Pattern Recognition Letters

Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments

Albert Samà[a#], Daniel Rodríguez-Martín[a#*], Carlos Pérez-López[a], Andreu Català[a], Sheila Alcaine[b], Berta Mestre[b], Anna Prats[b], M. Cruz Crespo[b], Àngels Bayés[b]

[a] *Technical Research Centre for Dependency Care and Autonomous Living (CETPD). Universitat Politècnica de Catalunya – BarcelonaTech (UPC). Rambla de l'Exposició 59-69 Vilanova i la Geltrú, 08800, Spain*
[b] *UParkinson, Centro Médico Teknon – Quirón-Salud, Passeig de la Bonanova 26, Barcelona, 08022, Spain*

ABSTRACT

Freezing of gait (FoG) is one of the most disturbing and incapacitating symptoms in Parkinson's Disease. It is defined as a sudden block in effective stepping, provoking anxiety, stress and falls. FoG is usually evaluated by means of different questionnaires; however, this method has shown to be not reliable, since it is subjective due to its dependence on patients' and caregivers' judgement. Several authors have analysed the usage of MEMS inertial systems to detect FoG with the aim of objectively evaluating it. So far, specific methods based on accelerometer's frequency response has been employed in many works; nonetheless, since they have been developed and tested in laboratory conditions, their performance is commonly poor when being used at patients' home. Therefore, this work proposes a new set of features that aims to detect FoG in real environments by using accelerometers. This set of features is compared with three previously reported approaches to detect FoG. The different feature sets are trained by means of several machine learning classifiers; furthermore, different window sizes are also evaluated. In addition, a greedy subset selection process is performed to reduce the computational load of the method and to enable a real-time implementation. Results show that the proposed method detects FoG at patients' home with 91.7% and 87.4% of sensitivity and specificity, respectively, enhancing the results of former methods between a 5% and 11% and providing a more balanced rate of true positives and true negatives.

1. Introduction

Freezing of Gait (FoG) is one of the most disabling gait disturbances among Parkinson's Disease (PD) symptoms being frequently considered as one of the cardinal symptoms in PD (Giladi & Fahn 1998). This symptom affects to almost a 50% of PD patients and up to an 80% in advanced stages (Macht et al. 2007). FoG is characterized by a sudden inability to perform an effective stepping that hampers the patient to initiate or continue locomotion during a short period of time (Moreau et al. 2008; Okuma 2006). One of its main consequences of this symptom is the loss of postural balance, which provokes falls (Bloem et al. 2004). Besides of the main physical problems caused by FoG, it can also cause complications such as social isolation, anxiety or depression (Lieberman 2006; Aarsland et al. 2007). In consequence, an accurate tracking of the incidence of FoG episodes along the day in terms of episode duration and frequency would be of great help to clinicians in order to assess the evolution of PD patients. This way, experts could provide assistance to patients by means of pharmacotherapy (e.g.: L-dopa) or physiotherapy (e.g.: strengthening exercises or cueing). Cueing are

external stimuli that have shown to reduce the duration of FoG episodes in several patients with PD (Nieuwboer et al. 2009; Arias & Cudeiro 2010; Lim et al. 2005). Among cueing systems, auditory systems are one of the most extended, since they are easy to implement (Jovanov et al. 2009) and they have shown promising effects in PD patients (Arias & Cudeiro 2010; Delval et al. 2014).

Evaluating FoG is a challenging problem due to two main factors. On the one hand, FoG is context-dependent, that is, it triggers when patients walk through narrow spaces, initiate or end gait, an obstacle impedes patients to follow their gait trajectory, or in turns (Schaafsma et al. 2003). On the other hand, patients are also conditioned by the presence of clinicians, reducing the frequency of episodes manifested compared to normal conditions (home environments) (Nieuwboer et al. 1998). This circumstance makes the correct evaluation of the symptom even more difficult. Regarding the clinical practice, FoG current evaluation is twofold: direct observation and specific questionnaires. The first method involves patients performing a series of activities in order to elicit FoG. These actions are, for example, walking through a narrow space, turns while walking and, Timed Up & Go tests (Podsiadlo & Richardson 1991; Schaafsma et al. 2003; Snijders et al. 2012;

Nonnekes et al. 2014). The drawback of these tests are given by the aforementioned stated: the patient is conditioned by performing movements in non-natural environments. Hence, clinicians might find that some patients freeze repeatedly while others do not freeze at all, not reflecting, thus, the actual severity of the symptom that patients would manifest in their Activities of Daily Living (ADL). The second method is performed by means of specific questionnaires (Giladi et al. 2000; Giladi et al. 2009; Nilsson et al. 2010). This method, although validated, might be inaccurate and biased due to the subjective answer of patients and caregivers. Additionally, self-assessment of FoG is not reliable due to the uncertain perspective of patients and caregivers to perceive the episodes. Self-assessment is also conditioned by memory loss, inattention or dementia of PD patients, leading to inaccurate recall (Papapetropoulos 2012). Evaluation at patient's home may improve the reliability; however, for economic and logistic reasons, clinicians are not able to visit patients' home regularly to perform an accurate map of the patient symptomatology.

Recent wearable technologies have enabled the possibility to assess in a more objective and cheaper way the evolution and severity of different chronic diseases (Wang et al. 2013; Chung et al. 2012; Mizuike et al. 2009; Fulk & Sazonov 2012). In the case of PD, there are several recent works that have been devoted to analyse and monitor PD symptoms (Salarian et al. 2007; Pastorino et al. 2011; Pérez-López, Samà, Rodríguez-Martín, Moreno-Aróstegui, et al. 2016; Pérez-López, Samà, Rodríguez-Martín, Català, et al. 2016). Inertial sensors based on Micro-Electro-Mechanized-Systems (MEMS), mainly triaxial accelerometers and gyroscopes, have opened the possibility to monitor motor symptoms with unobtrusive wearable devices. This way, patients can wear these devices in their ADL while being monitored without the presence of clinicians, preventing their interference from triggering FoG. Furthermore, these wearable sensors have a very low power consuming, which enable several days of uninterruptedly usage. In addition to this, the FoG monitoring output may be used to provide rhythmic cueing stimuli to PD patients (Bächlin et al. 2010) in order to reduce the frequency of the symptom. However, auditory cueing systems require real-time FoG detection with low latency time to immediately actuate after a FoG episode is detected.

FoG has been widely studied by means of inertial systems but, to the best of our knowledge, among those works focused on FoG detection, only that of Ahlrichs et al. analysed classification models based on inertial signals with patients performing activities of daily living (ADL) at their homes (Ahlrichs et al. 2016). However, Support Vector Machines (SVM) were the only classification method analysed and the output latency time was of a minute, being inadequate for cueing actuation. Other works reached a lower latency time, such as Moore et al. (6 seconds), Bächlin et al. (4 seconds) and Mazilu et al. (3 seconds); nonetheless, in these works, tests were performed under controlled conditions in laboratory environments. Consequently, when the algorithm is validated at patient's home environments, a high rate of false positive detections might be obtained due to many new activities and situations that the algorithm is evaluated with.

