

POLITECNICO OF TURIN

MASTER'S DEGREE IN COMPUTER ENGINEERING

DATA SCIENCE

---

Freezing of gait in Parkinson's  
disease: automatic early  
recognition of episodes from  
patients' inertial data

---

*Author:*

Gianluca AMPRIMO

*Supervisor:*

Dr. Gabriella OLMO

*Co-Supervisor:*

Luigi BORZÌ

October, 2020



# Abstract

Freezing of gait (FoG) in Parkinson’s disease is “a brief, episodic absence or marked reduction of forward progression of the feet despite the intention to walk” [79]. It is a common symptom in 50-80% of the patients [79], occurring most frequently during turning, when passing through narrow paths or overcoming an obstacle, and is strongly affected by cognitive aspects such as attention, stress or anxiety [61]. Having a strong correlation with falls [47], FoG is considered as one of the most dangerous symptoms, with severe effects on autonomy and Quality of Life (QoL). Due to the great variability of FoG features, treatments have to be patient-specific. This aspect, together with the need of further investigating its causes, requires to develop objective methodologies to assess FoG. Laboratory assessment typically includes FoG eliciting in an artificial setting, with an altered emotional state of the patient due to examination anxiety. Thus many literature studies have been recently focused on home evaluation systems [8]. In this regard, a popular option is based on a combination of data coming from wearable inertial sensors with machine learning algorithms, trained in order to detect FoG events. Although different, all these methodologies share a similar processing pipeline. The aim of this study was to explore a new approach to the problem, performing additional offline preprocessing to identify “regions of interest” in the data, where there is a higher probability to identify FoG. Such step was introduced to increase time efficiency during the analysis and the precision of the final classifier predictions. In this study, tri-axial acceleration data coming from two previous experiments (defined as “Phase 1” and “Phase 2”), involving a total of 85 subjects and a waist-mounted commercial smartphone for data collection, were employed. Phase 1 data were used to design the algorithm for the selection of “region of interest”, based on Continuous Wavelet Transform (CWT). On the selected pieces, a more traditional pipeline was implemented: the signals were segmented through a sliding window; from each segment a set of relevant spectral and temporal features was extracted and fed in input to several classifiers (K-NN, SVM, Random Forest) to compare their performances. Both multiclass and binary classification were explored. The best results were obtained using binary Support Vector Machine with RBF kernel, achieving accuracy, recall and precision of 95%, 84%, 90% respectively in a 10-fold stratified crossvalidation. Furthermore, the use of the implemented window of interest was found to increase specificity, in a false positive test over non-freezer patients, by 35 percentage points (from 62% to 97%).

# Acknowledgements

This work is the culmination of a long path, sometimes tortuous but always full of satisfactions. First of all, I want to express all my gratitude to my supervisor Dr. Gabriella Olmo, who proposed me this study, which joined together my personal interest for the biomedical field and my study path in machine learning. Thanks for the support and availability. Second but not less important, I want to wish all the best to my co-supervisor Luigi Borzì who introduced me to FoG recognition and provided me with plenty of useful suggestions, always letting me free to explore the subject, but coming to my help whenever I needed. Really thank you.

Then, I want to thank my parents, Maria and Valter, for having raised me as the man I am and for supporting me in my studies, often putting up with my moody behaviour during exam sessions. Many thanks to my sister Elisa and my brother-in-law Simone, for always saying the best of me and to the newcomer, my little nephew Samuel, who really filled me with joy while I was writing this thesis. Thanks also to my grandmother Mariuccia, who always wanted an engineer in her family and encouraged me to be one; to Luisa, for supplying me with good food when most needed and to my best friend Giuseppe, for providing me with wise advises when important decisions were on the plate.

At last but not least, to my life companion (girlfriend would not be enough) Marta, who shared this engineering path with me: life is difficult and scary, but with you by my side I can endure everything. Thank you.

