Detection of freezing of gait in people withParkinson’s disease using smartphones

Detection of freezing of gait in people with Parkinson's disease using smartphones

Luigi Borzí
Dept. of Control and Computing Engineering
Politecnico di Torino
Turin, Italy
luigi.borzi@polito.it

Gabriella Olmo
Dept. of Control and Computing Engineering
Politecnico di Torino
Turin, Italy

Carlo Alberto Artusi
Dept. of Neuroscience "Rita Levi Montalcini"
University of Turin
Turin, Italy

Leonardo Lopiano
Dept. of Neuroscience "Rita Levi Montalcini"
University of Turin
Turin, Italy

***Abstract*—Freezing of Gait (FOG) is one of the most troublesome motor symptoms associated with Parkinson's disease (PD), characterised by brief episodes of inability to step. It involves increased risk of falls and reduced quality of life, and correlates with motor fluctuations and progression of the disease. Hence, the knowledge of FOG event frequency, duration, daily distribution and response to drug therapy is fundamental for a reliable patient's assessment. In this study, we propose a FOG detection algorithm that takes as input inertial data from a single waist-mounted smartphone, and provides information about presence and duration of FOG episodes. Data acquisition was carried on 38 PD patients and 21 elderly subjects executing a standard 6-minute walking test. More than 3.5 hours of acceleration data have been collected. A combination of Support Vector Machine and k-Nearest Neighbour classifiers has been designed. Sensitivity of 95.4%, specificity of 98.8%, precision of 92.8% and accuracy of 98.3% in the 10-fold cross validation, and a detection rate of 84% in Leave-one-Subject-Out validation were obrained. These results, along with a good time resolution in the FOG duration identification and very efficient processing times, make the algorithm a promising tool for reliable FOG assessment during activities of daily living.**

***Index Terms*—Parkinson's disease, freezing of gait, inertial sensor, smartphone, machine learning**

I. Introduction

There is increasing evidence supporting the effectiveness of modern tele-health systems from both the social and the economic standpoint [1]. Remote monitoring can enable a constant, objective, reliable and cost-effective follow-up of patients, with huge potential applications in a progressively aging society with an increasing prevalence of neurodegenerative diseases. Parkinson's disease (PD) represents an optimal target for remote monitoring applications. The identification and continuous tracking of motor complications can be achieved using lightweight, wearable inertial sensors, allowing the clinicians to develop a complete clinical picture of each patient, in line with the modern paradigm of personalized medicine.

PD is the second most common neurodegenerative disorder, affecting about 3% of individuals over 65 years worldwide. Main motor symptoms of PD are rigidity, tremor, bradykinesia and postural instability [2]. Levodopa (L-dopa) is the gold standard therapy for controlling PD motor symptoms [3]. Nevertheless, in advanced stages, it can affect the natural evolution of the disease inducing involuntary movements as well as motor response fluctuations, i.e. alternate periods of good disease control (*on* state), with others of poor control and significant PD symptoms (*off* state) [4].

Freezing of Gait (FOG) is defined as a *brief, episodic absence or marked reduction of forward progression of the feet despite having intention to walk* [5]. FOG provokes a high risk of falls, reduces functional independence and impairs the quality of life of both patients and caregivers [6]. Several subtypes of FOG are recognized [7]–[11], depending on the main elicitation context (start of walking, turns, passing through narrow spaces, approaching destination, walking in open spaces), the manifestations (shuffling with small steps, trembling in place, complete akinesia) and the duration (cut-off points: 0-2 , 3-10,11-30, more than 30 seconds [12]). In [7], the authors performed a study on 19 PD patients to assess the main FOG subtypes. Trembling in place turned out to be the most typical FOG manifestation, both in *on* and in *off* conditions.

A thorough knowledge on FOG frequency, duration, nature, and response to drug therapy can provide the neurologists with useful information about motor fluctuations and progression of the disease, allowing them to adjust the therapy and evaluate its effectiveness. Nevertheless, the clinical assessment of FOG is challenging for several reasons. Its episodic nature makes it difficult to appreciate FOG events during the brief, pre-scheduled follow-up sessions in the medical office. The FOG occurrence is related to the time elapsed since the last L-dopa administration, is dependent on the patient's attention devoted to gait, and also on many subtle cognitive factors [13]–[15]. Moreover, even though FOG is influenced by anxiety and pain, a moderate emotional stress (e.g. a medical visit) may inhibit its manifestation. For these reasons, patients seldom

exhibit FOG during follow-up visits, even though they report having several episodes at home. Consequently, at present the FOG assessment is mainly based on anamnesis, questionnaires administered to the patients, and poorly reliable self-reports of patients themselves [16], [17]. From these considerations, it turns clear that long-term observations, possibly carried on during Activities of Daily Living (ADL), could yield a significantly improved assessment of this phenomenon. We believe that a primary condition to enable this scenario is the use of affordable, low-power and low-cost, easy-to-use instrumentation such as smartphones. This is in line with the recent guidelines defined by the Movement Disorder Society task force [18].

The aim of this study is to devise and experimentally evaluate a tool for detecting FoG events and estimating their duration, using inertial data acquired from a single smartphone positioned on lower back. In perspective, this will be performed during ADL, providing information about the number, duration and daily distribution of FOG episodes. Such tool is conceived as a sort of *electronic diary* replacing the patient's self-reports. However, in this preliminary phase, participants are called to walk back and forth in semi-supervised conditions. This allowed us to test the feasibility of our method, and to record a sufficient number of FOG events in a safe environment.

The rest of this paper is organized as follows. In Sect. II we discuss the state of the art in the field of automated FOG monitoring. In Sect.III we describe the experimental setup and the cohort of PD patients enrolled for our experiment. In Sect. IV the implemented algorithm is described. In Sect V the achieved results are presented and discussed, and in Sect. VI conclusions are drawn.

