily activities (e.g., orthopedic and neurological disorders). All patients were clinically assessed using the MDS-UPDRS III¹⁵ and HY.¹⁷ All measurements were performed in a clinically defined ON-state. All the subjects signed a written informed consent, and the study procedure was approved by the Ethical Committee of ASL1 (Azienda Sanitaria Locale, Massa e Carrara, Italy, n°1148/12.10.10) in accordance with the most recent Declaration of Helsinki.

Instrumentation

A wearable device based on an IMU was developed to objectively analyze the motor performance of the subjects involved in the study. The device, named SensFoot V2 (Fig. 1), is low-cost, low-power, non-invasive, small, lightweight, wireless, and easy to use. It is supplied by rechargeable LiPo batteries and enables the collection of data with 100 Hz sampling frequency. The device, which consists of an IMU integrated into the iNEMO-M1 board, is based on microelectromechanical sensors (MEMS) (three-axis gyroscope L3G4200D and six-axis geomagnetic module LSM303DLHC) and ARM-based 32-bit microcon-



FIGURE 1. SensFoot V2 wearable device.

troller STM32F103RE (STMicroelectronics, Italy). The system is integrated with a Bluetooth module (SPBT2632C2A, v3.0, STMicroelectronics) which wirelessly transmits acquired data to a remote personal computer for offline analysis. The device is fixed on the dorsum of the subject's foot with an elastic band to avoid movement between the foot and sensor.

Experimental Protocol

To assess the motor performance of the subjects, the experimental protocol was defined according to the neurologist and following the tasks described in the MDS-UPDRS III. In particular, four exercises were selected (tasks 3.7, 3.8, 3.10 of the MDS-UPDRS III), which are the only ones focused on lower limbs motion analysis: respectively, toe tapping with heel pin (TTHP), heel tapping (HEHE) and rotation (ROTA) and gait (GAIT).

First, the correct execution of each exercise was shown by the clinical staff, and every subject attended a short preliminary training to try all required movements. At the beginning of each exercise, the subject was asked to maintain a specific fixed position to allow 3 s of static acquisition, which represents a baseline for each trial. Specifically, regarding the tapping exercises (TTHP and HEHE), the subjects assumed a comfortable sitting posture, holding right angles between trunk and thigh and between thigh and shin; they stood with their arms at their sides for ROTA and GAIT. Then, the subjects were asked to perform the following exercises:

- *TTHP* the subjects tapped their toe on the floor, always keeping the heel in contact with the floor, for 10 s, performing the tapping action as quickly and widely as possible (MDS-UPDRS 3.7—toe tapping).
- *HEHE* the subjects tapped their heel on the floor, always keeping the forefoot raised from the floor, for 10 s, performing the tapping action as quickly and widely as possible (MDS-UPDRS 3.8—leg agility).
- *GAIT* the subjects started the gait with the non-sensorized foot and walked 15 m, at the preferred speed, in a linear way until reaching the finish line (MDS-UPDRS 3.10—gait).
- *ROTA* the subjects turned in clockwise/anticlockwise direction for 360°. The rotation was performed in a clockwise direction when SensFoot V2 was placed on the right foot and in an anticlockwise direction when SensFoot V2 was placed on the left foot (MDS-UPDRS 3.10—turn within gait).

Pre-processing

The inertial data acquired with SensFoot V2 were stored and processed offline *via* Matlab®R2016b (The MathWorks, Inc., Natick, MA, USA). Before computing parameters, data from the accelerometer and gyroscope were filtered with a fourth-order low-pass digital Butterworth filter with a 5 Hz cut-off frequency to eliminate high-frequency noise and tremor frequency bands,²⁰ typically identified between 3.5 and 7.5 Hz.²⁹ The choice for a cut-off frequency equal to 5-Hz represents a trade-off between the need to remove pathological tremor and to preserve significant information from the acquired signals. Indeed, a cut-off frequency less than 5 Hz could lead to delete useful information if the exercise is executed particularly fast. On the other hand, a cut-off frequency higher than 5 Hz could lead to maintain tremor noise together with the useful signal when Parkinsonian tremor occurs.

Differently, for gait analysis, the cut-off frequency was fixed to 3 Hz, since the cadence during a free walking is typically around 1–2 Hz. Custom-made algorithms were implemented to identify the characteristic times that allowed the signal segmentation for each exercise. Angular rates were integrated using the trapezoidal rule, with sub-intervals of integration equal to the inverse of the sensor-sampling rate ($\Delta t = 10$ ms), to calculate the movement amplitudes. A linear drift correction was applied step by step to avoid cumulative effects.