# Abbreviations

**CWT** Continuous Wavelet Transform. 26–30, 32, 35, 45, 53

**DBS** Deep Brain Stimulation. 6, 7, 12

**FN** False Negative. 41

**FoG** Freezing of Gait. 2, 4, 9–19, 21–26, 29–35, 37–39, 41, 42, 44–50, 52

**FP** False Positive. 41, 44, 49, 50, 52

**KNN** K-Nearest Neighbours. 38, 39, 42, 48, 49, 52

**LBs** Lewy Bodies. 2

**MI** Mutual Information. 34, 38

**PCA** Principal Component Analysis. 36

**PD** Parkinson's disease. 1–7, 9, 12, 13, 30

**QoL** Quality of life. 3–7, 14

**RF** Random Forest. 38, 40, 41, 43, 48, 49

**RI** Region of Interest. 26, 27, 29–31, 44, 46, 49, 52

**RIE** Regions of Interest Extractor. 25, 27, 29, 30, 32, 33, 43–46, 49, 50, 52, 53

**SVM** Support vector machine. 38, 39, 42, 48, 49, 52, 53

**TN** True Negative. 41

**TP** True Positive. 41

**UPDRS** Unified Parkinson's Disease Rating Scale. 14

**UPS** Ubiquitin-proteasome system. 2

# List of Figures

1	Inertial signals of activities interrupted by FoG, measured from different placements. Ref: [74] . . . . .	11
2	a)anterior, b)vertical, c)lateral acceleration signals from tri-axial accelerometer in a smartphone. . . . .	17
3	Spectrum of different activities compared to FoG. A clear shift to the 3-8Hz range can be appreciated. Ref: [69] . . . . .	18
4	Positioning of the smartphone employed for data recordings. Ref: [15]	20
5	1-minute extract of recorded inertial signals from freezing patient. a) Acceleration from accelerometer by components versus time; b) angular velocity from gyroscope by components versus time; c) activities recognised from video recordings versus time . . . . .	21
6	Most relevant inertial signals during a)walk, b)turn, c)stand, d)FoG. During walk, peaks of acceleration on x and z are synchronized; during turn angular velocity on x and z increases, acceleration on x decreases; in stand all signals are almost 0; during FoG altered signals are registered . . . . .	22
7	1-minute extract of recorded inertial signals from freezing patient in Phase 2. a) Acceleration from accelerometer by components vs time; b) angular velocity from gyroscope by components vs time; c) activities recognised from video recordings vs time . . . . .	23
8	Distribution of duration of FoG episodes in Phase 1 . . . . .	24
9	Distribution of duration of FoG episodes in Phase 2 . . . . .	24
10	Pipeline used for training the proposed FoG recognition algorithm	25
11	Testing pipeline for the proposed FoG recognition algorithm . . .	25
12	Schematic representation of CWT. A wavelet is scaled, shifted and convolved with a signal at a certain time. The resulting coefficient is mapped on a location(time)-scale(frequency) plane. Ref: [2] . .	27
13	Mexican hat wavelet. Ref: [2] . . . . .	28
14	Scalogram of $a_x(t)$ of a freezing patient. On y-axis both scale and the corresponding frequency are reported. As it can be appreciated, white pieces correspond to walking and turning. The red line corresponds to separation between walking and freezing band. . .	28
15	Internal pipeline of RIE. The numbers on the arrows correspond to steps listed in the text. . . . .	29
16	Regions of interest extracted from four different patients . . . . .	31
17	Relevance R vs features in optimal set $S$ . In the red box, the features scoring more than half the maximum value observed. . .	35
18	Distribution of training data variance among the computed principal components. In blue the cumulative variance, in red the variance explained by each component. . . . .	37
19	Scatter plot of training data with respect to principal components	37
20	Example of KNN algorithm. Ref: [51] . . . . .	39
21	A representation of hard margins SVM. Ref: [51] . . . . .	40

22	Type of windows that produced false positives in false positive test: a) the FP distribution in test without RIE; b) the FP distribution in test with RIE. Undefined activity 0, daily living mix -2 and stand 3 are responsible for most of the false positives. . . . .	50
----	---	----

## List of Tables

1	Comparison between Non Wearable (NWS) and Wearable (WS) systems from [73] . . . . .	16
2	Characteristics of inertial sensors in Samsung Galaxy s5 mini. Ref: [15] . . . . .	20
3	Statistics about FoG episodes duration . . . . .	23
4	Statistics about FoG episodes and their duration in Phase 2 . . . . .	24
5	Comparison of thresholding methods for RIE . . . . .	30
6	List of temporal features extracted from each window. . . . .	32
7	List of spectral features extracted from each window. . . . .	32
8	Mapping between activities and class labels . . . . .	33
9	Windows extracted from Phase 1 . . . . .	33
10	Windows in Phase 1 reduced dataset . . . . .	33
11	Feature set after selection . . . . .	35
12	Reduced dataset from Phase 1 and Phase 2 . . . . .	44
13	Optimal features from feature selection . . . . .	44
14	Execution time statistics. . . . .	45
15	Percentages of activities selected by RIE in Phase 1 . . . . .	46
16	Percentages of activities selected by RIE in Phase 2 . . . . .	46
17	Results of validation & test of multiclass classifiers trained on selected features . . . . .	47
18	Results of validation & test of multiclass classifiers trained on principal components . . . . .	47
19	Results of validation & test of binary classifiers trained on selected features . . . . .	47
20	Results of validation & test of binary classifiers trained on principal components . . . . .	47
21	Results of validation & test of multiclass classifiers (Phase 1+2) trained on selected features . . . . .	48
22	Results of validation & test of multiclass classifiers (Phase 1+2) trained on principal components . . . . .	48
23	Results of validation & test of binary classifiers (Phase 1+2) trained on selected features . . . . .	48
24	Results of validation & test of binary classifiers (Phase 1+2) trained on principal components . . . . .	49
25	Results of false positive test on non freezers from Phase 2 . . . . .	51

# Contents

<b>1</b>	<b>Parkinson's disease</b>	<b>1</b>
1.1	Incidence . . . . .	1
1.2	Pathogenesis . . . . .	2
1.3	Symptomatology . . . . .	2
1.3.1	Motor symptoms . . . . .	3
1.3.2	Non motor symptoms . . . . .	4
1.4	Treatment . . . . .	5
1.4.1	Pharmacological therapy . . . . .	6
1.4.2	Surgical therapy . . . . .	6
1.4.3	Motor rehabilitation . . . . .	7
1.5	Rating scales . . . . .	7
<b>2</b>	<b>Freezing of Gait</b>	<b>9</b>
2.1	Introduction . . . . .	9
2.2	Clinical features and etiology . . . . .	10
2.3	Treatment . . . . .	12
2.4	Assessment . . . . .	13
2.4.1	Subjective methods . . . . .	14
2.4.2	Objective methods . . . . .	15
2.5	Inertial methodologies for FoG recognition . . . . .	16
<b>3</b>	<b>Proposed detection algorithm</b>	<b>19</b>
3.1	Introduction . . . . .	19
3.2	Data description . . . . .	19
3.2.1	Data Phase 1 . . . . .	21
3.2.2	Data Phase 2 . . . . .	23
3.3	Proposed algorithm pipeline . . . . .	25
3.3.1	Regions of interest extractor (RIE) . . . . .	26
3.3.2	Windowing . . . . .	31
3.3.3	Features extraction . . . . .	32
3.3.4	Training set creation & feature selection . . . . .	33
3.3.5	Machine learning models . . . . .	38
3.4	Validation & Test . . . . .	41
<b>4</b>	<b>Results and discussion</b>	<b>45</b>
4.1	Statistics on RIE . . . . .	45
4.2	Results on Phase 1 data . . . . .	47
4.3	Results on Phase 2 data . . . . .	48
4.3.1	Validation and test results of retrained models . . . . .	48
4.3.2	False positive test results . . . . .	49
<b>5</b>	<b>Conclusions</b>	<b>52</b>
5.1	Achieved goals . . . . .	52
5.2	Future improvements . . . . .	52