II. Background

A variety of wearable sensors have been proposed for providing objective assessment of FOG. Yet, there is little agreement concerning the type of sensors to be used, their number, location on the body, experimental protocols for data acquisition, and signal processing algorithms. As for sensors, tri-axial accelerometers have been used either alone [19] or combined with gyroscopes [20]. As for locations, shin [21] and waist have been the most employed [22] as single location. Experimental protocols include Timed Up-and-Go (TUG) task [23], walking tasks with or without FOG provocation (e.g. passing through narrow spaces, walking across crowded halls, negotiate obstacles) [11] or dual tasking (e.g. carrying a glass of water while walking) [19]. Some authors address unconstrained ADL simulated in laboratory [19] or at patient's home [24]. All studies foresee video recordings of patients during data acquisition, with videos labelled by clinicians off-line. At last, smartphone use in FOG detection was recently explored in [23], [25], achieving performance comparable to other studies employing dedicated hardware. In 2018, [26] employed 3 inertial sensors attached on shank, thigh and lower back on 10 PD patients. The experimental protocol included both random and straight walking. Signals were segmented into 4 s-long windows, and features in both time and frequency domains were extracted and input to common Machine Learning (ML) algorithms, i.e. Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF). Results related to FOG detection were reported for both 10-fold cross validation and Leave-one-Subject-Out (LOSO). The first experiments led to sensitivity, specificity and accuracy of 92.4%, 95.6% and 95.3%, respectively; the latter yielded a sensitivity of 78.4%. In the same year, [27] equipped 21 PD patients with a single Inertial Measurement Unit (IMU) mounted on waist. Subjects were asked to perform TUG and simulate ADL. Inertial signals were segmented into 3.2 s-long windows, and spectral/time features were extracted. A 1D-Convolutional Neural Network (CNN) was built, leading to sensitivity, specificity and accuracy of 91.9%, 89.5%, 89.0% respectively. In 2019, [28] used 3-axial inertial sensors for FOG detection. Sensors were placed on the knee of 21 PD and 9 control subjects, performing TUG, turning with various amplitudes and walking with dual tasking. Recurrence Quantification Analysis was employed for feature extraction, and SVM for classification. Performance was evaluated using Monte Carlo validation, and the achieved error in FOG classification was less than 5%. In [29] the authors employ inertial data from the Daphnet Freezing of Gait Data Set. Signals were acquired by accelerometers placed on ankle, knee and lower back of 10 PD patients (8 of which manifested FOG during data acquisition). Patients were asked to walk, turn and simulate ADL. Signals were segmented into 4 s-long windows and time and frequency features were extracted. Different ML models were employed, namely RF, Multilayer Perceptron (MLP) and Hidden Markov Model (HMM). A LOSO validation was performed, leading to sensitivity and specificity of 95% and 75% respectively.

From this brief analysis, we can conclude that a lot of effort is being devoted to the problem of FOG detection using inertial sensors. The reported results point out the feasibility of catching FOG episodes with reasonable accuracy. However, it must be noticed that most studies perform data acquisition from inertial sensors mounted on more than one location, and this limits the possibility for patients to manage such systems by themselves in domestic environment. Moreover, signal segmentation is often performed on 4 s-long windows, and this increases the chance to miss short FOG episodes. This is a significant limitation, since early FOG manifestations are typically very short. Finally, most studies achieve suboptimal specificity, and none reports precision values. However, we believe that a very high specificity ($\geq 98\%$), together with a good precision ($\geq 80\%$), is mandatory in order to implement a reliable and detailed diary of FOG episode number, duration and circadian distribution. In fact, an excessive false-alarm rate could impair the clinical usefulness of such a diary. This opinion is also agreed upon by expert neurologists.

III. Materials and methods

The present study has been performed at the University Hospital *Città della Salute e della Scienza*, Turin, Italy, which hosts the Regional Reference Center for Parkinson's Disease

and Movement Disorders. In addition, an elderly subject cohort was enrolled at the *Orfanelle nursing home*, Chieri, Italy. The study has been conducted in accordance with the Declaration of Helsinki and approved by the local Ethics Committee. Participants received detailed information on the study purposes and execution, and written informed consent for observational study was obtained. Demographic and clinical data were noted anonymously. Patients agreed to the video-taping of the procedure after receiving suitable explanations and being guaranteed that he/she cannot be identified and the videotapes are not made available to persons different of the authorized ones. The experiments were carried out in hospital during the periodically scheduled outpatient visits; hence, the patients' safety was guaranteed by the presence of the medical staff.

A. Dataset

A total number of 38 PD subjects were included in the study. The inclusion criteria were a clinical diagnosis of PD with motor symptoms (either with a medical history of FOG events or not), and no major comorbidities or vision/cognitive impairments preventing them from accomplishing the required tasks. Subjects needing gait assistance (e.g. walking stick, crutch) were included in the study. Given that the experiment was conducted during pre-scheduled outpatients visits, all PD participants were in *daily on* state, meaning that they had taken their usual drug dose, and a variable time had elapsed since then. We also performed data acquisition on 21 elderly subjects (defined as *controls* in the rest of this paper). In this case, the inclusion criteria were no clinically evident sign of parkinsonism, no severe vision impairment, dementia, and other significant neurological disorders. Being enrolled in a retirement home, the age of control subjects was significantly superior to that of PD patients. However, even though no stratification by age was implemented, these controls have been included in the study, as they represent a worst-case situation for our algorithms in correctly classifying the gait features of PD patients. In fact, we can expect elderly people to exhibit challenging gait patterns, in terms of gait velocity, turn amplitude, wide use of walking aids (42% versus 21% in the PD patients cohort). Hence, the inclusion of this participants is useful for the algorithm validation.

Based on anamnesis and FOG questionnaires, 16 PD patients were subject to FOG events during their ADL (labelled as *anamnestic freezers*, aF), while the rest of subjects have never manifested FOG (*anamnestic non-freezers*, aNF). However, it is worth pointing out that only 10 PD patients exhibited FOG during the protocol execution. In the rest of this paper, we label as FOG+ the PD subjects who actually experienced at least one FOG episode during data acquisition, FOG- the other participants.The characteristics of the participants are summarized in Tab. I, in terms of age, disease duration, Hoehn and Yahr[1] score. Tab. II report the main demographic data of the control population.

[1]The Hoehn and Yahr scale is a commonly used method to evaluate the disease progression and disability degree. It includes 5 integer plus 2 intermediate stages, with 5 being the worst case.

TABLE I: Characteristics of PD patients with (aF) and without (aNF) an history of FOG.

Group	# subjects (% male)	Mean age (years±SD)	Disease duration (years±SD)	H&Y (mean±SD)
aF	16 (69)	74.6±4.5	9.8±3.5	2.5±0.7
aNF	22 (60)	68.1±9.7	8.3±4.4	2.2±0.4
FOG+	10 (60)	73.7 ± 4.2	6.8 ± 4.4	2.7 ± 0.7
FOG-	28 (63)	69.3 ± 9.5	9.1 ± 5	2.5 ± 0.8

TABLE II: Characteristics of control subjects.