In this paper, a new methodology to detect FoG is presented based on a single inertial system located at the waist. This methodology aims to detect FoG in order to monitor the symptom in the daily life of patients and to apply real-time cueing strategies. The proposed method is compared to other three approaches by testing six different classifiers with signals obtained from patients' at their residence. Once the best method is selected, then, a final model is found by reducing the real-time computational load. This way, the resulting classifier's inputs (i.e. features) are reduced in order to lighten the real-time implementation. Finally, the accuracy of the presented method is computed based on a leave-one-patient-out (LOPO) strategy in order to estimate in the most unbiased way its FoG detection capacity.

The paper is structured as follows: Section 2 is devoted to describe the related work on machine learning methods used to detect FoG. Section 3 describes the proposed approach and its validation. Section 4 presents the experiments and the data collection, while Section 5 shows and discusses our results. Finally, conclusions are drawn in Section 6.

2. Related work

It is well known that FoG is a disabling symptom for PD patients but also a poor understood symptom (Cowie et al. 2004; Okuma 2006; Giladi & Nieuwboer 2008). Clinical signs of FoG are different among patients so they manifest it differently. Although it is context-dependent, there is not a rule anticipating where a patient will freeze. Thus, specific tests might lead to uncertain outcomes, since some patients will manifest FoG but others might freeze in other conditions. However, there are some common features that characterize FoG in terms of inertial signals and that can be exploited by machine learning algorithms.

Moore and Bächlin (MB) approach, which is based on analyzing the power spectra in two frequency bands, has been employed in some works to analyze its accuracy in detecting FoG in different body parts (Bächlin et al. 2010; Zabaleta et al. 2008; Niazmand et al. 2011; Zhao et al. 2012; Mazilu et al. 2012; Tripoliti et al. 2013; Mazilu et al. 2013; Ahlrichs et al. 2016). According to Bächlin et al., best results are achieved with sensors placed at legs although waist is also a good location (Bächlin et al. 2009). In addition, it has been also combined with features obtained from other sensors that could contextualize or characterize better a FoG episode. For example, Zabaleta et al. developed a FoG detection algorithm based on gyroscopes attached to lower limbs. Their work consisted of combining the frequency features along with spectral densities features. They could detect the 82.7% of the FoG episodes although they only used 2 PD patients for this test. The group of Niazmand and Zhao implemented the MB algorithm in a system with 5 accelerometers embedded within a washable jogging pant (Niazmand et al. 2011). They obtained sensitivity and specificity over 85% in six PD patients in a short and controlled test in 2011. In 2012, they obtained results over 80% with eight PD patients in a real time system within the same pants (Zhao et al. 2012).

In 2012, Mazilu et al. performed an online algorithm with 3 accelerometers achieving results over 95% on sensitivity and specificity also using the MB algorithm combined with other features (Mazilu et al. 2012). After using *correlation based feature subset selection*, they achieved a reduced set of features, which was evaluated through different classifiers, including Random Forests, k-NN (1 and 2 neighbours), and multi-layer perceptron, among others. Best results were obtained with a window of 4 seconds and with Random Forest classifier, achieving 99.54% and 99.96% in sensitivity and specificity. This method was employed with a 10-fold cross validation (CV) and evaluated by windows through the Weka software; which presents the drawbacks of evaluating classifiers with data from a patient who also participated in the traning and, furthermore, leading to an overestimation of true negatives in long periods of FoG episodes absence (Hall et al. 2009). Nonetheless, tests were performed under controlled environments and with very few activities (basically walking). This way, the algorithm is only trained for

specific situations leading to uncertain behaviours in home ADL condition. With a user independent model, results decreased to a 66.25% and 95.38% on sensitivity and specificity, respectively.

There are other works in the literature that have studied FoG detection without the MB algorithm. For example, Mazilu et al. proposed an unsupervised feature learning in order to build an input vector for a tree-based classifier (Mazilu et al. 2013). In this work they employed two methods for selecting features and they compared them to the MB algorithm. First, they employed classical time-domain features and then they reduced features in order to analyse the behaviour of the tree-based classifier. Then, a Primary Component Analysis (PCA) was employed to extract new

latency. In addition, due to the first restriction, the method must have a low computational burden and must be implementable in low power-consumption microcontrollers in order to enable a long duration of the wearable device. These considerations introduce a trade-off between FoG detection accuracy and meeting the real-time and low computational-burden restrictions. However, given the nature of the problem, accuracy takes precedence.

The proposed approach is divided into different phases, which are depicted into the left part of Figure 1. First, inertial signals are captured from a waist-worn triaxial accelerometer. This position has been chosen, first, because it has been previously used to detect FoG with relative success (Moore et al. 2013), and, second, it also
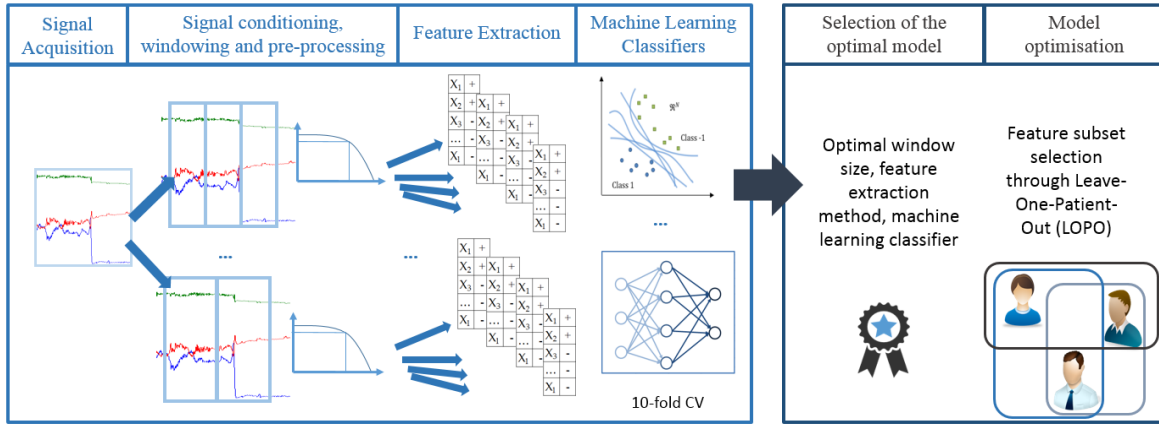

Figure 1. Proposed approach for Freezing of Gait detection. Specific feature extraction method is proposed while several window sizes and machine learning classifiers are compared. The model found by using the optimal window size and machine learning classifier is, then, reduced through a feature subset selection process.

features directly from the raw signal. Sensitivity and specificity were 77.7% and 87.56% for unsupervised features and 69.42% and 87.76% for classical time-domain features, respectively. In this case, however, they only compared *walk* vs FoG episodes, which may conduct to an ambiguous behaviour in other ADL situations. Nevertheless, results enhanced the MB approach by an 8.1% in terms of F1-score.

Tripoliti also tested different classifiers (Random forest, Decision Tree, Random tree and Naive Bayes) and different locations by computing entropy on the 3 axes of the sensors. A window-based evaluation was performed, achieving an accuracy of 96.11% with a sensitivity of 81.94% and specificity of 98.74%. with all the sensors available (6 accelerometers and 2 gyroscopes). The test was performed under controlled conditions and with a very specific test that included few activities (Tripoliti et al. 2013). Ahlrichs et al. performed an algorithm with different features and a SVM classifier (Ahlrichs et al. 2016). The specificity was validated with non-FoG patients, and the latency time was one minute, being useless for instantly actuate after the FoG episode appearance.