# 1 Parkinson's disease

Parkinson's disease (PD) is a widespread neurodegenerative disorder characterized by a slow but irreversible progression. It affects a significant portion of the elder world population, without distinction between races or cultures. Around 10 million people are affected worldwide, with most of them over the age of 60 and one in ten under 50 [30]. Men seem to be slightly more affected than women [30]. PD involves the death of the dopamine<sup>1</sup>-producing neurons in a region of the brain called *substantia nigra*, which results in an unbalance between inhibitory and excitatory mechanisms. The trigger of this phenomenon is unclear, even though several studies suggest both genetic and environmental causes. Patients with PD suffer from bradykinesia (i.e. slowness of movement), rigidity, tremor and postural instability [91]; they shows progressive speech and swallowing impairment [88] as well as several non-motor symptoms, including sleep disturbances, depression, psychosis, autonomic and gastrointestinal dysfunction, and dementia [52]. There are not exams available to identify PD prior to the manifestation of the symptoms [12]; the evaluation of the progression of the disease in a patient is performed by physicians through a series of clinical tests, commonly scored by the Unified Parkinson's Disease Rating Scale (UPDRS). The most widespread treatment on the market is levodopa, a drug designed to increase dopamine levels, which produces a temporary reduction of the symptoms [89]. However, adjusting levodopa administration is not trivial and the progression of the disease reduces its benefits (wear off). At the state of the art, no permanent solution for reverting the disease effects or its progression was found, hence all treatments are focused on improving the quality of life of the patients, relieving the symptoms [26].

## 1.1 Incidence

PD is the most common movement disorder besides essential tremor and the second most common neurodegenerative disease [5]. It affects approximately 0.3% of the world population, but this statistic increases to 3% in subjects above 65 years. In general, being related to aging, Parkinson's disease is more widespread in countries with an higher average age. Moreover, the number of people with PD is expected to increase by more than 50% by 2030 due to rising life expectancy [28]. Indeed, over the past generation, the global burden of Parkinson's disease has more than doubled (from 2.5 million individuals in 1996 to 6.1 million in 2016) as a result of increasing numbers of older people, with possible contributions from longer disease length and environmental factors. [82] Early onset of sporadic PD is rare, with about 4% of patients developing clinical signs of the disease before 50 years [5], which is compliant with a mean age of onset identified at 55 [25]. Much later is the mean age at diagnosis, 70 years both for men and women [107]. The incidence<sup>2</sup> is around 8 to 18 per 100000 person-year, even though the value varies observing different age intervals: it ranges from 0.5 per 100,000 in the 30-40 year category to 120 per 100,000 in the oldest age category (over 70) [107]. Also

---

<sup>1</sup>neurotransmitter required for a correct movement control [52]

<sup>2</sup>number of new cases per population at risk in a given time period

sex seems to have a relevant role in the disease: in 2016, for instance, 6.1 million people worldwide were affected by the illness, 2.9 million (47.5%) were women and 3.2 million (52.5%) were men [82]. The male/female ratio generally ranges from 1.37 to 3.7 and it increases with age, suggesting that twice as many men than women suffer from PD [107]. Some studies have also investigated the variability of incidence due to ethnicity: highest incidence was identified among Whites and Hispanic [24] [107], whereas both prevalence<sup>3</sup> and incidence are lower in Asians than in Whites [72].

## 1.2 Pathogenesis

Parkinson's disease is related to the death of dopaminergic neurons in the substantia nigra. By the time of death, this region of the brain has lost 50–70% of its neurons compared with the same region in healthy individuals [26]. Substantia nigra is a part of the Basal Ganglia, which play a role in the initiation of voluntary movements and in the comparison between motor commands and feedback from evolving motion [7]. The cause behind the degenerative process in substantia nigra is not clear. The most popular hypotheses in literature suggest it could be due either to misfolding and aggregation of proteins or to mitochondrial dysfunction [16] [110]. Another significant clinical trait of PD is the formation of intracytoplasmic inclusions known as Lewy Bodies (LBs) that are present in surviving neurons of the substantia nigra as well as in other affected brain areas [34]. These aggregates are made of abnormally folded proteins, in particular  $\alpha$ -synuclein, parkin and ubiquitin and they are thought to be generated by a malfunctioning of the Ubiquitin-proteasome system (UPS), which is responsible for the degradation of overabundant proteins and cell waste products. This failure in UPS is considered a possible culprit for the neurodegeneration caused by PD: toxic accumulation of intracellular proteins and aberrant proteins could be detrimental to neuronal survival. Supporting this, the accumulation of misfolded and aggregated  $\alpha$ -synuclein is thought to be the primary pathogenic event in familial PD linked to mutations or multiplication of the  $\alpha$ -synuclein gene [59].

## 1.3 Symptomatology

Symptoms of PD are classified as motor and non motor. The former are the most evident and well-known; most of the patients are identified only when the first motor features arise, even though it was estimated that, at this point, up to 80% of dopaminergic cells in the nigro-striatal system are already lost [99]. The cardinal motor symptoms can be summarized by the acronym TRAP: Tremor at rest, Rigidity, Akinesia (or Bradykinesia) and Postural instability. Also flexed posture and Freezing of Gait (FoG- sudden motor blocks) have been added to the peculiarities of parkinsonism, with PD being the most common form [52]. Non-motor symptoms vary from patient to patient and are often under appreciated features of PD, although they may manifest as pre-symptoms also 10 or more years before the diagnosis [1]. They are not directly related to alterations of dopaminergic

---

<sup>3</sup>proportion of population with a disease at a specific point in time

pathways, therefore they may develop even in subjects where motor symptoms are under control [59]. With the progression of PD, non-motor symptoms contribute to severe disability, therefore dealing with them is crucial to improve patients' Quality of life (QoL).

### 1.3.1 Motor symptoms

**Tremor** is the most visible manifestation of Parkinson's disease. It is the predominant symptom for patients that are affected by so called "tremor-dominant PD", whereas it can be completely absent in others [7]. Resting tremor is usually asymmetric with moderate amplitude, medium (4–6 Hz) frequency. It is characterized by an agonist-antagonist alternate contraction pattern that arises at rest, but disappears or decreases with action or during sleep. Overall, tremor tends to appear in the distal part of an extremity, indeed it is often manifested in the form of a "pill-rolling" movement of the hands, but it may also affect lips, chin, jaw and legs. Instead, it rarely involves the neck/head or voice [52]. The pathophysiology of rest tremor is largely unknown, but it is considered in general different from that of bradykinesia and rigidity; its magnitude is not related to dopamine deficiency and it does not respond readily to dopaminergic medicament. It was hypothesized a connection with altered activity in the basal ganglia circuit, which is affected by dopamine neurons death, and the cerebello-thalamo-cortical circuit, which is also involved in many other tremors [42]. In addition to rest tremor, many patients with PD also have postural tremor that is more prominent and disabling and may be the first manifestation of the disease. This type of tremor is called "re-emergent tremor" and it is often delayed when the patient assumes an outstretched horizontal position [52]. The occurrence of rest tremor varies among patients and with disease stage. This symptom is observed clinically in 75 % of patients with PD. Deep brain stimulation seems to provide positive effects on tremor control [45].