Group	# subjects (% male)	Mean age (years±SD)	Gait assistance
Control	21 (29)	85.6 ± 7.2	42.8%

B. Smartphone evaluation

In this study, a single Samsung S5 mini smartphone was employed. The characteristics of the embedded inertial sensors have been evaluated in terms of sample frequency, range and resolution, in order to assess their suitability for the specific data acquisition tasks (Tab. III). The reported sample frequency of 200 Hz is largely adequate, given that human activity acceleration signals lay in the 0-20 Hz band [30]. The accelerometer dynamic range of ± 2g is suitable for human movement analysis using waist-mounted inertial sensors, given that its amplitude ranges between±1 g while walking and ±2 g during running [30], [31]. Moreover, the gyroscope range of ± 2000 dps is in line with that reported on dedicated hardware (e.g. IMU) in literature [32]–[34]. As reported in Tab. III, the sensors included in the mentioned smartphone largely meet these requirements. Then, a suitable location of the smartphone on the body was investigated, and waist was identified as that ensuring maximum patient comfort, closeness to the body center of mass, and possibility to monitor main PD motor symptoms (e.g. bradykinesia, dyskinesia), also in accordance with [23], [35]. For the sake of completeness, also resolution is reported. At last, we have controlled that reported values were not limited neither by the Operative System or the application employed, by visual and subsequent computational analysis of the data exported in CSV files.

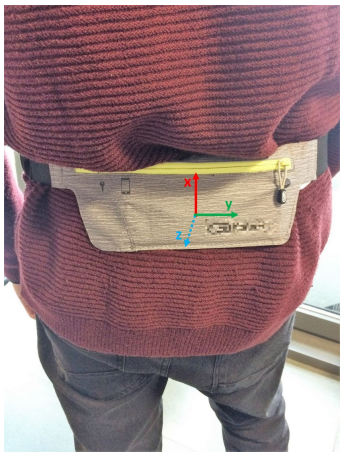
TABLE III: Samsung S5 mini sensors characteristics.

Sensor type	Range	Resolution	Sample frequency
Accelerometer	± 2 g	40 mg	200 Hz
Gyroscope	± 2000 dps	60 mdps	200 Hz

C. Experimental protocol

Following their periodic follow-up visit in the outpatients surgery, patients were asked to perform a Six-Minute Walking Test (6MWT) [36], a widely employed test thanks to its easy

Fig. 1: Smartphone belt, smartphone position and reference axes



setting up, patient tolerability and reproducibility [37]. They were equipped with the smartphone secured on their lower back (around the third lumbar vertebra) with an elastic band ensuring the adherence of the smartphone to the body (Fig. 1). They were asked to walk along a 10-meter hospital corridor at their preferred pace, possibly using their usual walking aids, then to turn 180° (alternating either direction) and return back to the starting point. This exercise encompasses several 180° turns and is repeated for 6 minutes so that to improve the chance of eliciting a FOG event. No pre-scheduled pause was planned during the test execution; nevertheless, the participants were free of quitting the test in any moment, interrupting it if tired, and possibly resuming it after a short rest. In order to ensure safety, the tests were all performed under the supervision of clinical personnel. For similar safety reasons, neither dual tasking or obstacle negotiation was included in the protocol. However, as discussed, turning is actually an effective FOG provocation task [38], [39] and *FOG while turning* is recognized as the most frequent subtype both in *on* and in *off* states [7]. The experimental conditions have forced us to perform the test on each patient after a variable time elapsed since the last L-dopa administration (*daily-on* condition, where a continuous transition between the two states can be appreciated). We recall that, even though much more frequent and of longer duration in the *off* condition, the most common FOG manifestations (e.g., trembling in place) are similar in either state. We remark that we have identified mainly this FOG subtype; however, classifying different FOG subtypes was not a goal of the present work. The activities performed during the protocol included walking, turning, standing, and FOG. The smartphone recorded and locally stored inertial data by means of SensorLog, a commercial app for Android 6.0 [40]. Once collected, data were exported in CSV format and processed offline using `MATLAB`, version 2018a for Windows 10. Moreover, the procedure was videotaped, with proper countermeasures in order to ensure the patients' privacy, to enable offline FOG evaluation by the clinicians [7].

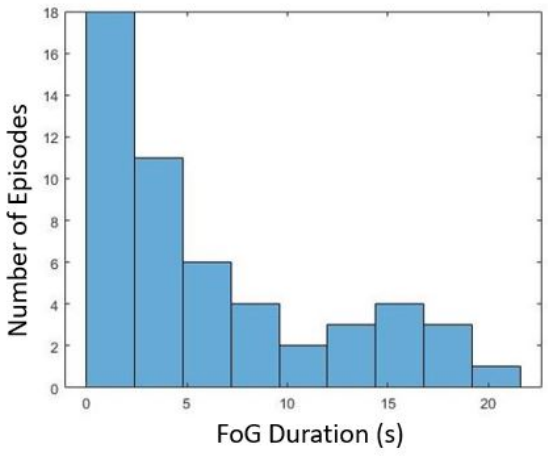
More than 3.5 hours of acceleration signals were recorded, gathering 52 FOG episodes (about 5 minutes) on 10 PD patients (6 $\pm$ 3 episode per patient). The histogram of the duration of captured FOG episodes is reported in Fig. 2. More than a half of the episodes exhibited a duration inferior to 5 s, even though some of them lasted 20 s or more. This is in line with the expected distribution of FOG event duration in *daily-on* conditions [7], [41]. In the present study, 23% and 42% of the episodes lasted less than 2 s and 3 s respectively.

IV. The proposed FOG detection algorithm

A supervised classification algorithm has been defined using different ML algoritms. Both multi-class and binary classifications were explored, in order to achieve the best accuracy. SVM, DT, k-Nearest Neighbor (KNN), Logistic Regression (LR), were employed for comparison with other reported studies. Moreover, an Artificial Neural Networks (ANN) approach was considered.

As already discussed, we deem specificity as the most important performance evaluation metric for the problem at hand. In fact, suboptimal specificity implies an excessive false positive rate (i.e. non-FOG misclassified as FOG events), and this would impair precision in unacceptable way (i.e. the probability that a detected FOG is actually a FOG event). Given that we want to use this information as a component of the patient's electronic diary, we believe that possibly missing some FOG event is more acceptable than providing the clinicians with unreliable information about FOG episode number, duration and distribution throughout the day. On the other hand, it is worth noticing that most literature studies foster sensitivity at the expenses of specificity [27], [29]. Actually, this can be motivated by the different main goal pursued by these studies, namely providing the patient with a cueing system triggered by FOG. In this case, the choice of favouring sensitivity is appropriate, as auditory or tactile cues may be tolerated even if no FOG really occurs. In any case, in order to prove clinically useful, a detection system must necessarily yield high precision besides high accuracy.

Fig. 2: Distribution of FOG episode duration.



A. Data pre-processing

The acceleration data recorded from the smartphone-embedded 3-axis accelerometer have been pre-processed prior to be subject to the feature extraction. In more detail, the following steps have been performed.

Recalibration. Due to possible incorrect positioning of the smartphone, gravity can contribute differently to the three components of the 3-D acceleration signal. Thus, for reproducibility, a recalibration process was performed using the method proposed in [42]. It consists in a correction of the accelerometer axis orientation by applying a quaternion rotation transformation to the device raw data.