The parameters measured for each exercise are reported in Table 2 and detailed in Appendix Section. Each subject was consecutively examined two times, both for right foot (R_FOOT) and left foot (L_FOOT). For comparison between groups, the mean value of the repeated measures was used.

Feature Selection

Three different datasets were created and analyzed from the obtained parameters. In the first case, a dataset composed only of 30 HC versus 30 PD, named dataset 2C₆₀, was designed. In the second case, the IH were considered as healthy subjects, which resulted in a dataset composed of 60 HC (30 HC plus 30 IH) and 30 PD, named dataset 2C₉₀. Finally, the three different groups of subjects were considered separately (i.e., 30 HC, 30 IH, and 30 PD), for dataset 3C₉₀. These three datasets were built to analyze how the system was able to distinguish among the three groups and how the performance changed when the IH were considered as healthy subjects or not considered at all.

The Kolmogorov–Smirnov test was applied to reveal if the acquired data had a parametric or non-parametric distribution. Since each extracted

TABLE 2. Extracted features during experimental sessions.

Exercise	Features	Acronym
GAIT	Gait time	GT_Time
	Number of strides during 15 m walk	GT_Strd
	Gait frequency	GT_Freq
	Stride time	GT_StrdT
	Swing time	GT_SWT
	Stance time	GT_STT
	Relative stance	GT_RS
ROTA	Dorsiflexion angular excursion of the foot	GT_Exc
	Rotation time to cover 360° turn	RO_Time
	Number of strides during 360° turn	RO_Strd
	Rotation frequency	RO_Freq
	Stance time	RO_STT
TTHP	Rotation relative stance: (stance time/rotation time)%	RO_RS
	Tapping frequency	TT_Freq
	Number of taps	TT_Taps
	Toe angle	TT_Exc
	Coefficient of variation of tapping frequency	TT_CVfreq
	Coefficient of variation of toe angle	TT_CVexc
HEHE	Energy expenditure	TT_IAV
	Average power in power spectral density (PSD)	HE_Power
	Fundamental frequency	HE_Freq
	Maximum peak in PSD	HE_Peak
	Energy expenditure	HE_IAV

parameter resulted in a nonparametric distribution, the Kruskal–Wallis test was used to investigate the statistical significance for each feature in discriminating among the three groups (i.e., dataset 3C₉₀), while a Wilcoxon rank sum test was applied to identify significant differences between two groups (i.e., datasets 2C₆₀ and 2C₉₀). The significance level was set at p value < 0.05 . In addition, Spearman's correlation coefficients were separately assessed for each foot to eliminate the features that resulted in a high degree of correlation. If two or more features gave a high correlation coefficient ($\rho > 0.85$), only one of the features was maintained. This dimensionality reduction was performed to decrease the size of the feature array and to avoid over-fitting in the following classification step.

Supervised Classifiers

Three supervised learning classifiers (i.e., SVM, RF, and NB) were implemented to distinguish among the different groups of subjects involved in this study according to the three defined datasets (2C₆₀, 2C₉₀, 3C₉₀).

Support Vector Machine

SVM maximizes predictive accuracy while automatically avoiding data over-fitting since it is based on the maximum margin hyperplane and support vectors are the closest elements to the decision surface. Beyond

the success of SVM in many classification problems,^{2,6,18,21,33,34} this approach was chosen for its good generalization performance²⁶ and its capability to perform well on big datasets with many attributes.¹⁹ SVM however suffers from high computational and large memory requirements, but these issues are not relevant for the proposed work because the analysis was made offline. SVM can implement different types of kernels (e.g., linear, quadratic, gaussian or RBF, polynomial) which define the function for generating the hyperplane.⁹ In this work, a built-in function of MATLAB was used for the SVM setting a third order polynomial kernel.

Random Forest

RF is an ensemble learning algorithm composed of multiple decision trees. The assignment of a new observation vector to a class is based on a majority vote of the different decisions provided by each tree constituting the forest. It is robust to noise and outliers because it is based on a random step in the process of creating the trees and selecting a splitting feature.⁸ RF was chosen in this work due to the good results demonstrated in previous works.^{2,4,34} Moreover this method performs well on large datasets and the parametrization is quite simple.⁸ Even if RFs are not computationally intensive the model can take a lot of memory,¹⁴ but since the proposed classification is supposed to work offline, these factors are not con-

sidered in this work. To implement the RF, a package for Matlab^{®35} based on the Breiman *et al.* algorithm⁸ was used.