**Rigidity** in Parkinson's disease patients is characterized by increased muscle tone to palpation at rest, decreased distension to passive movement, increased resistance to stretching and ease of the shortening reaction [86]. Rigidity is perceived throughout the full range of movement and it may be associated with pain: a painful shoulder is a frequent initial manifestation of PD although it is often misdiagnosed as arthritis, bursitis or rotator cuff injury [52]. Similarly to tremor, no direct correlation has been shown between dopamine deficiency and rigidity [86]. The clinical features of rigidity suggest a complex pathophysiological cause [86], related to changes in the passive mechanical properties of joints, tendons and muscles and aberrations in peripheral sensory inputs that may affect the response to muscle stretch [87].

**Bradykinesia** is characterised by reduced speed when starting and performing a single movement and progressive reduction of its amplitude, up to total cessation during repetitive simple movements [86]. It impairs in particular fine motor activities: in fact, it is often evaluated asking the patient to perform tasks

as opening and closing the hand, tapping thumb and index fingers, or tapping the foot on the ground [63]. Due to its nature, this symptom has substantial effects on daily living and QoL. Bradykinesia also manifests as loss of spontaneous movements and gesturing, drooling because of altered swallowing, monotonic and hypophonic dysarthria, loss of facial expression (hypomimia) and reduced blinking, and decreased arm swing during walk [52]. A peculiar trait of bradykinesia in PD is that patients are still able to correctly perform motor tasks when receiving a visual/auditory cue, for example catching a ball thrown at them. This phenomenon (kinesia paradoxa) suggests that patients with PD have intact motor programmes but struggle to access them without an external trigger [52]. Bradykinesia is the symptom that best correlates with dopaminergic deficiency [36] and the derived unbalance between inhibitory and excitatory mechanism in Basal Ganglia [86]. Although muscle weakness and other PD motor symptoms may contribute, the principal deficit was identified in an insufficient recruitment of muscle fibers during the start of movement [10].

**Postural instability** is a symptom that arises during the advanced stages of the disease and it is caused by a loss of straightening reflexes [52]. It largely contributes, together with FoG, to motor impairment, increasing the risk of falls and hip fractures, in particular in elder patients. [61]. The pathophysiology of axial postural abnormalities in PD is poorly understood, and several central and peripheral causes have been proposed, including asymmetry of the Basal Ganglia outflow, rigidity, dystonia, abnormal processing of vestibular or proprioceptive afferents, abnormal spatial cognition, focal myopathy in the paraspinal muscles, spinal and soft tissue changes, and side effects of dopaminergic and nondopaminergic drugs. [61] To assess abnormal postural stability, pull test is employed: the clinician quickly pulls the patient backward by the shoulders and quantifies the degree of retropulsion. If the patient takes more than two steps backwards or does not show postural response, this indicates an abnormal postural response [52]. Dopaminergic therapy and deep brain stimulation can improve some axial signs but usually do not robustly ameliorate postural instability [52].

### 1.3.2 Non motor symptoms

Non motor symptoms can be classified in autonomic dysfunction, cognitive and neurobehavioral disorders, and sensory and sleep abnormalities [52].

**Autonomic dysfunction** may either arise before the diagnosis or be caused by medications. It includes a variety of conditions, for example orthostatic hypotension, sweating dysfunction, sphincter dysfunction, bladder and erectile dysfunction. [52] Orthostatic hypotension affects 30–40% of patients and can produce dizziness, visual disturbances and altered cognition that may precede loss of consciousness when the patient is assuming the upright posture [99]. Sphincter and, in general, gastrointestinal dysfunctions are due to slowing of mobility of the gastrointestinal tract; 70-80% of patients suffer from slow-transit constipation [54]. A relevant related issue is dysphagia which increases the risk for polmonite ab ingeris, a common cause of death in PD patients [56]. Urinary control disturbances

include urinary frequency, urgency and incontinence and are often correlated with the progression of PD [109]. Erectile dysfunction is common in male subjects [99].

**Cognitive/neurobehavioural disorders** cover a wide range of different psychological pathologies as depression, apathy, dementia, anxiety, psychosis and hallucinations. They used to be underrecognized and undertreated, but awareness of their impact on the quality of life of patients is growing. Depression is a very common condition in patients with PD, with prevalence between 20% and 70%. However, pharmacologic treatment with antidepressant medications and cognitive behavioral interventions may significantly ameliorate its effects [14]. Apathy is often associated to depression, even though in some cases it may represent a separate phenomenon. It is defined as a lack of motivation, associated with reduced goal-oriented behavior and emotional expression. Up to 40% of PD patients suffer from apathy, in particular older men with drastic motor impairment, worse executive dysfunction, and a higher risk of dementia [14]. About dementia, it is generally developed in later stages of the illness. Nevertheless, also earlier symptoms related to impairment of cognitive ability (e.g. planning or organizing goal-directed behaviour), visuospatial dysfunction, impaired speech fluency and memory impairment are observed as precursors of PD [86]. Generalised anxiety, panic attacks, and social phobias are common in PD, about 40% of patients have anxiety that manifests as apprehensiveness, nervousness, irritability, and feelings of impending disaster as well as palpitations, hyperventilation, and insomnia [86]. Psychosis in the form of hallucinations, delusions and paranoia can occur in up to 30% of PD patients [26] and they may be associated with the assumption of dopaminergic medications. These drugs are also considered responsible for obsessive-compulsive and impulsive behaviour shown by some subjects, such as craving (especially for sweets), binge eating, hypersexuality, pathological gambling, compulsive shopping and punding, characterised by obsession with repetitive handling, examining, sorting and arranging of objects. These behavioural symptoms are referred to as “hedonistic homeostatic dysregulation” [52].