Filtering. Acceleration signals were band-pass filtered (0.5-15 Hz) with a fourth-order zero-lag bandpass Butterworth filter, in order to keep the frequency components of interest and removing signal mean value and high off-band noise. Cut-off frequencies were set taking into account that the acceleration signal during locomotion lies in the band 0.5-3 Hz, whereas the signal during FOG episodes lies in the band 3-8 Hz [43], [44].

Labelling. The clinical labelling of FOG episodes is not trivial, especially when short events are involved, and an universally recognized gold standard method is debated. In this work, following common practice, we had each session of data acquisition labelled by an expert clinician, using a `Matlab` Graphic User Interface (GUI), either directly or by examining the videotapes. The manual labelling of activities other than FOG by means of inertial data is not trivial as well. We took care of manually labelling the activities included in the protocol using the same GUI employed for FOG identification. In Fig. 3, a segment of inertial data is shown. Standing is characterized by a reduced movement intensity. The vertical and anterior-posterior components of acceleration exhibit a regular pattern during walking, with main peaks representing contact of feet on the ground. Turning is characterized by an appreciable reduction in acceleration amplitude (both x- and z-axis component), together with a significant increase in the angular velocity around x-axis, followed by a reduction until zero value is reached again. In this work, the validation of our labelling has been performed using the videotapes. However, we are planning to carry out this validation using state-of-the art gait analysis systems; this is left for future developments.

Segmentation. We have considered a window duration able to encompass a characteristic signal segment (at least 1 s), while enabling the detection of short FOG events. In fact, the temporal resolution achieved by the algorithm, i.e. the capability of identifying short-duration FOG episodes, is inversely related to the window duration. A 2s-long window represents a trade off between resolution and computational efficiency, keeping in mind that a possible evolution of our tool is the real-time execution, and that the shortest FOG episodes that carry significant clinical information exhibit a duration of about 2 s [8]. Short episodes are deemed important, as they may predict a worsening of the disease and an impairment of the patient's motor conditions.

B. Feature extraction and selection

Candidate features taken into account in the present work have been identified after a thorough literature research on FOG detection based on inertial data. Moreover, additional features were defined after a direct analysis of signal patterns and spectral content differences between FOG and other activities included in the protocol. Table IV reports the entire set of features extracted from inertial signal. The features marked with an asterisk have been explicitly defined in the present study. In order to identify the most suitable feature subset for the classification of FOG episodes, all the addressed candidate features have been evaluated in each signal window for all subjects, and keeping the three signal components (vertical, medio-lateral, anterior-posterior) separated. Then, the Pearson correlation coefficient r was computed between each feature and the target (i.e. FOG or non-FOG in case of binary classification; FOG, walk, turn, stand in case of multiclass classification). Feature selection was performed keeping those features which exhibit at least moderate correlation with the target (i.e. $r > 0.4$). Among those, a further selection was performed so that only non-redundant features were kept. Table IV lists the final subset of features used for the training of different ML classifiers. A brief description of the features defined in this study is provided below.

TABLE IV: Feature set extracted in the present study. The asterisk denotes features explicitly defined for the present work.

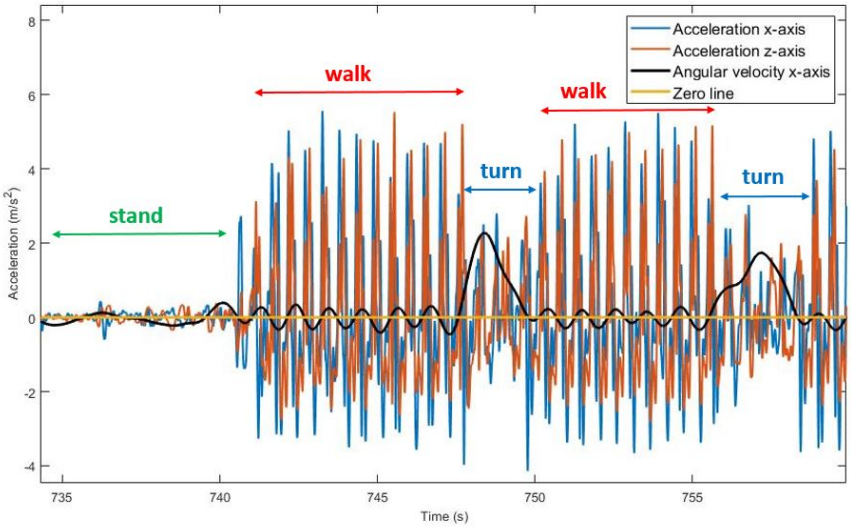
	Study	Feature
Time Domain	[45]	Mean
	[26], [45]	Standard Deviation
	[46]	Variance
	[46]	Maximum amplitude
	[26]	Root Mean Square
	[26], [47], [48]	Entropy
	[45]	Correlation between axes
	Present study	Zero crossing rate*
	Present study	Number of peaks*
Frequency Domain	[29]	Total Power
	[49]	Power in locomotor band
	[49]	Power in Freeze band
	[49]	Freeze index
	[29]	Spectral entropy
	[45], [48]	Kurtosis
	[45], [48]	Skewness
	Present study	Dominant frequency*
	Present study	Freeze ratio*

TABLE V: Selected features.

Feature	r	p-value
std x-axis	0.56	< 0.0001
std z -axis	0.55	< 0.0001
Dominant frequency x-axis	0.48	< 0.0001
zero crossing rate z-axis	0.44	< 0.0001
freeze ratio z-axis	0.43	< 0.0001

Zero crossing rate. The number of times the acceleration signal crosses the zero-line.

Fig. 3: An acceleration signal segment, encompassing activities included in the protocol.



Number of peaks. The number of peaks with an amplitude greater than the standard deviation of the signal in the window.

Dominant frequency. The frequency value corresponding to the maximum of Fast Fourier Transform (FFT) in the window considered.

Freeze ratio. The ratio between the power in the 3-10 Hz band and the total power, computed using FFT of the signal.

C. Training and test

The initial training set included all the time-windows related to the three acceleration components, computed on the measured data of each PD patient. However, given that we wanted to train several models and optimize their parameters, we had to limit the computational burden and memory requirement of this phase. To this end, we defined a *reduced training set*, encompassing all the FOG episodes and a 15-s segment for each activity of each subject (Tab. VI). This reduced set keeps the variability both in FOG manifestations and in other activity execution with respect to the complete set, while limiting the computational complexity. Feature extraction was then performed on the reduced training set.

TABLE VI: Reduced training set used for the training of different algorithms.