3. FoG detection approach

This section presents the approach followed to detect FoG based on machine learning classification models that are applied to inertial signals. In addition, this section also describes the reduction of the final FoG detection model.

Two main considerations are envisaged in the design of the FoG detection method. First, the method must enable the monitoring of the symptom during the daily life of patients. Second, it must enable the activation of external auditory cueing stimuli. As a result of the second consideration, the machine learning approach must be implementable in real-time with low

enables the monitoring of other relevant information for PD (such as gait parameters (Sayeed et al. 2015), dyskinesia (Pérez-López, Samà, Rodríguez-Martín, Moreno-Aróstegui, et al. 2016) and on/off motor states presence (Pérez-López, Samà, Rodríguez-Martín, Català, et al. 2016)). Second, signals are conditioned and specific features for FoG detection are obtained as a function of different window lengths. These features are then entered into several machine learning classifiers.

Finally, the most suitable model for FoG detection is optimised for real-time implementation. This way, given the most accurate window size, feature extraction, and machine learning classification method, a feature reduction of the corresponding model is performed. This feature reduction method is described in subsection 3.4 and is represented in the right part of Figure 1.

The proposed approach is compared to three previously reported works. These works consist of other feature extraction methods and specific classifiers. This way, previously reported features are tested with the all the different window lengths and machine learning classifiers used by the proposed approach. The other features employed are presented in Section 4.

3.1. Signal conditioning and windowing

FoG detection method employs the acceleration signals collected by a waist-worn device. This device is a wearable inertial system called 9x2, a small and light (77x37x21mm3 and 78g with battery) device (Figure 2), which is located at the left side of the waist (Rodríguez-Martín et al. 2013). This inertial measurement unit stored accelerometer data at 40Hz. This frequency is enough for the analysis of human movement in PD patients (Zhou & Hu 2008) and, furthermore, 99% of the frequency content of gait is contained below 20 Hz (Antonsson & Mann 1985). Finally,

freezing of gait symptom has been observed to provide harmonics in acceleration signals between 3 and 8Hz (Moore et al. 2008).

More formally, let $x_1, \ldots, x_M$, $y_1, \ldots, y_M$, and $z_1, \ldots, z_M$ be the measurements obtained by the accelerometer in each of its three axes at discrete times $t_1, \ldots, t_M$. As shown in Figure 2, the proposed FoG detection method conditions the signals captured by the accelerometer with a second-order low-pass Butterworth filter, which enables the removing of high-frequency noise. The implementation of the Butterworth filter is given by two vectors $\boldsymbol{a} \in \mathbb{R}^2$ and $\boldsymbol{b} \in \mathbb{R}^3$ that enable the filtering of the signals with a low computational cost, since each new filtered value is obtained through the current and the last two filtered and non-filtered samples, i.e. for X axis $x'_i = \sum_{j=1}^{3} b_i x_{i-j+1} - \sum_{j=1}^{2} a_i x'_{i-j}$.

Filtered signals are then windowed, i.e. they are splitted into windows of N consecutive samples, being each window started every $N/2$ samples. The sequence of samples contained in a certain window w, where $w = \{1, \ldots, \lfloor 2M/N \rfloor\}$ and $\lfloor \cdot \rfloor$ is the integer floor operator, starting at sample $s_w = 1 + (w - 1) \cdot N/2$ is represented by:

$$A^w = \{(x'_{s_w}, y'_{s_w}, z'_{s_w}), \ldots, (x'_{s_w+N-1}, y'_{s_w+N-1}, z'_{s_w+N-1})\} \quad (1)$$

The number of samples N to be used in a window is not limited; however, we consider a small number of possibilities due to our aim of implementing the FoG detector in real-time. Thus, window length is required to be implementable for a low-consume microcontroller and, furthermore, it should be a power of 2 in order to facilitate the discrete Fourier transform based on the FFT algorithm. In consequence, window length values explored are $N=\{32, 64, 128, 256\}$, which correspond to 0.8, 1.6, 3.2 and 6.4 seconds, respectively, since 40 samples are obtained per second.

The filtered samples contained into a window are, then, characterized according to a specific set of features, which are presented in the next subsection. In this paper, this set of features to detect FoG is proposed and it is compared to three other sets previously published in the literature.

Figure 2. The 9x2 and its orientation on waist.

3.2. Feature extraction

A total of 55 features are proposed to characterize FoG, which are listed in Table 1. These features aim to represent different aspects of PD patients activities and movements and, furthermore, different characteristics of gait. Means, standard deviations (Rodriguez-Martin et al. 2013), skewness, kurtosis and integrals (Bouten et al. 1994) of each axis are used to represent the orientation of the sensor, the quantity of movement and the third and fourth statistical moments on each axis. Correlations (Reyes Ortiz 2015) and auto-regression coefficients (Khan et al. 2010) are employed to include information on the linear relation among axis and the shape of the signal. Also, the differences between the means of the different accelerometer axes are included to consider the relative orientation of the inertial system (Rodriguez-Martin et al. 2013) with respect to previous windows, which enables the identification of postural changes. Skewness and kurtosis are also obtained from the magnitude signal.

In addition, frequency features (Najafi et al. 2003; Rodriguez-Martin et al. 2013; Pérez-López, Samà, Rodríguez-Martín, Moreno-Aróstegui, et al. 2016; Samà et al. 2012) are used to include information related to determine different postures or activities which cannot be a FoG episode, such as walking, sit, stand, lying, and PT. These features are obtained by means of the FFT, which is applied to each window A^w. The result is composed of a set of $3 \cdot N$ complex values $(X_1, Y_1, Z_1), \ldots, (X_N, Y_N, Z_N)$ representing the amplitude and phase of each complex exponential in which the signal can be decomposed according to the following equation:

$$X_h^w = \sum_{n=1}^{N} x'_{s_w+n-1} e^{-\frac{i2\pi hn}{N}} \quad (2)$$

which corresponds to the window w of the X axis, although it is also obtained for axes Y and Z, and where $h = 1, \ldots, N$.

From the values obtained by the FFT algorithm in Equation (2), several features are obtained related to the spectral content of the windowed signal. Firstly, specific frequency bands are identified. Given a range of frequencies $[f_1, f_2]$ with $0 < f_1 \leq f_2 \leq 20$, the set H_{f_1,f_2}^W is formed based on the absolute values of the complex harmonics within this frequency range:

$$H_{f_1,f_2}^W = \left\{ \begin{array}{l} \left|X^W_{\lfloor f_1 * \frac{N}{40} \rfloor}\right| + \left|Y^W_{\lfloor f_1 * \frac{N}{40} \rfloor}\right| + \left|Z^W_{\lfloor f_1 * \frac{N}{40} \rfloor}\right|, \\[6pt] \left|X^W_{\lfloor f_1 * \frac{N}{40} \rfloor+1}\right| + \left|Y^W_{\lfloor f_1 * \frac{N}{40} \rfloor+1}\right| + \left|Z^W_{\lfloor f_1 * \frac{N}{40} \rfloor+1}\right|, \\[6pt] \ldots \\[6pt] \left|X^W_{\lfloor f_2 * \frac{N}{40} \rfloor}\right| + \left|Y^W_{\lfloor f_2 * \frac{N}{40} \rfloor}\right| + \left|Z^W_{\lfloor f_2 * \frac{N}{40} \rfloor}\right| \end{array} \right\} \quad (3)$$

Secondly, frequency features are employed, since the energy spectra within specific frequency ranges correspond to determined activities. The first frequency range or frequency band used is $H_{0.04,0.68}^w$ and corresponds to the posture transition band (Najafi et al. 2003). A second band employed is $H_{3,8}^w$, which has been shown to be related with FoG (Moore et al. 2008). Finally, $H_{0.68,3}^w$ is proposed since it corresponds to walking frequency content (Rodriguez-Martin et al. 2013). From these three bands, standard deviation, skewness and kurtosis are obtained (Mazilu et al. 2013). The remaining frequency features used are the maximum, the second maximum harmonic and their distance, which are considered relevant since they establish the main dominant frequency and how the frequency content is distributed.