**Sleep and sensory disturbances** are common in PD patients. Nocturnal sleep disturbances occur in 60–98% of patients and are related to disease severity and levodopa intake. During daytime some patients, in extreme cases, experience sudden irresistible sleep attacks [26]. Among sensory disturbance, anosmia or hyposmia are found in at least 80% of patients even many years before the diagnosis [29]. 40–85% of the patient report sore limbs and experience different types of pain: musculoskeletal, oral, thoracal, abdominal and genital [99].

## 1.4 Treatment

There is not a therapy, at the moment, capable of stopping Parkinson’s disease progression, reversing its symptoms or producing a neuroprotective action on still undamaged dopaminergic neurons [26]. Only symptomatic treatment is available and it aims at improving the QoL of the patients: with a life expectancy from diagnosis of 17 years [46], they should carefully plan with their doctors a long-

term medical strategy. This strategy should take into account the age of the patient, the presence of cognitive impairment, comorbidities and the response of the subject to the therapy. The treatment should be designed not only to reduce the motor symptoms, but also the non motor ones caused either by the illness or the medications. In early PD the timing when to start the drug treatment is quite complex to define and the decision should be taken, with direct involvement of the patient, balancing physical impairment against the drug-related complications [26]. The two most common therapy are the pharmacological therapy and the surgical therapy, even though also muscular rehabilitation and physiotherapy prove beneficial in alleviating symptoms and musculoskeletal pain [61].

#### 1.4.1 Pharmacological therapy

The pharmacological therapy is designed to fix, even if temporary, the unbalance between inhibitory and excitatory mechanisms, compensating the drop in dopamine levels due to the degeneration of dopaminergic neurons.

**Levodopa** or L-DOPA is a dopamine precursor that is extensively employed in the treatment of Parkinson's disease [26]. It is normally administered together with carbidopa<sup>4</sup> for reducing the induced side effects (e.g. nausea) and maximizing levodopa transport into the central nervous system [68]. Levodopa was considered for many years the gold standard because produces, in particular at early stages, a good response in the patient. However, with the disease progression the majority of subjects experience shorter duration of response to individual doses (wearing-off symptoms), an alternation of good and poor response to the drug (on-off symptoms), involuntary movements of the head, trunk or limbs (dyskinesias), other motor side effects and psychosis [99].

**Dopamin agonists and MAO-B inhibitor** are a possible alternative to Levodopa. Dopamin agonists, like apomorphine, proved to be less likely to produce dyskinesias and the wearing-off phenomenon than levodopa, but they are more likely to cause hallucinations, confusion, and psychosis, especially in the elderly. Moreover, levodopa produces overall greater symptomatic benefits [31]. MAO-B inhibitors offer mildly improvements, can be administered together a levodopa therapy [31] and were able to improve PD motor symptoms, with effects that persist up to 7 years or more [50].

#### 1.4.2 Surgical therapy

The most common surgical intervention is Deep Brain Stimulation (DBS), which is normally performed at advanced stages of the disease, when severe impairments to QoL are present. DBS requires an electrode to be inserted through the skull to stimulate the globus pallidus, subthalamic nucleus, or thalamus. A pacemaker-like stimulator is implanted under the skin and wires connect the device to the electrode, in order to send, when necessary, impulses to reduce abnormal electrical

---

<sup>4</sup>aromatic amino acid decarboxylase (AADC) inhibitor

signals in the brain. This procedure can reduce bradykinesia, tremor and rigidity, but also drug-related motor complication. However, there is concern about the increased incidence of psychiatric side effects, especially depression, following DBS. Patients with cognitive impairment or significant depression are, therefore, not suitable for this treatment [26].

### 1.4.3 Motor rehabilitation

Rehabilitation aims at maximizing motor and cognitive functions and minimizing secondary complications, in order to optimize independence, safety, and QoL of the patient. Several rehabilitative approaches have been proposed, like non-specific physiotherapy (e.g. muscle strengthening and stretching, balance and postural exercises), occupational therapy, treadmill and robotic training, dance and martial arts therapy, multidisciplinary approaches including speech and cognitive therapy, motor imagery and action observation therapy, virtual reality and telerehabilitation. Such treatments tend to produce short-term improvements and cannot replace medicaments or surgery. Nevertheless, physical exercise is generally accepted as an adjuvant, because it has also a positive impact on non-motor symptoms [61].

## 1.5 Rating scales

Parkinson's disease (PD) is diagnosed and staged only on the basis of clinical tests, that involve the evaluation of the motor and non motor symptoms shown by the patient. A need for a unified international standard for the assessment of the disease progression led to the definition of several rating scales, among which Hoehn and Yahr (H & Y) scale and Unified Parkinson's Disease Rating scale (UPDRS) are the most largely employed and acknowledged.

**Hoehn and Yahr scale** is commonly used to compare groups of patients and to perform a gross estimate of disease progression, ranging from stage 0 (no signs of disease) to stage 5 (wheelchair bound or bedridden unless assisted) [52]. H & Y stages seem to correlate well with progression of motor complications, reduction of QoL, and neuroimaging studies of dopaminergic loss [40]. Its main issues are [53]:

- it does not describe in details the motor impairment of the patient;
- it does not include any information about non motor symptoms;
- it is quite insensitive to changes in a patient's clinical state, therefore it is not suitable for individual use.

**UPDRS** is the most used rating scale for Parkinson's disease and it consists in a questionnaire divided into sections that is filled out by the doctor administering the test and also the subject involved. It was defined in 1987, but it was revised in 2001 to solve some ambiguities and the lack of detailed instructions in some sections [40]. The current version (known as MDS-UPDRS) is organised in four parts:

- “Evaluation of mental activity, behaviour and mood”, which evaluates non-motor aspects and is filled out by the patient and/or the caregiver;
- “Self-evaluation of activities of daily living”, which evaluates motor aspects of daily living and is filled out by the patient and/or caregiver;
- “Evaluation of motor function”, which evaluates the current motor conditions of the patient according to a series of physical tests (e.g. pull test) performed by the physician, who is in charge of filling out this part of the questionnaire;
- “Evaluation of complications of therapy”, which is focused on assessing the effects produced by the current therapy on the patients in the weeks prior the test and it is compiled both according to the doctor’s observations and information from patient and/or caregiver.