Naïve Bayes

NB are probabilistic learning algorithms, based on Bayes' Theorem. They calculate the probability of each category for a given sample and then output the category with the highest probability. This algorithm require little data for training and little storage space, while being quick in the training phase. Moreover, it does not have lot of parameters to be set. However, the strong assumption that characterizes the NB is the conditional independence of the features (i.e., all attributes are independent given the value of the class variable) and even if it shows good performance also when there are dependencies between attributes, the performance can decrease when there is a strong correlation between two or more features.¹ Due to the described features and its probabilistic approach, which is different from SVM and RF, the NB classifier is used in this work, in particular a built-in function of MATLAB was used to implement the NB.

These classifiers were trained and tested to evaluate their ability to correctly recognize the health status of the subjects. Three different classifications were implemented, and relative confusion matrices were calculated for 2C₆₀, 2C₉₀, and 3C₉₀.

A ten-fold cross-validation method was used, which ensured that unknown samples that were never presented to the system during the training phase were always available for testing. For each fold, the classifier was employed to train a model and assess its performance. The method was then used to calculate the average test error over all the folds.

The performance of the classifiers were evaluated in terms of sensitivity or recall (i.e., percentage of cases that are correctly identified as true, Eq. (1)), specificity (i.e., percentage of cases that are correctly identified as false, Eq. (2)), precision (i.e. percentage of cases correctly identified as true with respect to all diagnosed as true, Eq. (3)), accuracy (i.e., percentage of cases that are correctly identified over all subjects, Eq. (4)), and

F-measure (i.e., a weighted average of the specificity and sensitivity, Eq. (5)). To obtain these measurements, the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values were calculated.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$F\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

The methodological approach implemented in this work for comparative assessment of motor performance is summarised in the flow chart reported in Fig. 2.

RESULTS

The results obtained from inertial data analysis and the motor performance assessment using three different supervised classification (SVM, RF, NB) approaches on three different datasets (2C₆₀, 2C₉₀, 3C₉₀) are reported in this section.

Feature Selection

The 23 extracted features, both for right and left sides, from lower limb tasks, are reported as mean values and standard deviations for each class in a table uploaded as Supplementary Data online. The *p*-values calculated with significance tests, considering the three different datasets, are also reported, and statistical differences are marked with *.

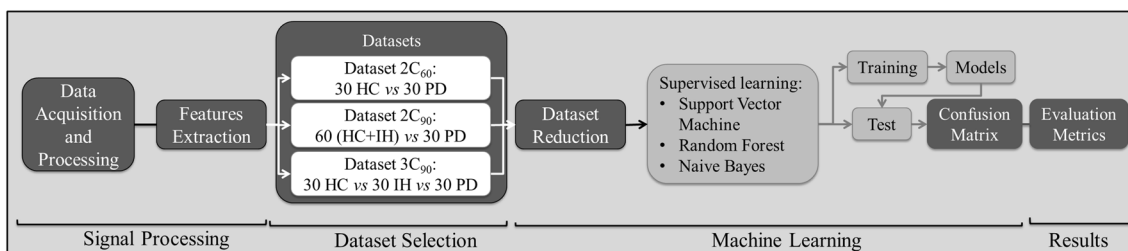


FIGURE 2. Flow chart showing the implemented methodology for motor assessment of different groups of people.

For dataset 2C₉₀, 13 parameters per limb resulted in significant differences between the two classes. In addition, in dataset 3C₉₀, the L_FOOT RO_RS could detect differences among the three groups, and in dataset 2C₆₀, an additional significant feature was L_FOOT RO_Freq, which resulted in 15 parameters per limb.

Since some parameters are highly correlated with other parameters, and redundancy of information can decrease the accuracy of group detection, according to the Spearman's correlation coefficients, the final datasets were respectively reduced to 12 features per limb for 2C₆₀, 10 features per limb for 2C₉₀, and 11 features per limb for 3C₉₀ by deleting RO_STT, TT_Freq, and HE_Power.

In summary, the features used to identify differences among all the datasets were GT_Time, GT_Strd, GT_Exc, RO_Time, RO_Strd, TT_Taps, TT_IAV, HE_Freq, HE_Peak, and HE_IAV. Additionally, RO_RS contributed in 3C₉₀ and 2C₆₀, while RO_Freq was included only for 2C₆₀.

Classification Results

The classification results, reported in Table 3, consider the right foot and left foot separately as well as both feet together.

Generally, results derived from both feet are slightly better than those obtained using a single limb, but there are also some cases in which a single foot worked better than two feet. Nonetheless, the changes obtained by using a single limb or both are not significant, and they could be caused by the specific datasets.