Finally, the Principal Component Analysis (PCA) has also been employed, as described below. Let $\boldsymbol{r}_{0.04,8}^w$ be the row vector-form of the set $H_{0.04,8}^w$. PCA has been used to reduce the dimensionality of its harmonics in any windowed signal. More specifically, PCA has been applied to a set of $\boldsymbol{r}_{0.04,8}^w$ vectors organised in a matrix $M = \left[\boldsymbol{r}_{0.04,8}^1{}^T \; \boldsymbol{r}_{0.04,8}^2{}^T, \ldots, \boldsymbol{r}_{0.04,8}^L{}^T \right]^T$. This L windows are a randomly-selected subset of the training data composed by a 10% of the data from each patient, thus containing both FoG episodes and other activities. PCA is applied to M in order to obtain the orthonormal change of basis matrix V that brings each pattern $\boldsymbol{r}_{0.04,8}^w$ from the original space to the latent one. PCA is applied to matrix $\boldsymbol{M}$ by first centering the data, then obtaining the covariance matrix, and then applying a Singular Value Decomposition (SVD) so that $\boldsymbol{U\Sigma V}$ decomposition is obtained. In the SVD process, it was identified that only the first 3 latent variables contributed with data (i.e. their singular values

were significantly higher than the rest). In consequence, only the first three components of the latent representation of $r_{0.04,8}^{w}$, i.e. the first three values of $V \cdot r_{0.04,8}^{w}$, are employed as features (feature group 11, last row in Table 1). Note that the PCA process is only applied once in the training process and a real-time implementation of the feature-extraction process would only require a partial implementation of $V \cdot r_{0.04,8}^{w}$.

Table 1. List of features employed in the proposed approach

Group: No. features	Feature group	Description
g_1: 5	Means	Mean of axis X, Y and Z Mean of the difference between X and Z axis and Y and X axis
g_2:3	Difference among mean values	Difference in the mean value of axis X between current window and predecessor. Similarly, difference in axes Y–Z and Z–X
g_3:3	Standard deviations	Standard deviation in axes X, Y and Z (time domain)
g_4:3	Correlations	Correlation between X and Y axis, Y–Z, and X–Z
g_5:5	Frequency standard deviation	Standard deviation of the absolute value of the harmonics in the following frequency bands: 0.04-0.68 Hz, 0.68-3 Hz, 3-8 Hz, 8-20 Hz and 0.1-8 Hz
g_6:4	Highest harmonics and centre of mass	Maximum and 2nd maximum harmonic amplitude, their frequency distance, and the frequency centre of mass
g_7:7	Skewness	Skewness of I) X, Y and Z axis (separately); II) of the magnitude signal, III) of the absolute value of the harmonics in the frequency bands 0.04-0.68, 0.68-3 and 3-8 Hz
g_8:7	Kurtosis	Kurtosis of I) X, Y and Z axis (separately); II) magnitude signal, III) of the absolute value of the harmonics in the frequency bands 0.04-0.68, 0.68-3 and 3-8 Hz
g_9:3	Integrals	Discrete summation of all values from X, Y and Z axis
g_{10}:12	Auto-regression coefficients	4 auto-regression coefficients obtained by the Bourg method from axis X, Y and Z
g_{11}.3	Principal Component Values	First three principal component values

3.3. Machine learning classifiers

Features extracted to represent each signal window, either the proposed ones or the previously reported on the literature, are used as the input of a supervised learning classifier with the aim of obtaining a robust FoG detection. Note that the label corresponding to each window, which consists on either FoG or no-FoG, is also employed by the learning classifier.

In this paper, several classifiers are tested in order to establish the most suitable. Since FoG detection is a bi-classification problem, SVM (Cristianini & Shawe-Taylor 2000) and Logistic regression (Hosmer et al. 2013) are employed due to their suitability and commonly good performance. On the other hand, Naive Bayes, k-NN, MLP (Haykin 1998) and Random Forest (Breiman 2001) have also been utilized in FoG detection works (Mazilu et al. 2012; Tripoliti et al. 2013); in consequence they are also evaluated. The classifiers that have been tested and their corresponding evaluated hyper-parameters are presented in Table 2. The supervised learning experiments have been performed through Weka software (Hall et al. 2009). Feature subset selection was implemented in Matlab, and LibSVM package was used to work with SVM (Chang & Lin 2011).

Table 2. List of classifiers and parameters tested

Classifier	Main Parameter
k-NN	K: 1, 3, 10, 20, 30, 40, 50
Random Forest	#trees: 1,3,5,10,20
Logistic Regression	-
Naive Bayes	-
Multilayer Perceptron	Learning Rate: 0.3, 0.5, 1 #Hidden Layers: 1,2,3,5,10
Support Vector Machines	Kernel: Linear, 2nd D. Poly, 3rd D. Poly, RBF C: 10^{-3}, 10^{-2}, …, 10^3 γ (RBF): 10^{-3}, 10^{-2}, …, 10^3

In order to compare the different classifiers, a stratified 10-fold CV approach has been applied to all of them. As the next subsection describes, a final LOPO process is used to evaluate the FoG detection abilities of the most suitable classifier.

3.4. Reduction of the optimal model and LOPO process

The classifier that is found to maximise FoG detection accuracy is optimised for its implementation into a low-power-consumption microcontroller. To this end, the number of features used is reduced through a greedy feature subset selection method. In addition to this, a LOPO schema is used in order to determine in a non-biased way the FoG detection capacity of the classifier.

More concretely, the best combination of window size, feature extraction and machine learning classifier is considered. As Section 5 describes, the set of features that most suitably characterizes FoG is the proposed in this paper, which comprises 55 values that are listed in Table 1.

From this initial feature set, LOPO is applied to estimate the FoG detection ability. More specifically, data from all patients except for one are used to train a classification model based on CV. The resulting model is, then, used to predict the data from the remaining patient, from which specificity and sensitivity values are obtained. This process is repeated as many times as patients are available. As a result of the initial LOPO process, the average specificity and sensitivity among patients using the initial set is obtained.