Each item in the test receives a score between 0 and 4, with a total score that can range from 0 (no disability) to 199 (total disability). [96]

## 2 Freezing of Gait

### 2.1 Introduction

Freezing of Gait (FoG) is an episodic and complex event related to gait<sup>5</sup> impairment in Parkinson’s disease patients. FoG in the literature is referred to as “a brief, episodic absence or marked reduction of forward progression of the feet despite the intention to walk” [79]. Such definition tries to convey the nature of a phenomenon that is actually characterized by a high variability in its manifestation. Freezing of Gait can occur as an inability to start moving (*start hesitation*) or an arrest while walking, as well as episodes of shuffling forward with steps that are millimetres to a couple of centimetres in length [79]. Even though freezing events can happen also while performing a straight walk in a comfortable setting, it is more often triggered or worsened by challenging situations or provocative environments, such as changing direction (*turning hesitation*), approaching narrow doorways (*tight quarter hesitation*) or destinations (*destination hesitation*), moving into crowded spaces, walking on a slippery surface, crossing thresholds or changes in floor, stepping into an elevator or entering a revolving door [61]. Moreover, dual tasking (e.g. walking and talking together) can increase the probability of freezing [98], as well as emotional factors such as stress or anxiety, due to increased cognitive load [61]. FoG is one of the most debilitating motor symptoms of Parkinson’s disease (PD) as it may lead to a loss of independence and an increased risk of falls [47], which are not only harmful for patients, but represent a significant contribution to health care costs for society [13]. Freezing events are generally regarded as a feature of late PD that involves 80% of subjects in advanced stages [47]; however, it was observed that it may also occur in very early stages, affecting up to 26% of patients not yet exposed to levodopa [13]. It happens more frequently in men than in women and less frequently in patients in which the main symptom is tremor [60]. Nevertheless, not all patient experience FoG but the possibility that it will manifest in all advanced PD patients after a long enough washout period from medications has never been ruled out [74]. Some studies, however, criticise the division of patients in non-freezers versus freezers because of the difficulties of an objective assessment and suggest, instead, to classify patients along a continuous spectrum of freezing severity, ranging from no freezing at all at one end, to severe FoG at the other end [95]. Due to its episodic and unpredictable nature, FoG assessment is a complex task. First of all, patients are usually not well aware of how freezing verifies and this reduces the validity of administered questionnaires [95] [78]. Moreover, they are less likely to exhibit freezing in a controlled and unfamiliar setting like a hospital or a laboratory [95], due also to the greater attention they pay to the motor task they are asked to perform [79]. The possibility of carrying out an objective at home evaluation would be crucial to better study the mechanism underlying this phenomenon [8].

---

<sup>5</sup>locomotion achieved by motion of lower limbs

## 2.2 Clinical features and etiology

Freezing events are generally quite short: most episodes last less than 10 seconds and only a few last more than 30 seconds [92]. They are more frequent during off state, which is defined as a withdrawal from antiparkinson medication for at least 12 hours [13]. Some relevant features are observed in affected patients:

- the foot or toe does not leave the ground or only barely clears the support surface during gait [79];
- inertial acceleration signals have a frequency spectrum content shifted to the 3-8 Hz band, with respect to normal walking band between 0.5-3 Hz [69]. This peculiarity is correlated to an alternate trembling of the legs during the event [79];
- increase in cadence but shorter path length before FoG [79] [4];
- a feeling of being “glued to the floor” [13];
- the freezing event is commonly relieved by visual or auditory cues [61] [39];
- FoG can be asymmetrical, affecting mainly one foot or being elicited more easily by turning in one direction [79].

In addition to these, patients with severe FoG have a generally impaired and altered gait cycle, with higher variability both in stride timing and amplitude, which are easily captured as abnormal variations in inertial signals [95] [74], as shown in Figure 1. This phenomenon is amplified in particular for dual task tests, remarking the importance of attention to walking in freezers [98].

Overall, three different manifestation patterns are identified: trembling in place, shuffling forward and akinesia.

**Trembling in place** is, as already mentioned, an alternating tremor of the legs (knees) [79] and it is associated to an effort to overcome the motor block. It is the most frequent type [92].

**Shuffling forward** means that the patient is not able to take a normal-length forward step, but performs a series of short steps. Walking with short steps was also identified as a possible trigger for in-lab elicitation of fog [77].

**Akinesia** is an extreme form of bradikinesia that consists in complete absence of movement of the limbs or trunk. It is actually the less frequent of the three [13].

Freezing can sometimes involve also gait-unrelated tasks: motor blocks have been reported to occur in alternating repetitive movements of the fingers and during speech. These events resemble FoG in severity and frequency content [79].