Concerning the use of features coming from both feet, among the three classifiers, the best results were achieved with RF and NB when considering the 2C₆₀

dataset (accuracy and *F*-measure both equal to 0.97 for RF and to 0.95 for NB), while RF gave better performances when the other two datasets were given as input to the machine learning algorithms (accuracy and *F*-measure both equal to 0.77 for 3C₉₀ and to 0.98 and 0.97 for 2C₉₀, respectively). In all cases, SVM generally appeared as the worst classifier to be used for this purpose, considering the acquired data (Fig. 3).

In Table 4, the results from the RF classifier for the 3C₉₀ dataset are reported in detail for each class and for both feet. The results show that PD patients are well identified from the other two classes, whereas the worst results are seen for IH subjects that could be misclassified, as indicated by the confusion matrix in Fig. 4.

DISCUSSION

This work aimed to investigate how the acquisitions made with inertial sensors, worn on the feet during the performance of MDS-UPDRS III scale tasks, allowed the extraction of significant motor parameters to identify differences between three groups of people (i.e., healthy subjects, people with idiopathic hyposmia, and PD patients) with good accuracy. The motion data were acquired *via* SensFoot V2 inertial sensor while subjects executed an experimental protocol composed of four motor tasks from the MDS-UPDRS III scale, which is typically used to evaluate PD patients during neurological exams. Acquired data were transmitted to a PC where they were filtered and processed through appropriate algorithms to calculate 23 features per limb that could represent subjects' motor performances. A data analysis based on significance and correlation tests was implemented to obtain

TABLE 3. Comparative results from SVM, RF, and NB classifiers in terms of Precision, Recall, Specificity, Accuracy, and *F*-measure for the three datasets (3C₉₀, 2C₉₀, 2C₆₀) calculated on right foot, left foot, and both feet.

	Precision			Recall			Specificity			Accuracy			<i>F</i> -Measure		
	SVM	RF	NB	SVM	RF	NB	SVM	RF	NB	SVM	RF	NB	SVM	RF	NB
3C ₉₀															
R_FOOT	0.576	0.735	0.766	0.544	0.722	0.756	0.772	0.861	0.878	0.544	0.722	0.756	0.560	0.729	0.761
L_FOOT	0.660	0.784	0.726	0.633	0.778	0.722	0.817	0.889	0.861	0.633	0.778	0.722	0.647	0.781	0.724
FEET	0.738	0.774	0.760	0.689	0.767	0.756	0.844	0.883	0.878	0.689	0.767	0.756	0.713	0.770	0.758
2C ₉₀															
R_FOOT	0.926	1.000	0.897	0.833	0.867	0.867	0.967	1.000	0.950	0.922	0.956	0.922	0.877	0.929	0.881
L_FOOT	0.957	0.893	0.897	0.733	0.833	0.867	0.983	0.950	0.950	0.900	0.911	0.922	0.830	0.862	0.881
FEET	1.000	1.000	0.929	0.767	0.933	0.867	1.000	1.000	0.967	0.922	0.978	0.933	0.868	0.966	0.897
2C ₆₀															
R_FOOT	0.875	0.966	0.967	0.933	0.933	0.967	0.867	0.967	0.967	0.900	0.950	0.967	0.903	0.949	0.967
L_FOOT	0.964	0.964	0.964	0.900	0.900	0.900	0.967	0.967	0.967	0.933	0.933	0.933	0.931	0.931	0.931
FEET	0.966	0.967	0.966	0.933	0.967	0.933	0.967	0.967	0.967	0.950	0.967	0.950	0.949	0.967	0.949

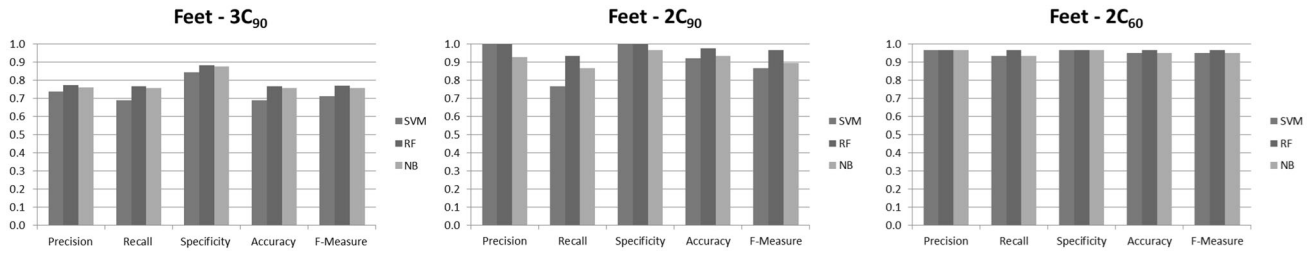


FIGURE 3. Comparative classification results for SVM, RF, and NB classifiers considering both feet.