Activity	Signal length (min)	Total windows computed
Walk	9.5	568
Turn	8.5	508
Stand	10	598
FOG	4.4	262

The features reported in Tab. V were extracted on the reduced training set, and given as input to different ML models. Many combinations of the parameters were explored and a tuning procedure was set in order to minimize the misclassification error (i.e. optimize the classification accuracy). To this end, a Bayesian Optimization algorithm, available in Matlab, was employed. As for SVM, kernel function, kernel scale, boxconstraint and multi-class method were tuned. Number of neighbors, distance metric and distance weight were optimized for KNN algorithm. The number of hidden layers, number of neurons per hidden layers and type of transfer function were optimized for ANN.

The results achieved by the several considered algorithms are listed in Tab. VII for both binary and multi-class classification methods. One-vs-one approach was not explored for Linear Regression, as it is not feasible. It can be noticed that

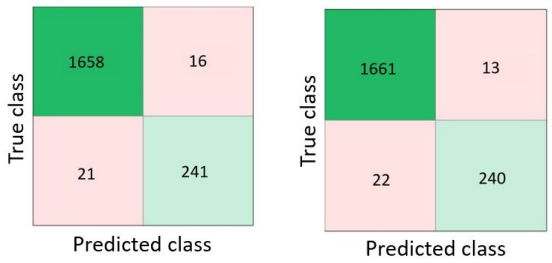
TABLE VII: 10-fold cross validation performance of different ML models

Classifier	Accuracy (%) One-vs-One	Accuracy (%) One-vs-All
Decision Tree	88.6	95.7
Linear Regression	n.r.	95.2
Linear SVM	86.5	95.4
Quadratic SVM	91.0	98.0
k-Nearest Neighbor	90.6	98.1
Artificial Neural Network	97.9	63.5

quadratic-SVM and KNN yield the best results in terms of One-vs-All accuracy, with similar performance. For the sake of completeness, the confusion matrices are shown in Fig. 4. This let us to question if a combination of such two classifiers is able to further increase the performance. To this end, we performed an additional 10-fold cross validation on the same training set, using a combination of the two classifiers. the achieved accuracy turned out to be improved, reaching a value as high as 98.4%.

Then, we performed a test on Control and FOG- subjects, employing quadratic-SVM, KNN and their combination. Each

Fig. 4: Confusion matrices relative to the best classifiers.



(a) Quadratic SVM (b) KNN

processed window was labelled as "FOG" if and only if both models classified it as part of a FOG episode. The results are reported in Tab. VIII. Please notice that this test was performed only on such groups in order to assess specificity, as such groups are known to not experiment FOG. It can be noticed that the combination yields increased specificity. Even though slight, this improvement highly affects the reliability of the system, as even a small improvement in specificity implies a substantial reduction of false positive rate when many true negatives are included in the data set. Hence, we selected the quadratic-SVM and KNN classifier combination as the best configuration for the problem at hand.

TABLE VIII: Specificity yielded by the two best classifiers and their combination.

Group	q-SVM (%)	KNN (%)	Combination (%)
Control	98.45	99.81	99.89
PD FOG-	97.8	98.2	99.66

V. Results and Discussion

In this section, we report the performance achieved by our combined classifier. We recall that the main objective of this work is to verify whether it is possible to automatically identify number and duration of FOG events, with prospective applications during ADL. Table IX reports the performance achieved by our classification method. The 10-fold cross-validation on the reduced dataset yields information on the detection effectiveness when all the inertial data variability is taken into account. Tests performed on all population samples allows one to evaluate the false positive rate, so as to appreciate the detection reliability. When dealing with control subjects, the data used for validation is not employed in the training phase. As for PD patients, a small segment of each patient's signal was included in the cross-validation phase, i.e. 15 s of each activity for each patient, out of the complete duration of the protocol. This may slightly affect specificity for FOG+ and FOG- patients. However, we believe that the inclusion of 38 PD patients with different age, motor conditions, disease progression and gait assistance needs implies a satisfactory variability degree in the data, and that further testing on new subjects is likely not to impair significantly the algorithm performance (as also shown by the test results performed on the unknown control subjects).

We have performed a Leave-One-Subject-Out (LOSO) validation. LOSO consists in training the classifiers with FOG episodes from all FOG+ patients except one, which is used for testing. Such validation is able to assess robustness, i.e. the capability of the algorithm to detect FOG episodes belonging to a patient whose inertial data was not included in the training stage (generalization capability). The resulting detection rate of 100% , with (84.1 $\pm$ 15.6)% of FOG detected in each patient, is a very promising result.

TABLE IX: Performance of the implemented detector. CV=cross-validation.

Dataset	Evaluation method	Performance	Value (%)
Reduced dataset (All PDs)	10-fold CV	Sensitivity	95.4
		Specificity	98.8
		Precision	92.8
		Accuracy	98.3
FOG+	LOSO	Detection rate	100
Control	Test	Specificity	99.89
FOG-	Test	Specificity	99.66
FOG+	Test	Specificity	95.2

Finally, we investigated the distribution of false positives along the signals. To this end, we counted the number of consecutive and non-consecutive windows erroneously classified by the system; the results are shown in Tab. X. It can be appreciated that no consecutive false events were detected either in control or FOG- groups, whereas only 2 events involved two consecutive windows in FOG+ subjects. This result is rather impressive. In fact, if we limit the algorithm to detect FOG episodes slightly longer than 2 s, specificity approaches 100%, and the system can detect FOG with the highest possible precision. The practical implications of this result deserve thorough discussion with the expert neurologists; this is left for future developments of our research.

TABLE X: Final classifier testing

Group	Signal length	False positives (consecutive)	Percentage on total windows
Control	58.3 min	4 (0)	0.11 (0)%
PD FOG-	107 min	22 (0)	0.34 (0)%
PD FOG+	33.6 min	15 (2)	0.74 (0.1)%

The slightly inferior performance achieved for FOG+ subjects can be explained in terms of the different gait pattern of these latter with respect of FOG- and elderly subjects. FOG+ patients showed a reduced movement intensity, together with more fluctuating acceleration signal, clearly appreciable in the frequency domain. Fig. 5 puts into evidence a shift towards high frequencies of all signal components (red circles). This may impair the detection performance, as some features used for classification (i.e. zero-crossing rate, freeze ratio and dominant frequency) highly depend on the signal frequency

content. Nevertheless, the system demonstrated sound performance in terms of sensitivity and precision. We are confident that a patient-specific training would gradually weaken such differences, with significant improvement of the classification performance.

As already discussed, the time resolution in determining the duration of a FOG episode depends on the window length. For a 2s-long window, a maximum error of 1 s at the beginning and 1 s at the end of each FOG episode may occur, thus leading to a time resolution of 2 s, sufficient for the purpose of the present study. To better clarify this situation, Fig. 6 reports a typical FOG episode (red square), along with the detected window (black line). The signals to the left and right of the FOG event contain turn and walk, respectively. In this case, the time resolution in determining the FOG duration is about 1s.