Secondly, the 55 features are divided into 11 groups of features according to the same division in which they are presented in Table 1, i.e. G={$g_1, \dots, g_{11}$}. Each one of these feature groups is individually removed; then, LOPO is applied, resulting in the average specificity and sensitivity among patients. The average geometric mean between specificity and sensitivity, i.e. $M = \sqrt{Specifity \cdot Sensitivity}$ for each after removing each group of features is then considered. At this point, M^G, $M^{G-g_1}, M^{G-g_2}, \dots, M^{G-g_{11}}$ have been obtained. For those groups satisfying $M^{G-g_i} \le M^G + \delta$ the process is stopped. Alternatively, those feature groups that improve the geometric mean above δ are explored, and the process continues iteratively by removing another group, e.g. $M^{G-g_i-g_2}, M^{G-g_i-g_3}$, etc. Values of δ tested are 0.1%, 0.25%, 0.5%, 1%, 2%, and 5%.

The final model to be implemented in a low power-consumption microcontroller for real-time FoG detection is that one that maximises the geometric mean metric through the described LOPO process.

4. Experiments

This section describes the data acquisition process including the inclusion criteria and the data test protocol. In addition, it is also presented the features set in which the proposed method is going to be compared to.

4.1. Data acquisition

Data from 15 PD patients that form part of the *Freezing in Parkinson's Disease: Improving Quality of Life with an Automatic Control System* (MASPARK) project data collection, have been employed in this work. These patients participated in a data collection protocol carried out by Centro Médico Teknon (Spain) which was approved by the local ethics committee. All patients gave informed written consent to the study. The inclusion criteria for the participants was a Hoehn & Yahr stage above 2 in OFF state, not having dementia according to DSM IV criteria, having a FoG-Q score above or equal to 6, and, finally, to have presented a total amount of FoG time during the data collection longer than one minute.

Data from the first 6 PD patients who participated in the database have been used to obtain the optimal window size, feature extraction and machine learning classifier through a 10-fold CV process, as described in Section 3. The complete dataset has been used in the LOPO process to reduce the features used by the classifiers, as described in subsection 4.3.

The mean age of the patients was 71.5±12.7 years old and the mean Hoehn & Yahr scale was 2.79±0.29. Four patients needed walking assistance in OFF and the mean FoG-Q test index was 14.9±6.23, being the lowest score 6 and the highest score 23.

Data collection was performed at patients' home, where FoG episodes use to occur with more frequency (Nieuwboer et al. 1998). The protocol test consisted of two parts of approximately 20 minutes each. Given that medication has a strong effect on the severity and frequency of FoG episodes in many patients (Giladi et al. 2000), the first test was performed early in the morning after the withdrawal of the first daily medication intake. Once the first part was completed, the medication was taken. The second part was performed when the effect of the medication was evident.

Both parts of the test comprised 4 different activities: firstly, showing the patients' home, secondly, a FoG provocation test in which the patient was told to go through a narrow space and, afterwards, turning back, over several times. The third activity consisted of going outdoors for a short walk, and, finally, the last one required the patient to perform a dual task activity, e.g. reading something while carrying an object. Within the second part of the test, a false positive protocol activity for FoG was also performed. This protocol consisted of performing activities in which their inertial frequency response may be similar to a FoG episode, for example, brushing teeth, painting, and erasing in a sheet of paper.

Acceleration signals were captured while patients executed the described protocol wearing the 9x2 Inertial Measurement Unit on the left side of the waist. All tests were video-recorded in order to create a gold-standard for labelling the activities and the FoG episodes. Within the MASPARK project, clinical experts labelled all activities, postures and the type of FoG episodes.

4.2. Comparison with previously reported feature extraction methods

As previously described, features proposed in the predecessor Section were compared to those reported in other research works. This subsection describes the features used to this end.

The first tested set of previously reported features was reported by Mazilu et at (Mazilu et al. 2012), who employed 15 features as the input for different classifiers. They used the mean, standard deviation and variance of a given window from each axis X, Y and Z. They also employed the entropy of the frequency components in a window and the sum of the frequency components, i.e. $\left|X_{0,20}^w\right|$, $\left|Y_{0,20}^w\right|$ and $\left|Z_{0,20}^w\right|$, for each axis individually. Finally, the freezing index (FI) and power index (PI) proposed by Moore et al. (Moore et al. 2008) and Bächlin et al. (Bächlin et al. 2009), respectively, were calculated for all three axes together. The last two indices are represented by $FI = \left|P_{3,8}^w\right| / \left|P_{0.5,3}^w\right|$ and $PI = \left|P_{0.5,8}^w\right|$, respectively.

The second set of features tested in this paper are those employed by Tripoliti et al. (Tripoliti et al. 2013). After pre-processing the signals, they extracted the entropy of the accelerometer and gyroscope signals from each axis in different body parts as features to detect FoG. In this paper, however, only accelerometer features and waist location are tested in order to be comparable to our single sensor approach. This way, three features are obtained.

Finally, features presented by Moore et al. and Bächlin et al. are also tested. In the work of Bächlin et al., they optimized the freezing index and power index by maximizing the minimum of sensitivity and specificity given with a certain value for both FI and PI. Thus, after finding an optimal value, a threshold based rule was designed. In our paper, in order to compare the performance of the same classifiers for different feature sets, the threshold-based classifier is not used. Instead of this, Moore and Bächlin features are the input for different classifiers maximizing the accuracy, which enables a more generalised non-linear classification.

5. Results

In this section, the results from the 10-fold CV results using different window sizes, features and classifiers are first presented. Second, the improvement of the most suitable model, in order to reduce the computational resources needed, through feature reduction and LOPO process is reported, with the most significant results for each feature group extraction. Finally, these results are discussed.

5.1. Window size, feature extraction and classifier results

Table 3 shows a summary of the results achieved in the 10-fold CV using different classifiers and window sizes. Given that many parameters are tested and too many results have been obtained to be completely reported, the geometric mean of the sensitivity and specificity of each method over the different patients is reported in this table. Also, this mean is provided for the best parameter configuration for each classifier, which is repeated among the different window sizes and feature sets employed. The last row includes the results corresponding to the reduced feature for comparison purposes (see subsection 5.2 for the specific features used).

The best result obtained (89.60%) belongs to CETpD features with 64-samples window size and a SVM classifier. This result stands for a 91.81% on sensitivity and an 87.45% on specificity, which is considered by the authors as an appropriately balancing of both sensitivity and specificity. These results are significantly better ($p < 0.01$) according to a t-test applied to the means of the results obtained among each group of features (among the different window and for the best classifier). Mazilu et al. features obtain also a good score (83%) although results of sensitivity are 94% and specificity 73.7% are much more unbalanced. These lower values may be due to the features used by Mazilu et al. since, although they represent the signal behaviour in each axis, they do not include cross-axis information (as in correlation), which may be important in recognizing FoG since the signal variation among them may result characteristic. Finally, Tripoliti et al. and Moore & Bächlin features obtain a geometric mean of 77% (98.8% on sensitivity and 60.31% on specificity) and 77% (92.6% on sensitivity and 64.26% on specificity), respectively. These results are clearly lower than the obtained from the proposed features and have a lack of any practical usage given the low specificity values, which would result in a high rate of false positive detections.

Last row of Table 3 reports the geometric mean values corresponding to the reduced feature set that result from the subset selection process applied to the CETpD features and described in

subection 5.2. The resulting feature set has been tested with the same 6 classifiers under the same conditions than the remaining features presented in Table 3. The feature set reduction eliminates redundant sets of features leading to a less computational burden algorithm while keeping a similar performance.