TABLE 4. Results for 3C₉₀ dataset, applying RF classifier on both feet in terms of precision/class, recall/class, F-measure/class.

	Precision	Recall	F-measure
PD	1.000	0.933	0.966
IH	0.655	0.633	0.644
HC	0.667	0.733	0.698

		Predicted Condition		
		PD	IH	HC
True Condition	PD	28	2	0
	IH	0	19	11
	HC	0	8	22

FIGURE 4. Confusion matrix from RF classifier for 3C₉₀ dataset for FEET. Correct predictions are reported in green, while incorrect predictions are reported in red.

an adequate feature array to use for the final classification step. In this study, multiple comparisons were conducted, including the use of three different supervised machine learning approaches (i.e., SVM, RF, NB) to be applied on three datasets (i.e., 2C₆₀, 2C₉₀, 3C₉₀), for two-group or three-group classification, based on data from a single limb or both (i.e., R_FOOT, L_FOOT, FEET).

The study results reveal that the unobtrusive system used in this work enables us to distinguish between PD patients and HC with high accuracy and precision within a two-group classification. Moreover, beyond the state of the art, acquiring motor data from IH subjects as well, a further two-group classification considering IH as healthy people can be achieved with good results. Ultimately, a preliminary approach for three-group classification is proposed. In this case, IH subjects can be clustered into an additional specific class having intermediate motor performance between HC and PD patients. Thus, IH subjects can be identifiable as people at risk of developing the pathology in a prodromal phase. Further, this issue can be achieved by using non-invasive inexpensive tools (i.e., a two-step

approach that includes an olfactory test and motor assessment with inertial sensors) instead of traditional imaging techniques.

The results show that the chosen protocol (i.e., toe tapping, leg agility, gait, and rotation) allows extraction of important parameters to identify the motor characteristics for each group. The set of significant and non-correlated features, selected according to the statistical analysis, was composed of parameters computed from all four exercises; thus, each of the four tasks is important to identify the motor performance of the individuals. This finding will encourage researchers to use a comprehensive protocol for motor performance assessment which includes tapping exercises, in contrast with most works that analyzed gait only. ^{2,4,18,25,27,33,34}

Using the selected feature array, the system was able to distinguish with high accuracy and precision HC vs. PD patients using all the applied classifiers, both considering a single foot or two feet. Values obtained by analyzing the feet together were generally slightly better than those calculated for a single limb, although, for a few cases, a single foot seemed to work better than two feet. However, the differences were minimal and not significant. Further, from a clinical point of view, since the PD onset is typically asymmetrical,¹¹ the analysis of a single body side is not justifiable because it does not permit an exhaustive analysis of motor performance, which is particularly crucial in the first stage of the pathology. Thus, the analysis of both feet is recommended to identify the disease even in the first stages, when the motor symptoms could affect only one side.

Concerning the use of both feet, in this work, the best results for two-class classification were achieved by excluding the IH from the dataset (2C₆₀) with RF and NB (0.97 recall and F-measure for RF, and 0.93 recall and 0.95 F-measure for NB). These findings agree with other previous works in literature, which did not discriminate between the two limbs. RF seems to be the best classifier in our work, with 0.97 accuracy, as found also by Arora *et al.* (0.98 accuracy)⁴ and by Wahid *et al.* (0.93 accuracy).³⁴ Even if SVM appears as the less suitable classifier in this study, the values

obtained by applying this algorithm (0.93 sensitivity, 0.97 specificity, 0.95 accuracy) are comparable to or better than those found in literature by Tien *et al.* (0.94 sensitivity, 0.96 specificity),³³ Kugler and Jaremenko (0.90 sensitivity, 0.90 specificity),²¹ Wahid *et al.* (0.89 accuracy),³⁴ Alam *et al.* (0.96 accuracy),² and Khorasani and Daliri (0.90 accuracy).¹⁸ Moreover, even though NB is not frequently applied in literature in similar studies, the results obtained in this work are promising (0.95 accuracy) and significantly better than those achieved by Wahid *et al.* (0.84 accuracy),³⁴ who also employed this probabilistic approach.