Due to the small feature set and to the basic classifiers implemented, the computational time (i.e. time employed in pre-processing, feature extraction and feature computation) is rather acceptable. As an example, the testing of 6-min-acceleration requires 4 s. This makes possible to devise a real-time implementation, hence to trigger a cueing system. However, this would require some modifications to the algorithm settings so as to further improve sensitivity, as discussed in Sect. IV. This application is left for possible future developments.

Finally, the proposed method is compared with similar recent studies addressing FOG with wearable sensors and ML techniques. The results are shown in Tab. XI. We remark that these comparisons must be interpreted with caution, given that differences between the patients conditions and experimental set-up may affect the results. However, it can be noticed that

Fig. 5: FFT of the acceleration signals: walking activity of a FOG- (above) and FOG+ (below) subject.

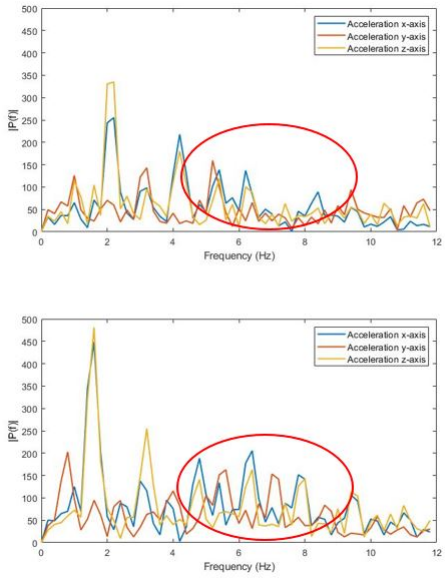
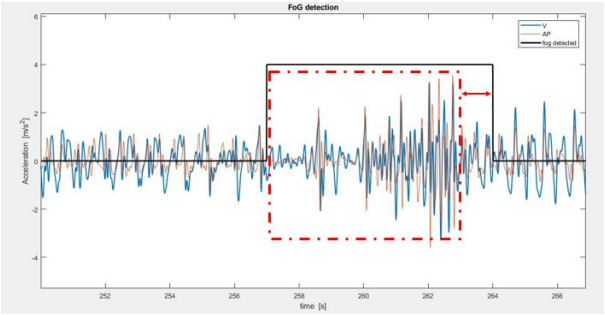


Fig. 6: Time resolution in detecting FOG.



all performance metrics of the proposed method outperform the algorithms under comparison. Moreover, the high precision achieved by our algorithm (i.e. the probability that a window classified as FOG actually contains FOG) enhances the sensitivity significance. The suboptimal specificity reported in other studies (max 95.6%, to the best of our knowledge) and the fact that precision values are seldom reported, may denote an excessive false positive rate, which would make questionable the practical usefulness of such systems in detecting FOG events. However, it must be pointed out that the algorithms under comparison were meant to trigger auditory cues, thus a different trade-off between sensitivity and specificity was addressed.

A limitation of the presented study is that few FOG episodes (namely, 54 episodes, for a total FOG duration of about 5 minutes) have been recorded. Actually, our experimental protocol did not address any kind of FOG provocation test (except for turning); all PD participants were in *daily on* state, and no selection based on previous history of FOG was applied on PD subjects. On the other hand, in [23], [25] about 10 minutes of FOG events have been registered thanks to FOG provocation (e.g. dual tasking). In [22], 20 minutes of FOG have been collected, with all patients being in *off* state. In [24], [27], up to 93 FOG minutes are recorded due to a complex experimental protocol which implements the tests under L-dopa suppression; however, only 12 minutes are related to patients in *on* state. Given the practical experimental conditions of our feasibility study, it was not possible for us to enroll patients in *off* state or under L-dopa suppression; this kind of experiments is left for future development.

Furthermore, we have been able to capture only two specific FOG subtypes: elicitation context "turning", manifestation "trembling in place" (most cases) or "shuffling with small steps" (few cases). This could have influenced the acceleration signal topology, allowing a relatively easy detection; for example, complete akinesia may be more difficult to distinguish from standing than a trembling FOG type. Nevertheless, we remark that the goal of our study was not to classify different FOG subtypes, and that the subtypes we have been able to identify are the most common ones.

In any case, we believe that the quite large number of involved participants, providing noticeable variability in the execution

TABLE XI: Comparison between the proposed method and similar recent studies.
n. r. = not reported.

Author	**[26]**	**[27]**	**[29]**	**[28]**	**Present study**
Year	2018	2018	2019	2019	2020
Dataset	10 PD	21 PD	10 PD	21 PD, 9 Control	38 PD, 21 Control
Device type	Inertial sensors	IMU	Inertial sensors	Inertial sensors	Smartphone
Device location	Shank, thigh, lower back	Waist	Shank, thigh, lower back	Knee	Lower back
Model	RF	1-D CNN	1-D CNN	SVM	SVM+KNN
Window length (s)	4	3.2	4	6	2
Sensitivity (%)	92.4	91.9	n.r.	n.r.	95.4
Specificity (%)	95.6	81.9	n.r.	n.r.	98.8
Precision (%)	n.r.	n.r.	n.r.	n.r.	92.8
Accuracy (%)	95.3	89.0	n.r.	95	98.3
Detection rate in LOPO (%)	78.4	n.r.	95	n.r.	84.3

of the test, improves the significance of our results. Finally, the results have been worked out adopting a patient-independent approach. If we implement a subject-dependent training of the ML algorithms, the performance will significantly improve. This approach is left for future developments.

VI. Conclusion and future work

Home monitoring of FOG episodes would provide valuable information to check PD progression, motor fluctuations and response to drug therapy. To this end, the algorithm designed in this study yielded promising performance in terms of robustness, low number of false positives, limited computational burden and high generalization capability. At present, the method has been applied on data measured on patients and elderly subjects performing a 6MWT. A straightforward development can be to implement a number of short data acquisition sessions at different daytimes; the acquired data could be analyzed at the end of each day, achieving information about FOG circadian distribution and duration. We plan to include more FOG episodes in the dataset, recruiting PD subjects with previous history of FOG, in *off* state while data acquisition, during the L-dopa suppression test, and possibly including other FOG provocation tasks in the experimental protocol. With such an expanded data set, the classification of different FOG subtypes may also be addressed. We plan to implement a more thorough analysis of the PD patient's gait features, in order to recognize possible pre-FOG events from subtle gait variations. In order to achieve a continuous and long-term monitoring of patient during ADL, we plan to broaden the set of activities performed during the test, so as to encompass most ADL situations. Furthermore, a patient-specific training of the classificator could be addresses, leading to further improvement in the detection performance. A real-time implementation of the algorithm can be foreseen, given the computational efficiency of the proposed algorithm. We plan to include more monitoring tasks in the same wearable device, so as to provide a thorough patient follow-up. Finally, the reduced latency from FOG episode manifestation to its detection, our tool could be possibly employed also for enabling some kind cues (e.g. auditory, tactile) in real-time, in order to solve the FOG episode itself. In order to be able both to provide a real-time feedback to patients and to perform state-of-the-art signal processing, we plan to use smarphone for a real-time soft processing and for data trasmission to a Raspberry Pi for hard processing tasks (e.g. Deep Learning applied to time-frequency images, obtained computing Convolutional Wavelet Transform of the inertial signals). We strongly believe this approach to be the most realistic and efficient in a home environment, in view of a well defined and practical remote-monitorng of patients.