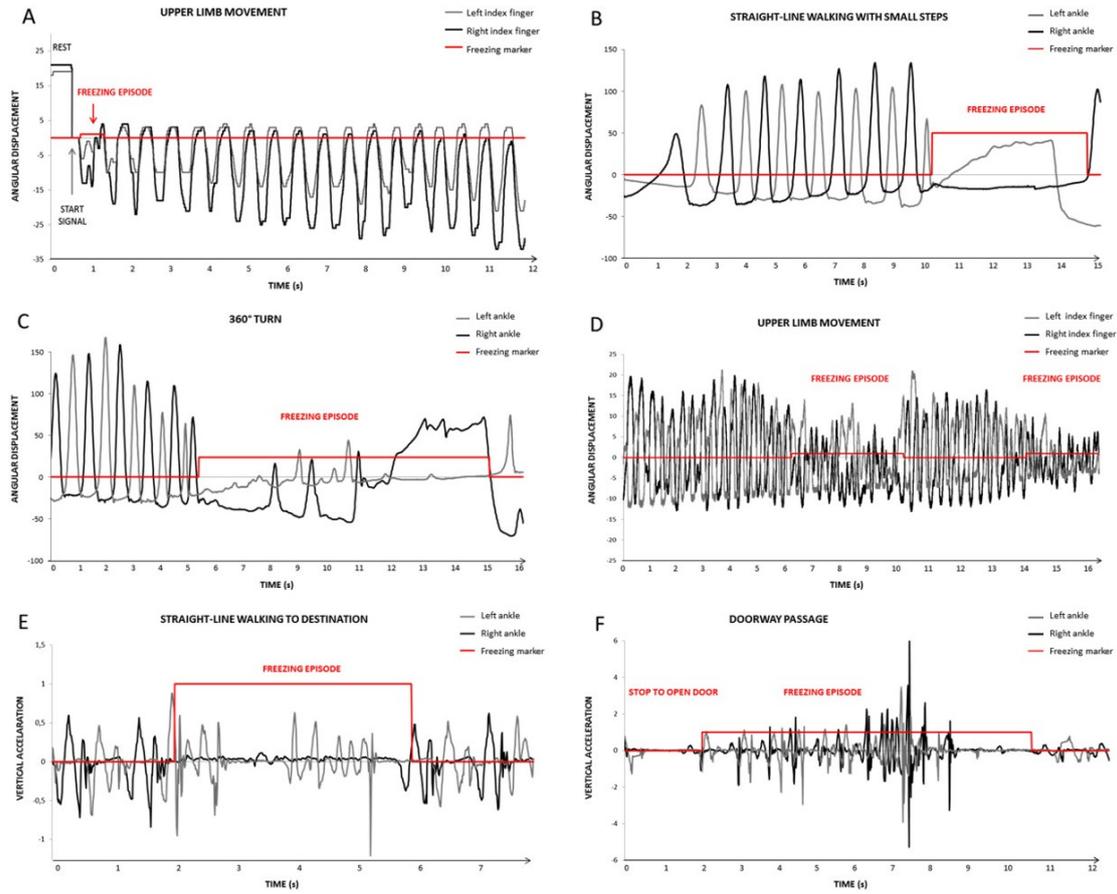


Figure 1: Inertial signals of activities interrupted by FoG, measured from different placements. Ref: [74]

Freezing of Gait etiology is poorly understood, but the trend has shifted toward considering a multisystem disfunctioning at the base of this symptom, rather than a pure motor-related disfunctioning. This was due to the raising awareness of the impact of cognitive/psychological alterations on FoG elicitation [61] [47] [37]. Five main hypotheses are being explored [79] [48]:

- **Abnormal gait pattern generation**, which is based on the observation of altered gait rhythmicity and gait cycle coordination, as well as shortening of steps, suggests that the main cause of FoG is due to abnormal output from the central pattern generators of the spinal cord. This results in high-frequency oscillations in both the lower and upper limbs during freezing episodes;
- **A problem with central drive and automaticity of movement**, which considers an impairment in automaticity and it is supported by the evidence that dual tasking can elicit the freezing event. Hence, FoG may be the result of disruption of the basal ganglia–supplementary motor area loop for self-initiated movement and this would also explain the relieving effect of video/auditory cues, which compensate this loss via the cerebellum–dorsal premotor cortex to maintain a central drive for locomotion;

- **Abnormal coupling of posture with gait**, which identifies the cause of freezing events in an unbalance between the anticipatory postural adjustment needed for shifting body weight in gait, and stepping. This hypothesis is supported by the trembling observed in legs, which is interpreted as a repeated but unsuccessful tentative of postural adjustment, caused by a disrupted basal ganglia mechanisms for preparing a motor programme;
- **A perceptual malfunction**, which suggests an exaggerated response to action-relevant visual information, due to the fact that patients affected by FoG decrease their gait speed and stride length to a much greater degree as a doorway is approached and tend to misjudge its size while walking. However, such effect is absent when the patient is seated, therefore this hypothesis requires further investigation and it is generally not regarded as a possible single cause;
- **A consequence of frontal executive dysfunction**, which considers the main cause of FoG to be an impairment in frontal executive functionalities, such as set-shifting, attention, problem solving, and response inhibition. Indeed, freezing is often triggered by challenging walking tasks that require the aforementioned properties and a fast switch in motor programmes. Nevertheless, this hypothesis requires further revision because not all PD patients with executive dysfunction show FoG.

## 2.3 Treatment

Designing an effective therapy for FoG, either pharmacological or surgical, is a cumbersome task. FoG has a variable and complex response to levodopa. Overall, three different FoG categories are identified:

- “OFF” FoG, which is the most common [78] and occurs when the patient is in a withdrawal phase from levodopa, hence it generally improves increasing the dosage or the frequency of administration of the medicament [48] [79] [92];
- “ON” FoG, which seems actually to be levodopa-induced [48] [79].
- “levodopa-unresponsive” FoG, which does not respond to the treatment and can verify both in OFF and ON state [48].

Differentiating between OFF and ON FoG is fundamental for tuning the drug therapy [95]. Dopaminergic agonists as an alternative to levodopa for “levodopa-unresponsive” FoG produced disappointing results and they could also worsen ON FoG [78]. MAO-B inhibitors have been associated with a decreased likelihood of developing FoG in a large randomised, controlled study, but they rarely reduce freezing once it has developed [79]. Also for DBS, available results are characterized by a great variability, are fairly small and limited, with a brief follow-up (on average 1 year after surgery) [78]. One study [76] showed that freezing severity improved in DBS-treated patients after 1 year, but in different subgroup of patients with PD, FoG was actually induced by surgery [32]. Some patients develop

or have worsened FoG and other axial motor problems several years after electrodes were implanted, however this could be a result of natural disease progression [78]. Adjusting the stimulator settings (e.g. decreasing stimulation voltage or frequency) improved the results in some cases [78]. Physiotherapy and physical exercise, as for other symptoms of Parkinson’s disease, can ameliorate FoG. Even patients with mild freezing should be educated about the phenomenon, especially about the risk of falls, various provoking circumstances, and preventive measures to follow, like conscious movement strategies to increase step amplitude, retaining stepping rhythm, making lateral weight shifts, directing attention to gait, and making wide arcs when turning [78]. A relevant rehabilitation strategy is related to the use of visual, auditory or somatosensory external cues that, as previously mentioned, can prevent or unlock freezing episodes [39] [84]. This mechanism could be exploited also for real-time FoG prevention, when combined with models for automatic detection [100]. Instruments for this kind of approach include laser projecting canes, metronome-like signals sent through headphones, smartglasses and headsets, tactor-stimulating sensors, often included as single components in a more complex assessment system [100]. The main drawback observed is the wear-off of the benefits due to the subject getting used to the continuous stimulation, hence the importance of avoiding cuing due to false positive events [100].