Thus, the overall results of this work show good discrimination among HC and PD groups based on the analysis of lower limb motor performances acquired by wearable inertial sensors. The developed system could be applied for quantitative assessment of motion to support neurologists for objective PD diagnosis during clinical examinations.

In contrast with previous works, in this study a group of IH subjects, which are people at risk of developing PD, was included. Since they can be reasonably considered as healthy subjects, because the pathology is latent in them and will be developed only in 10–12% of IH within 5 years,^{23,28} the dataset 2C₉₀ was defined and assessed, evaluating 60 subjects as healthy vs. the 30 diagnosed PD. Considering the RF classifier and the use of both feet for the assessment, the results were high in terms of accuracy (0.98), sensitivity (0.93), and particularly for specificity (1.00).

Ultimately, in dataset 3C₉₀, where each group was assigned to a different class, the classifier performances decreased. However, this worsening was primarily related to a difficulty in discriminating between HC and IH, as shown in Fig. 4, where some IH can be confused with HC, and some healthy individuals can be misclassified as having idiopathic hyposmia. Importantly, the average results achieved in our work with both feet (0.77 recall, 0.77 precision) are better than those found by the only similar work in literature that analyzed three classes (0.70 recall, 0.73 precision),³³ even if the considered groups are different. Tien *et al.*, in fact, considered PD with and without gait disturbances and HC. In our study, PD patients are well identified (0.93 class recall, 1.00 class precision) also within 3C₉₀. Thus, the system is able to recognize the motor impairments related to the pathology. The main issue is the limited ability to discriminate between HC and IH, although this is clinically justifiable because an IH subject is an healthy with a deficiency in smell sense, and only a reduced part of the IH will actually develop the disease. On the other hand, for other IH subjects, hyposmia might not be related to PD, and worsening in motor performance might not be revealed. Nevertheless, in the three-group classification, IH subjects

are mainly identifiable as a separate group. Thus, the system seems promising in assessing small motor capability variations as well.

However, some limitations are disclosed in this study. First, the correlation with clinical scales was not considered in this work. Subjects were divided in two or three groups without considering, for instance, the severity of Parkinson's Disease or the level of idiopathic hyposmia. Thus, it was not possible to analyze whether the PD patients that were incorrectly classified as IH subjects were the ones with a low severity of pathology. Furthermore, it was not determined whether IH individuals that were incorrectly classified as HC actually had a slight olfactory impairment. Moreover, although the number of considered subjects is comparable or higher with respect to previous works,^{2,4,6,18,21,25,33,34} it is not exhaustive for the assumption that the feature array selected after the significance and correlation tests could be the definitive one to identify differences from people's motor performance. To validate the set of extracted parameters, it appears mandatory to increase the number of subjects to be considered for motor assessment. However, despite the improvements that can be applied in future works, the results obtained from the classification system are high in terms of accuracy within the scope of this paper. Thus, the presented system could reasonably support the neurologist in objective evaluation of subjects at the beginning of the disease, as well, when the typical symptoms are not yet evident but the neurodegenerative process has already begun and the motor performance is just beginning to worsen.

The analysis implemented in this work on motor data acquired by a wearable inertial device for lower limb motion assessment allowed the selection of a feature array of parameters that can identify differences between different groups of people. The comparative classification using three different supervised approaches suggests that Random Forest could be the best classifier to differentiate among healthy controls, hyposmic subjects, and PD patients, both in two-group and three-group classification. The results obtained for a two-class analysis are high (0.97 accuracy), comparable to, or better than, other works in literature, while, to the best of our knowledge, this is the first work in which motor assessment of IH subjects, as people at risk for developing PD, is evaluated. The system can recognize IH as a separate group in a three-class classification, even if some difficulties are experienced in distinguishing between IH and HC. Furthermore, this result is clinically justifiable because only a reduced subset of IH subjects will be affected by PD in the following years. Finally, the proposed system could support the neurologist for a quantitative assessment of patients' motor performances, identifying slight

worsening in motion capabilities for objective diagnosis of PD.

In future works, some improvements can be pursued to remove the present limitations. First, the dataset could be enlarged to confirm if the feature array selected in this work is truly adequate to discriminate among the groups. Then, upper limb motor performance could be analyzed as well, because the PD onset and development involve the entire body; thus, crucial additional information could be obtained. Further, clinical information (e.g., PD stage) can be correlated to biomechanical outcomes to improve the assessment of people involved in the study. Finally, a follow-up analysis of the IH subjects could be conducted, particularly for those that were misclassified, to examine if they will develop the disease or if their olfactory impairment is unrelated to PD.

ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (<https://doi.org/10.1007/s10439-018-2104-9>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGMENTS

This work was financially supported by DAPHNE project (Regione Toscana PAR FAS 2007-2013, Bando FAS SALUTE 2014, CUP J52I16000170002).

REFERENCES

- ¹Al-Aidaros, K. M., A. Abu Bakar, and Z. Othman. Naïve Bayes variants in classification learning. *Proc.—2010 Int. Conf. Inf. Retr. Knowl. Manag. Explor. Invis. World, CAMP'10* 276–281, 2010. <https://doi.org/10.1109/infrkm.2010.5466902>.
- ²Alam, M. N., A. Garg, T. T. K. Munia, R. Fazel-Rezai, and K. Tavakolian. Vertical ground reaction force marker for Parkinson's disease. *PLoS ONE* 12:1–13, 2017.
- ³Antonini, A., R. Benti, S. De Notaris, S. Tesi, A. Zecchinelli, G. Sacilotto, N. Meucci, M. Canesi, C. Mariani, G. Pezzoli, *et al.* 123I-Ioflupane/SPECT binding to striatal dopamine transporter (DAT) uptake in patients with Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. *Neurol. Sci.* 24:149–150, 2003.
- ⁴Arora, S., V. Venkataraman, S. Donohue, K. M. Biglan, E. R. Dorsey, and M. A. Little. High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones. *Acoust. Speech Signal Process.* 2014. <https://doi.org/10.1109/icassp.2014.6854280>.
- ⁵Atkinson-Clement, C., S. Pinto, A. Eusebio, and O. Coulon. Diffusion tensor imaging in Parkinson's disease: review and meta-analysis. *NeuroImage Clin.* 16:98–110, 2017.
- ⁶Barth, J., J. Klucken, P. Kugler, T. Kammerer, R. Steidl, J. Winkler, J. Hornegger, and B. Eskofier. Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's Disease. *Eng. Med. Biol. Soc.* 2011. <https://doi.org/10.1109/IEMBS.2011.6090226>.
- ⁷Berg, D., A. E. Lang, R. B. Postuma, W. Maetzler, G. Deuschl, T. Gasser, A. Siderowf, A. H. Schapira, W. Oertel, J. A. Obeso, C. W. Olanow, W. Poewe, and M. Stern. Changing the research criteria for the diagnosis of Parkinson's disease: obstacles and opportunities. *Lancet Neurol.* 12:514–524, 2013.
- ⁸Breiman, L. Random Forest. *Mach. Learn.* 45:5–32, 2001.
- ⁹Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2:121–167, 1998.
- ¹⁰Dorsey, E. R., 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* 68(5):384–386, 2007.
- ¹¹Fahn, S. Clinical aspects of Parkinson disease. In: Parkinson's disease: molecular and therapeutic insights from model systems, edited by R. Nass, and S. Przedborski. Amsterdam: Elsevier Inc., 2008, pp. 3–48.
- ¹²Galantucci, S., F. Agosta, E. Stefanova, S. Basaia, M. P. Van Den Heuvel, T. Stojković, E. Canu, I. Stanković, V. Spica, M. Copetti, *et al.* Structural brain connectome and cognitive impairment in Parkinson disease. *Radiology* 283:515–525, 2016.
- ¹³Gelb, D. J., E. Oliver, and S. Gilman. Criteria for the diagnosis of Parkinson's Disease. *Arch. Neurol.* 56:33–39, 1999.
- ¹⁴Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. Random forests for land cover classification. *Pattern Recognit. Lett.* 27:294–300, 2006.
- ¹⁵Goetz, C. G., B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. Stern, R. Dodel, B. Dubois, R. G. Holloway, J. Jankovic, J. Kulisevsky, A. E. Lang, A. J. Lees, S. Leurgans, P. A. LeWitt, D. Nyenhuis, C. W. Olanow, O. Rascol, A. Schrag, J. A. Teresi, J. J. van Hilten, and N. LaPelle. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.* 23:2129–2170, 2008.
- ¹⁶Haugen, J., M. L. T. M. Muller, V. Kotagal, R. L. Albin, R. A. Koeppe, P. J. H. Scott, K. A. Frey, and N. I. Bohnen. Prevalence of impaired odor identification in Parkinson disease with imaging evidence of nigrostriatal denervation. *J. Neural Transm.* 123:421–424, 2016.
- ¹⁷Hoehn, M. M., and M. D. Yahr. Parkinsonism: onset, progression, and mortality. *Neurology* 17:427–442, 1967.
- ¹⁸Khorasani, A., and M. R. Daliri. HMM for Classification of Parkinson's Disease based on the raw gait data. *J. Med. Syst.* 38:147, 2014.
- ¹⁹Kim, J., B. S. Kim, and S. Savarese. Comparing image classification methods: K-nearest-neighbor and support-vector-machines. Stevens Point: World Scientific and Engineering Academy and Society (WSEAS), pp. 133–138, 2012.
- ²⁰Kim, J.-W., Y. Kwon, Y.-M. Kim, H.-Y. Chung, G.-M. Eom, J.-H. Jun, J.-W. Lee, S.-B. Koh, B. K. Park, and D.-K. Kwon. Analysis of lower limb bradykinesia in Parkinson's disease patients. *Geriatr. Gerontol. Int.* 12:257–264, 2012.