VII. Acknoledgements

The Authors would like to thank the *Orfanelle Nursing Home*, Chieri, for their valuable support in selecting elderly subjects for our study.

References

[1] E. A. Krupinski and J. Bernard, "Standards and guidelines in telemedicine and telehealth," *Heathcare*, vol. 2, pp. 74–93, 2014.
[2] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," 2008.
[3] C. W. Olanow, K. Kieburtz, O. Rascol, W. Poewe, A. H. Schapira, M. M. Emre, H. Nissinen, M. Leinonen, and F. Stocchi, "Factors predictive of the development of levodopa-induced dyskinesia and wearing off in parkinson's disease," *Movement Disorders*, vol. 28, no. 8, pp. 1064–1071, 2013.
[4] C. C. Aquino and S. H. Fox, "Clinical spectrum of levodopa induced complications," *Movement Disorders*, vol. 30, no. 1, pp. 80–89, 2015.
[5] E. Heremans, A. Nieuwboer, and S. Vercruysse, "Freezing of gait in Parkinson's disease: where are we now?," 2013.
[6] A. Weiss, T. Herman, N. Giladi, and J. M. Hausdorff, "New evidence for gait abnormalities among Parkinson's disease patients who suffer from freezing of gait: insights using a body-fixed sensor worn for 3 days," *Journal of Neural Transmission*, , no. 3, pp. 403–410, 2015.
[7] J. D. Schaafsma, Y. Balash, T. Gurevich, A. L. Bartels, J. M. Hausdorff, and N. Giladi, "Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease," *European Journal of Neurology*, vol. 10, no. 4, pp. 391–398, 2003.
[8] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *Journal of Neuroscience Methods*, vol. 167, no. 2, pp. 340–348, 2008.

[9] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, "Wearable assistant for Parkinsons disease patients with the freezing of gait symptom," *IEEE Trans. on Information Tech. in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2010.

[10] A. Nieuwboer and N. Giladi, "Characterizing freezing of gait in Parkinson's disease: Models of an episodic phenomenon," 2013.

[11] S. Mazilu, "Gait, wrist and sensors: Detecting freezing of gait in Parkinson's disease from wrist movement," pp. 583–588, 2015.

[12] N. Giladi H., Shabtai, E. S. Simon, S. Biran, J. Tal, and A. D. Korczyn, "Construction of freezing of gait questionnaire for patients with Parkinsonism," *Parkinsonism and Related Disorders*, vol. 6, no. 3, pp. 165–170, 2000.

[13] C. Barthel, E. Mallia, B. Debû, B. R. Bloem, and M. U. Ferraye, "The practicalities of assessing freezing of gait," 2016.

[14] E. Heremans, A. Nieuwboer, J. Spildooren, J. Vandenbossche, N. Deroost, E. Soetens, E. Kerckhofs, and S. Vercruysse, "Cognitive aspects of freezing of gait in Parkinson's disease: A challenge for rehabilitation," in *Journal of Neural Transmission*, 2013, number 4, pp. 543–557.

[15] K. A. Ehgoetz Martens, C. G. Ellard, and Q. J. Almeida, "Does anxiety cause freezing of gait in Parkinson's disease?," *PLoS ONE*, vol. 9, no. 9, pp. e106561, 2014.

[16] N. Giladi, J. Tal, T. Azulay, O. Rascol, D. J. Brooks, E. Melamed, W. Oertel, W. H. Poewe, F. Stocchi, and E. Tolosa, "Validation of the freezing of gait questionnaire in patients with Parkinson's disease," *Movement Disorders*, vol. 24, no. 5, pp. 655–661, 2009.

[17] A. Nieuwboer, L. Rochester, T. Herman, W. Vandenberghe, G. E. Emil, T. Thomaes, and N. Giladi, "Reliability of the new freezing of gait questionnaire: Agreement between patients with Parkinson's disease and their carers," *Gait and Posture*, vol. 30, no. 4, pp. 459–463, 2009.

[18] Alberto J Espay, Jeffery M Hausdorff, Alvaro Sanchez-Ferro, Jochen Klucken, Aristide Merola, Paolo Bonato, Serene S Paul, and on behalf of the MDS Technology Task Force, "A Roadmap for Implementation of Patient-Centered Digital Outcome Measures in Parkinson's disease Obtained Using Mobile Health Technologies," *Movement Disorders Clinical Practice*, pp. 1–7, 2019.

[19] C. Ahlrichs, A. Samà, M. Lawo, J. Cabestany, D. Rodríguez-Martín, C. Pérez-López, D. Sweeney, L. R. Quinlan, G. Laighin, T. Counihan, P. Browne, L. Hadas, G. Vainstein, A. Costa, R. Annicchiarico, S. Alcaine, B. Mestre, P. Quispe, A. Bayes, and A. Rodríguez-Molinero, "Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients," *Medical and Biological Eng. and Computing*, vol. 54, no. 1, pp. 223–233, 2016.

[20] Hanbyul Kim, Hong Ji Lee, Woongwoo Lee, Sungjun Kwon, Sang Kyong Kim, Hyo Seon Jeon, Hyeyoung Park, Chae Won Shin, Won Jin Yi, B. S. Jeon, and K. S. Park, "Unconstrained detection of freezing of gait in Parkinson's disease patients using smartphone," in *37th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, 2015, vol. 2015, pp. 3751–3754.

[21] C. A. Coste, B. Sijobert, R. Pissard-Gibollet, M. Pasquier, B. Espiau, and C. Geny, "Detection of freezing of gait in Parkinson disease: Preliminary results," *Sensors*, vol. 14, no. 4, pp. 6819–6827, 2014.

[22] S. Rezvanian and T. Lockhart, "Towards real-time detection of freezing of gait using wavelet transform on wireless accelerometer data," *Sensors*, vol. 16, no. 4, pp. 475, 2016.

[23] M. Capecci, L. Pepa, F. Verdini, and M. G. Ceravolo, "A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease," *Gait and Posture*, vol. 50, pp. 28–33, 2016.