Table 3. Geometric mean of sensitivity and specificity in detecting FoG for the different feature extraction methods, window sizes and machine learning classifiers. Best results are presented here among the different hyper-parameters tested. *Ran. Forest* stands for Random Forests, *ANN* for Artificial Neural Networks (multilayer perceptron) and Logistic Regr. for Logistic regression. The complete table of results is available as the extra material of this paper. In this supplementary material, the hyper-parameters corresponding to each result can be found.

	Window Size	k-NN	Ran. Forest	Logistic Regr.	Naive Bayes	ANN	SVM
CETPD features	32	70.80	78.48	58.52	79.24	79.26	89.48
	64	76.86	80.76	66.73	79.18	80.42	89.60
	128	81.98	82.25	71.04	79.42	85.16	89.47
	256	83.17	83.41	74.69	79.18	83.60	89.19
Mazilu et al. features	32	77.62	81.06	55.52	78.41	75.55	82.49
	64	78.32	81.17	57.55	80.71	77.40	83.45
	128	80.68	81.89	58.52	81.27	73.76	83.65
	256	80.52	80.70	57.99	79.51	75.28	83.58
Tripoliti features	32	4.11	7.95	0	76.68	0	77.11
	64	52.63	51.67	0	77.57	0	77.45
	128	56.02	56.15	0	77.74	0	76.91
	256	72.03	54.77	0	77.20	0	74.65
Moore & Bächlin features	32	58.35	60.26	2.90	5.44	0	77.18
	64	65.94	67.10	1.97	0	0	45.90
	128	67.78	66.85	0	0	0	46.14
	256	62.86	64.72	0	71.26	0	43.62
CETpD subset of features	*128*	*79.69*	*80.72*	*75.75*	*81.35*	*81.36*	*89.63*

On the other hand, window size seems to slightly affect to some classifiers though with a minor relevance. For example, CETpD, Mazilu and Tripoliti features' results obtained by the SVM are not altered by the window size. However, in Moore and Bächlin approach, the algorithm performance is maximized with 32 samples per window; although two exceptions are found since k-NN and random forest classifiers provide the best performance with 256 samples. These results may be due to the sensor position, given that patients wore the sensor on the waist. In a previous paper in which different sensor locations were compared through the Moore approach (Bächlin et al. 2009), waist location was found to miss short FoG episodes with respect to ankle location. Thus, given that only non-short episodes (longer than a second) are detected, the window size is not influential with regards to the accuracy.

Finally, the comparison between classifiers shows that the best one is the SVM classifier with 89.60% on the geometric mean of sensitivity and specificity compared to the 85.16% achieved with the multilayer perceptron, which is the second better classifier. In average, Random Forests are the 2nd best performance classifier, though an 83.41% is achieved as highest value. In addition, SVM is the best classifier for any feature set achieving the highest value after performing different window size tests.

Given the results presented in Table 3, feature reduction and LOPO process have been applied to the proposed feature extraction and the SVM classifier with 128 samples.

5.2. Feature reduction and LOPO process

The results of the feature subset reduction process for δ=1% is represented in Figure 3 as a directed graph. The graph's initial node corresponds to the complete set of features, and the remaining nodes represent those subsets of features that satisfied the condition $M^{G-g_{i_1}-\cdots-g_{i_k}} > M^{G-g_{i_1}-\cdots-g_{i_{k-1}}} + \delta$, i.e. those subsets for which having removed a specific group of features improves the LOPO accuracy in at least 1%. Directed edges represent the relation among nodes, i.e. the successor has a group of features less than the predecessor. Each node include the average geometric of the specificity and sensitivity among patients. The feature subset selection process provided the same group selection for $\delta < 1\%$, while $\delta > 1\%$ resulted in slightly lower geometric mean values.

Figure 3 shows that there are four groups of features that can be consistently removed, which are the means, kurtosis, integrals and autoregression coefficients. After removing them, the classifier improves by 4.32% its average geometric mean with respect to the initial node. The number of features decreases from 55 to 28, which reduces the computational load of the feature extraction process. This way, a real-time implementation of the method will benefit through, first, the reduction of the Support Vectors (SV) dimensionality (in terms of columns) and the reduction of the resources needed to compute the features. As a result, the method is more affordable for a real-time implementation, which is very sensible in low-consume microcontrollers as those that wearable devices require.

The results provided by the LOPO process (geometric mean of 85.15%, 84.49 on sensitivity and 85.83 on specificity) do not achieve the same values presented in Table 3 (89.63%) due to the differences in the evaluation process described in Section 3. First, Table 3 results were obtained through a stratified 10-fold CV, so data from all patients were likely to be in the training and validation folds. Instead, LOPO process performs the same 10-fold CV without using the data from a patient (last process in Figure 1). Then, these data are used to obtain the specificity and sensitivity. In consequence, LOPO process has a higher degree of difficulty and, as a result, presents lower performances, which in our case are about 4.3%. An additional test has been carried out to reduce the number of SV. A method devoted to this task is the *Separable Case Approximation* (SCA) (Geebelen et al. 2012). This method starts from an already trained regular SVM (i.e. with soft-margin and kernelised, as one of the resulting models from our LOPO process), to identify the training patterns that are misclassified by the model. Then, the misclassified patterns are either removed from the training dataset or their class labels are flipped, and a new SVM without soft-margin is trained with the new dataset. By doing this, a very similar hyperplane separation is obtained and the number of SV of the new SVM is significantly reduced. In our case, the original SVM from a given patient originally had 7412 SV, from a training dataset with more than 20,000 patterns. After applying SCA, the resulting SVM had 992 SV. The original SVM with the complete set of features would require to allocate 1.6 MB of memory to save it, while the reduced one requires less than a half, 0.79 MB. Given that the flash memory of the 9x2-device's microcontroller used in the experiments (STM32F415RG) is of 1 MB, the improvements done by the feature and SV reduction enable the real-time implementation of the method.

5.3. Discussion

The results achieved by the method presented in this work obtained by optimising the window size and the machine learning classifier show that it is the set that provides the highest accuracy and geometric mean, outperforming those reported by Mazilu et al., Tripoliti et al. and Moore and Bachlin works.
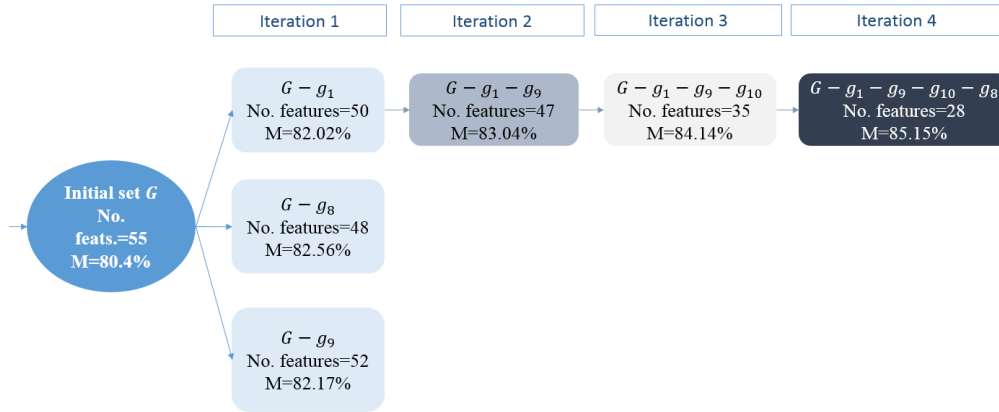

Figure 3. Feature reduction and LOPO results for $\delta = 1\%$. The graph's initial node corresponds to the complete set of features. Each successor node represents a group of features (among those presented in Table 1) removed from the predecessor node and satisfying the condition $M^{G-g_{i_1}- \dots -g_{i_k}} > M^{G-g_{i_1}- \dots -g_{i_{k-1}}} + \delta$. Each node includes the average geometric of the specificity and sensitivity among patients. The complete list of feature groups are defined in Table 1.