- ²¹Kugler, P., and C. Jaremenko. Automatic recognition of Parkinson's disease using surface electromyography during standardized gait tests. *Eng. Med. Biol. Soc.* 2013. <https://doi.org/10.1109/EMBC.2013.6610865>.
- ²²Maremmanni, C., F. Cavallo, C. Purcaro, G. Rossi, S. Salvadori, E. Rovini, D. Esposito, A. Pieroni, S. Ramat, P. Vanni, B. Fattori, and G. Meco. Combining olfactory test and motion analysis sensors in Parkinson's disease pre-clinical diagnosis: a pilot study. *Acta Neurol. Scand.* 137:204–211, 2018.
- ²³Maremmanni, C., G. Rossi, N. Tambasco, B. Fattori, A. Pieroni, S. Ramat, A. Napolitano, P. Vanni, P. Serra, P. Piersanti, M. Zanetti, M. Coltelli, M. Orsini, R. Marconi, C. Purcaro, A. Rossi, P. Calabresi, and G. Meco. The validity and reliability of the Italian olfactory identification test (IOIT) in healthy subjects and in Parkinson's disease patients. *Parkinsonism Relat. Disord.* 18:788–793, 2012.
- ²⁴Palma, J.-A., and H. Kaufmann. Autonomic disorders predicting Parkinson disease. *Parkinsonism Relat. Disord.* 20:S94–S98, 2014.
- ²⁵Palmerini, L., S. Mellone, G. Avanzolini, F. Valzania, and L. Chiari. Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21:664–673, 2013.
- ²⁶Patel, S., 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 Trans. Inf. Technol. Biomed.* 13:864–873, 2009.
- ²⁷Perumal, S. V., and R. Sankar. Gait and tremor assessment for patients with Parkinson's disease using wearable sensors. *ICT Express* 2:168–174, 2016.
- ²⁸Ponsen, M. M., D. Stoffers, E. C. Wolters, J. Booij, and H. W. Berendse. Olfactory testing combined with dopamine transporter imaging as a method to detect prodromal Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 81:396–399, 2010.
- ²⁹Rigas, G., A. T. Tzallas, M. G. Tsipouras, P. Bougia, E. E. Tripoliti, D. Baga, D. I. Fotiadis, S. G. Tsouli, and S. Konitsiotis. Assessment of tremor activity in the Parkinson's Disease using a set of wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* 16:478–487, 2012.
- ³⁰Rovini, E., C. Maremmanni, and F. Cavallo. How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review. *Front. Neurosci.* 11:555, 2017.
- ³¹Sarkar, S., J. Raymick, and S. Imam. Neuroprotective and therapeutic strategies against Parkinson's disease: recent perspectives. *Int. J. Mol. Sci.* 17(6):904, 2016.
- ³²Schapira, A. H., R. K. Chaudhuri, and P. Jenner. Non-motor features of Parkinson disease. *Nat. Rev. Neurosci.* 18:435–450, 2017.
- ³³Tien, I., S. D. Glaser, and M. J. Aminoff. Characterization of gait abnormalities in Parkinson's disease using a wireless inertial sensor system. *Eng. Med. Biol. Soc.* 2010. <https://doi.org/10.1109/IEMBS.2010.5627904>.
- ³⁴Wahid, F., R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland. Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE J. Biomed. Health Inform.* 19:1794–1802, 2015.
- ³⁵Yan, K. YAN-PRTools., 2016. <https://it.mathworks.com/matlabcentral/fileexchange/56723-yan-prtools>.
- ³⁶Yang, K., W.-X. Xiong, F.-T. Liu, Y.-M. Sun, S. Luo, Z.-T. Ding, J.-J. Wu, and J. Wang. Objective and quantitative assessment of motor function in Parkinson's disease—from the perspective of practical applications. *Ann. Transl. Med.* 4:90, 2016.