[24] D. Rodríguez-Martín, A. Samà, C. Pérez-López, A. Català, J. M. M. Arostegui, J. Cabestany, A. Bayés, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, T. J. Counihan, P. Browne, L. R. Quinlan, G. Laighin, D. Sweeney, H. Lewy, J. Azuri, G. Vainstein, R. Annicchiarico, A. Costa, and A. Rodríguez-Molinero, ," *PLoS ONE*, vol. 12, no. 2, pp. e0171764, 2017.

[25] H. B. Kim, H. J. Lee, W. W. Lee, S. K. Kim, H. S. Jeon, H. Y. Park, C. W. Shin, W. J. Yi, B. Jeon, and K. S. Park, "Validation of Freezing of Gait monitoring using smartphone," *Telemedicine and e-Health*, vol. 24, no. 12, pp. 1–6, 2018.

[26] Yi Xia, ZhiMing Yao, Yixiang Lu, Dexiang Zhang, and Nan Cheng, "A Machine Learning Approach to Detecting of Freezing of Gait in Parkinson's Disease Patients," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 4, pp. 647–654, 2018.

[27] J. Camps et. al., "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit," *Knowledge-Based Systems*, vol. 139, pp. 119–131, 2018.

[28] Taylor Chomiak, Wenbiao Xian, Zhong Pei, and Bin Hu, "A novel single-sensor-based method for the detection of gait-cycle breakdown and freezing of gait in Parkinson's disease," *Journal of Neural Transmission*, pp. 1–8, jun 2019.

[29] Rubén San-Segundo, Roque Torres-Sánchez, Jessica Hodgins, and Fernando De la Torre, "Increasing Robustness in the Detection of Freezing of Gait in Parkinson's Disease," *Electronics*, vol. 8, no. 2, pp. 119, 2019.

[30] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, "Direct measurement of human movement by accelerometry," *Medical Engineering & Physics*, vol. 30, no. 10, pp. 1364–1386, dec 2008.

[31] Vytautas Grigas, Valdas Eidukynas, and Aurelijus Domeika, "Acceleration based evaluation of the human walking and running parameters 425 . Acceleration based evaluation of the human walking and running parameters," 2014.

[32] T. Iluz, E. Gazit, T. Herman, E. Sprecher, M. Brozgol, N. Giladi, A. Mirelman, and J. M. Hausdorff, "Automated detection of missteps during community ambulation in patients with Parkinson's disease: a new approach for quantifying fall risk in the community setting," *Journal of NeuroEng. and Rehabilitation*, pp. 1–9, 2014.

[33] A. Suppa, A. Kita, G. Leodori, A. Zampogna, E. Nicolini, P. Lorenzi, R. Rao, and F. Irrera, "L-dopa and freezing of gait in Parkinson's disease: Objective assessment through a wearable wireless system," *Frontiers in Neurology*, vol. 8, no. AUG, 2017.

[34] Kristina Bettecken, Felix Bernhard, Jennifer Sartor, Markus A. Hobert, Marc Hofmann, Till Gladow, Janet M.T. Van Uem, Inga Liepelt-Scarfone, and Walter Maetzler, "No relevant association of kinematic gait parameters with Health-related Quality of Life in Parkinson's disease," *PLoS ONE*, vol. 12, no. 5, pp. 1–11, 2017.

[35] Carlos Pérez-López, Albert Samà, Daniel Rodríguez-Martín, Andreu Català, Joan Cabestany, Eva De Mingo, and Alejandro Rodríguez-Molinero, "Monitoring Motor Fluctuations in Parkinson's Disease Using a Waist-Worn Inertial Sensor," *Springer LNCS*, vol. 9094, pp. 461–474, 2015.

[36] S. H. Salzman, "The 6-min walk test," *Chest Journal*, vol. 135, no. 5, pp. 1345–1352, 2009.

[37] T. M. Steffen, "Six-minute walk test (6mwt)," 2012.

[38] M. Mancini, A. Weiss, T. Herman, and J. M. Hausdorff, "Turn around freezing: Community-living turning behavior in people with Parkinson's disease," *Frontiers in Neurology*, vol. 9, no. 1, pp. 1–9, 2018.

[39] Martina Mancini, Bastiaan R Bloem, Fay B Horak, and Simon J G Lewis, "Clinical and Methodological Challenges for Assessing Freezing of Gait : Future Perspectives Assessing the Presence and Severity of FOG in Clinical Practice," *Movement Disorders*, vol. 34, no. 6, pp. 783–790, 2019.

[40] AAVV, "Sensor log," 2014.

[41] P. Lamberti, S. Armenise, V. Castaldo, M. De Mari, G. Iliceto, P. Tronci, and L. Serlenga, "Freezing gait in Parkinson's disease," *European Neurology*, vol. 38, no. 4, pp. 297–301, 1997.

[42] Marco D. Tundo, Edward Lemaire, and Natalie Baddour, "Correcting Smartphone orientation for accelerometer-based analysis," *MeMeA 2013 - IEEE International Symposium on Medical Measurements and Applications, Proceedings*, pp. 58–62, 2013.

[43] Erika Rovini, Carlo Maremmani, and Filippo Cavallo, "How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review," *Frontiers in Neuroscience*, vol. 11, pp. 555, oct 2017.

[44] Sinziana Mazilu, Alberto Calatroni, Eran Gazit, Anat Mirelman, Jeffrey M Hausdorff, and Gerhard Tr, "Prediction of Freezing of Gait in Parkinson ' s From Physiological Wearables : An Exploratory Study," *IEEE J Biomed Health Inform*, vol. 19, no. 6, pp. 1843–1854, 2015.

[45] Daniel Rodríguez-Martín, Albert Samà, Carlos Pérez-López, Andreu Català, and Joan M.Moreno Arostegui, "Home detection of freezing of gait using Support Vector Machines through a single waist-worn triaxial accelerometer," *PLoS ONE*, vol. 12, no. 2, pp. e0171764, 2017.

[46] Kemal Polat, "Freezing of Gait (FoG) Detection using Logistic Regression in Parkinson ' s disease from Acceleration signals," *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, pp. 1–4, 2019.

[47] Bo Li, Yuqian Zhang, Liang Tang, Chao Gao, and Dongyun Gu, "Automatic Detection System for Freezing of Gait in Parkinson's Disease Based on the Clustering Algorithm," *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic*

and Automation Control Conference, IMCEC 2018, pp. 1640–1644, 2018.

[48] Terry Taewoong Um, Franz Michael Josef Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić, "Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks," 2017.

[49] S. Mazilu, U. Blanke, M. Hardegger, G. Tr, E. Gazit, M. Dorfman, and J. M. Hausdorff, "Gatiassist: A wearable assistant for gait training and rehabilitation in Parkinson' s disease," 2014.