Table 4. Summary of results. Results reported in previous publications and those obtained in this paper are presented. Conditions employed to perform each method, including number of patients, number of sensors, protocol test and evaluation method are also described.

	Reported results reported in the previous paper (Sensitivity, Specificity)	Results obtained in this work Accuracy (sens., spec.) through 10-fold CV	No. of sensors used in the previous paper	No. of sensors used in this work	No. of patients in the previous paper	Number of patients in this work (10-fold CV)	Protocol test in the previous paper	Protocol test in this work	Evaluation system in the previous paper	Evaluation system in this work
Rodriguez et al.	-	89.6% (91.77%, 87.45%)	Waist		-		-		-	
Mazilu et al.	66.25%, 95.38%	83.65% (94.75%, 73.72%)	Waist		8		Walking and turns	Home and outdoors activities (walk, stand, sit, posture transitions, stairs, dual task, home cleaning, brushing teeth, draw, laptop affairs)	Window evaluation	Window Evaluation
Tripoliti et al.	81.94% , 98.74%	77.45% (98.21%, 61.1%)	Waist, legs, wrists and chest	Waist	5	6	Lying, walk, turns, drinking water and sit		Window evaluation	
Bächlin et al.	73.1%, 81.6%	77.18% (92.59%, 64.25%)	Ankle		8		Walk, turns, carrying light objects		Window evaluation	

Mazilu et al. presented accuracies in their work that are lower than those achieved by their same features with our dataset of signals. They reported a sensitivity and specificity over 95% by using a 10-fold CV and 3 accelerometers located on the shank, thigh and lower back. Two reasons are envisaged to explain our lower results. First, their evaluation dataset was formed by random windows including data of a patient into the training set and the evaluation set. This method leads to an overestimation of results, since the classifier learns the behaviour of the episodes performed by the patient who is evaluated. In the case of the LOPO process, Mazilu et al. report a high specificity but a very low sensitivity (66%). Second, the environment (i.e. patient's home in our case, and laboratory conditions in theirs) and the activities performed by patients, which were basically walking and turning in Mazilu's experiments, made it easier for the classifiers in Mazilu's work to detect FoG, so rather simpler features were enough to achieve such performance. On the contrary, the more challenging activities done by patients in our experiments may have provided many false positive situations. In this sense, it is noted that their method achieved a geometric mean of 77.23% through a LOPO evaluation, which is 8% lower than the reported for our approach.

Similarly, the results of Tripoliti reported in their paper (81.9% on sensitivity and 98.74 on specificity) may have been decreased by the use of a single sensor (they used up to 6), and the set of activities, which was limited (walking, turning and drinking from a glass).

To summarise, it is observed that the presented approach either presents a similar accuracy or outperforms previous works, while having the advantage of using a single sensor and being validated through a LOPO process. In addition, the method obtained is implementable in real-time, which are important advantages for the monitoring of PD and the administration of cueing.

6. Conclusions

In this work, a machine learning algorithm to detect FoG with a single inertial system in the waist based on a triaxial accelerometer

is proposed. This algorithm consists of a novel set of features used as the input of a learning classifier. The algorithm is tested with inertial data from a dataset of signals collected within patients' home, and it is evaluated by using several classifiers, different parameters for each classifier, and different window sizes in order to identify the most suitable one. Furthermore, the proposed method has been compared to other sets of features from relevant works that achieved significant results in detecting FoG. Results show a significant increase on FoG detection accuracy in the proposed approach with respect to the other implemented feature sets.

The algorithm presented in this work enables the real-time detection of FoG in ambulatory conditions for PD patients, opening the possibility to administering cueing stimuli. However, further work to validate the method is needed by including more patients. In this sense, within MASPARK project, it is expected to increase this database with the coming pilots of the project.

Acknowledgments

Conflicts of Interest

Albert Samà, Carlos Pérez-López, and Andreu Català are shareholders of Sense4Care, which is a spin-off company of Universitat Politècnica de Catalunya that may commercialize the results of this research device in a near future. These authors declare that the possible commercialization of the product is a research outcome, not being the design, the analysis, the interpretation of the results or the conclusions affected by commercial interests.

References

Aarsland, D. et al., 2007. Neuropsychiatric symptoms in patients with Parkinson's disease and dementia: frequency, profile and associated care giver stress. *Journal of neurology, neurosurgery, and psychiatry*, 78, pp.36–42.

Ahlrichs, C. et al., 2016. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. *Medical & Biological Engineering & Computing*, 54(1), pp.223–233.

Antonsson, E.K. & Mann, R.W., 1985. The frequency content of gait. *Journal of biomechanics*, 18(1), pp.39–47.

Arias, P. & Cudeiro, J., 2010. Effect of rhythmic auditory stimulation on gait in Parkinsonian patients with and without freezing of gait. *PloS one*, 5(3), p.8.

Bächlin, M. et al., 2010. A wearable system to assist walking of Parkinson's disease patients. *Methods of information in medicine*, 49(1), pp.88–95.

Bächlin, M. et al., 2009. Online Detection of Freezing of Gait in Parkinson's Disease Patients: A Performance Characterization. *Proceedings of the 4th International ICST Conference on Body Area Networks*.

Bloem, B.R. et al., 2004. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. *Movement disorders : official journal of the Movement Disorder Society*, 19(8), pp.871–884.

Bouten, C. V et al., 1994. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine and science in sports and exercise*, 26(12), pp.1516–1523.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp.5–32. Available at: http://link.springer.com/10.1023/A:1010933404324.

Chang, C.-C. & Lin, C.-J., 2011. LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp.1–27.

Chung, P.-C. et al., 2012. Gait analysis for patients with Alzheimer'S disease using a triaxial accelerometer. In *IEEE International Symposium on Circuits and Systems*. IEEE, pp. 1323–1326.

Cowie, D. et al., 2004. Freezing of gait in Parkinson's disease. *Journal of Neurology*, 251 Suppl(4), pp.28–31.

Cristianini, N. & Shawe-Taylor, J., 2000. *An introduction to support Vector Machines: and other kernel-based learning methods* C. U. Press, ed.,

Delval, A. et al., 2014. Auditory cueing of gait initiation in Parkinson's disease patients with freezing of gait. *Clinical Neurophysiology*, 125(8), pp.1675–1681.

Fulk, G.D. & Sazonov, E., 2012. Using Sensors to Measure Activity in People with Stroke. *Topics in Stroke Rehabilitation*, 18(6), pp.746–757.

Geebelen, D., Suykens, J.A.K. & Vandewalle, J., 2012. Reducing the number of support vectors of SVM classifiers using the smoothed separable case approximation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(4), pp.682–688.

Giladi, N. et al., 2000. Construction of freezing of gait questionnaire for patients with Parkinsonism. *Parkinsonism & related disorders*, 6(3), pp.165–170.

Giladi, N. et al., 2009. Validation of the freezing of gait questionnaire in patients with Parkinson's disease. *Movement Disorders*, 24(5), pp.655–661.

Giladi, N. & Fahn, S., 1998. Freezing Phenomenon, the Fifth Cardinal Sign of Parkinsonism. In *Movement Disorders*. pp. 329–335.

Giladi, N. & Nieuwboer, A., 2008. Understanding and treating freezing of gait in parkinsonism, proposed working definition, and setting the stage. *Movement Disorders*, 23(2), pp.423–425.

Hall, M. et al., 2009. The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, 11(1), p.10.

Haykin, S., 1998. *Neural Networks* 2nd ed., Prentice Hall.

Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X., 2013. *Applied Logistic Regression* 3rd ed., Wiley.

Jovanov, E. et al., 2009. deFOG--A real time system for detection and unfreezing of gait of Parkinson's patients. In *31st Annual International Conference of the IEEE EMBS*. pp. 5151–5154.

Khan, A.M. et al., 2010. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 14(5), pp.1166–72.

Lieberman, A., 2006. Are Freezing of Gait (FOG) and panic related? *Journal of the Neurological Sciences*, 248(1–2), pp.219–222.

Lim, I. et al., 2005. Effects of external rhythmical cueing on gait in patients with Parkinson's disease: a systematic review. *Clinical rehabilitation*, 19(7), pp.695–713.

Macht, M. et al., 2007. Predictors of freezing in Parkinson's disease: a survey of 6,620 patients. *Movement disorders : official journal of the Movement Disorder Society*, 22(7), pp.953–956.

Mazilu, S. et al., 2013. Feature Learning for Detection and Prediction of Freezing of Gait in Parkinson's Disease. In *Proceedings of the 9th international conference on Machine Learning and Data Mining in Pattern Recognition*. pp. 144–158.

Mazilu, S. et al., 2012. Online Detection of Freezing of Gait with Smartphones and Machine Learning Techniques. *Proceedings of the 6th International ICST Conference on Pervasive Computing Technologies for Healthcare*, (3), pp.123–130.

Mizuike, C., Ohgi, S. & Morita, S., 2009. Analysis of stroke patient walking dynamics using a tri-axial accelerometer. *Gait & posture*, 30(1), pp.60–64.

Moore, S.T. et al., 2013. Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry. *Journal of neuroengineering and rehabilitation*, 10(1), p.19.

Moore, S.T., MacDougall, H.G. & Ondo, W.G., 2008. Ambulatory monitoring of freezing of gait in Parkinson's disease. *Journal of neuroscience methods*, 167(2), pp.340–348.

Moreau, C. et al., 2008. Externally provoked freezing of gait in open runways

in advanced Parkinson's disease results from motor and mental collapse. *Journal of neural transmission*, 115(10), pp.1431–1436.

Najafi, B. et al., 2003. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *IEEE Transactions on Biomedical Engineering*, 50(6), pp.711–723.

Niazmand, K. et al., 2011. Freezing of Gait detection in Parkinson's disease using accelerometer based smart clothes. In *IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, pp. 201–204.

Nieuwboer, A. et al., 1998. A frequency and correlation analysis of motor deficits in Parkinson patients. *Disability and rehabilitation*, 20(4), pp.142–150.

Nieuwboer, A. et al., 2009. The short-term effects of different cueing modalities on turn speed in people with Parkinson's disease. *Neurorehabilitation and neural repair*, 23(8), pp.831–836.

Nilsson, M.H. et al., 2010. Development and testing of a self administered version of the Freezing of Gait Questionnaire. *BMC neurology*, 10, p.85.

Nonnekes, J. et al., 2014. Short rapid steps to provoke freezing of gait in Parkinson's disease. *Journal of Neurology*, 261, pp.1763–1767.

Okuma, Y., 2006. Freezing of gait in Parkinson's disease. *Journal of neurology*, 253(7), p.VII27-32.

Papapetropoulos, S.S., 2012. Patient Diaries As a Clinical Endpoint in Parkinson's Disease Clinical Trials. *CNS Neuroscience & Therapeutics*, 18(5), pp.380–387.

Pastorino, M. et al., 2011. Assessment of Bradykinesia in Parkinson's disease patients through a multi-parametric system. In *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 1810–1813.

Pérez-López, C., Samà, A., Rodríguez-Martín, D., Català, A., et al., 2016. Assessing Motor Fluctuations in Parkinson's Disease Patients Based on a Single Inertial Sensor. *Sensors*, 16(12), p.2132.

Pérez-López, C., Samà, A., Rodríguez-Martín, D., Moreno-Aróstegui, J.M., et al., 2016. Dopaminergic-induced dyskinesia assessment based on a single belt-worn accelerometer. *Artificial Intelligence in Medicine*, 67, pp.47–56.

Podsiadlo, D. & Richardson, S., 1991. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39(2), pp.142–8.

Reyes Ortiz, J.L., 2015. *Smartphone-Based Human Activity Recognition*, Springer International Publishing. Available at: http://link.springer.com/10.1007/978-3-319-14274-6.

Rodriguez-Martin, D. et al., 2013. SVM-based posture identification with a single waist-located triaxial accelerometer. *Expert Systems with Applications*, 40(18), pp.7203–7211.

Rodríguez-Martín, D. et al., 2013. A Wearable Inertial Measurement Unit for Long-Term Monitoring in the Dependency Care Area. *Sensors*, 13, pp.14079–14104.

Salarian, A. et al., 2007. Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. *IEEE transactions on bio-medical engineering*, 54(2), pp.313–322.

Samà, A. et al., 2012. Dyskinesia and motor state detection in Parkinson's Disease patients with a single movement sensor. In *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012*. pp. 1194–1197.

Sayeed, T. et al., 2015. Adapted step length estimators for patients with Parkinson's disease using a lateral belt worn accelerometer. *Technology and health care : official journal of the European Society for Engineering and Medicine*, 23(2), pp.179–94.

Schaafsma, J.D. et al., 2003. Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease. *European Journal of Neurology*, 10(4), pp.391–398.

Snijders, A.H. et al., 2012. Freezer or non-freezer: clinical assessment of freezing of gait. *Parkinsonism & related disorders*, 18(2), pp.149–54.

Tripoliti, E.E. et al., 2013. Automatic detection of freezing of gait events in patients with Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 110(1), pp.12–26.

Wang, W.-H. et al., 2013. Inertial-Sensor-Based Balance Analysis System for Patients with Alzheimer's Disease. In *Conference on Technologies and Applications of Artificial Intelligence*. IEEE, pp. 128–133.

Zabaleta, H., Keller, T. & Fimbel, E.J., 2008. Gait analysis in frequency domain for freezing detection in patients with Parkinson's disease. *Gerontechnology*, 7(2), p.4.

Zhao, Y. et al., 2012. Online FOG Identification in Parkinson's disease with a time-frequency combined Algorithm. *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, 25(Bhi), pp.192–195.

Zhou, H. & Hu, H., 2008. Human motion tracking for rehabilitation—A survey. *Biomedical Signal Processing and Control*, 3(1), pp.1–18.