## 2.4 Assessment

Because of its clinical features, assessing Freezing of Gait is a challenging task for physicians. Besides its episodic and unpredictable nature, which requires to develop ad hoc protocols and instruments for its evaluation, freezing is characterized by a great variability among patients and several factors that can ameliorate (e.g. attention to gait, levodopa ON state), or worsen its occurrence (dual tasking, stress, anxiety, levodopa OFF state) [8] [47]. Assessment can be either performed in laboratory or at the patient’s house.

**In laboratory assessment** allows to evaluate FoG in patients both in ON and OFF state [57] and the usage of complex motion capturing systems such as multiple video recordings, optoelectronic sensors, floor sensors [73]. The main limitation of in-lab assessment is that the likelihood of a FoG manifestation is strongly reduced due to the artificial setting: wide halls or laboratories of hospitals are often unlike the patient’s house, where most of the freezing events occur. The inhibitory action of increased attention on gait and other emotional components like anxiety or stress contributes as well to this phenomenon [78].

**At home assessment** is based on either an evaluation made by the patient and/or the caregiver through questionnaire or diaries or on an objective evaluation through the usage of low-cost, low-consumption and portable sensors, among which wearable inertial sensors are a popular option [15]. The latter approach is very promising when combined with machine learning models trained on features extracted from inertial signals, but still largely under investigation [8]. This will be further discussed in section 2.5. At home evaluation allows to observe FoG

during so-called Activities of Daily Living (ADL) performed by the patient. Distinguishing the freezing events with respect to these activities is challenging, but it is of fundamental importance when aiming at designing FoG assessment (and prevention) solutions [8]. Moreover, given the familiar setting and its natural obstacles, the probability of observing freezing is much higher with respect to the in-lab methodology.

Moreover, FoG assessment methods can be generally classified as subjective or objective [8].

### 2.4.1 Subjective methods

Subjective evaluations are based on information coming from the patient or his/her caregiver and observations made by a qualified examiner. Some of them can be performed at home (diaries), others require the interaction with a physician in hospital (history-taking, scales and questionnaires, clinical examination).

**History-taking** is a preliminary assessment approach that consists in an interview with a clinician, who asks the subject to do a self evaluation of FoG. The patients are usually not familiar with the terminology, so alternative approaches should be employed: asking if he/she has ever felt a “glued-to-the-ground” feeling or if he/she has experienced sudden blocks while walking; showing video or performing a demonstration of what FoG looks like; asking about falls frequency, being the two events often related [8] [95].

**Scales and questionnaires** are another common approach that can be used by the clinician to evaluate FoG severity in a patient. The revised version of UPDRS contains a section about the presence and severity of FoG both OFF and ON medication. However, it only allows the examiner to get some insight into the presence and burden of FoG in daily life, and not to finely characterize its features [8]. Two specific questionnaire about FoG known as FOGQ [38] and NFOGQ [75] were defined to solve the aforementioned issues. The FOGQ can help clinicians screen for the presence of FoG, and also to assess the subjective severity [95]. The NFOGQ is very useful as it measures the severity of FoG in terms of frequency of occurrence, intensity and duration of the longest FoG episodes, and subjective impact on QoL and activities of daily living [8]. However, these questionnaires only assess FoG during turning and gait, and no other circumstances that commonly cause FoG, such as negotiating narrow passages or performing a dual task. In addition, they do not document the the environment in which FoG occurs. Finally, the treatment effect (ON or OFF state) in which FoG predominantly occurs is not scored [95].

**Diaries** can be used by patients as a powerful and cheap instrument for keeping track of FoG in daily living. A well structured diary records information about episodes such as time of day, triggering conditions (e.g. walking through a doorway), and whether the episode led to a fall or a near-fall. All these data can be

exploited to tune the medical treatment. The main drawback is that only patients with intact cognitive abilities can keep a diary and, depending on the purpose and duration over which it needs to be filled, it requires followup by a researcher or therapist every week to ensure correct completion [8].

**Clinical examination** is performed in-lab. The patient is asked to perform some motor activities, for example Timed Get Up and Go (TUG), in which he has to get up from a sitting position, walk a short distance, turn around, walk back to the chair and sit down again [73], or the even simpler Six Minute Walking Test (6MWT). Such tests are usually complicated to elicit FoG, using different tricks as dual tasking, either motor (e.g. carrying an object while walking) or cognitive (e.g. talking while walking), surpassing obstacles, turning of a wide angle, passing through narrow paths [13].

#### 2.4.2 Objective methods

Objective methods employ different kind of sensors (non wearable and wearable) and instrumentation for recording and evaluating FoG episodes, in most of the cases in a laboratory setting [8].

**Video recordings** are the gold standard for offline FoG recognition, because they allow to capture the event during the walking tests and study it later [8], applying if necessary image processing techniques for extracting further information [73]. Viewing the patient from multiple angles simultaneously may also be helpful in understanding the patient’s gait patterns [21]. Video tapes are often used in combination with the other techniques.

**Floor sensors** are placed along the floor on the so called “force platforms” or instrumented walkways, in which gait is measured by pressure or force sensors when the subject walks on them. There are two types of floor sensors: force plate and pressure measurement systems [73]. Force plates measure the force applied to the ground by the feet; not only the downward force exerted, but also braking and acceleration force and force directed mediolaterally [21].

**EMG systems** measure electrical manifestation of muscle during voluntary or involuntary contraction. If the posture of the patient or his movement dynamics are altered, it follows that the mechanics of movement will be abnormal too; hence, the electrical activity driving that movement will also be altered. However, muscle activity can be effectively assessed only in conjunction with a motion-capture system or force-plate data [21].

**Wearable sensors** are employed in several studies for gait analysis and FoG evaluation [101] [94], with a rise in their popularity due to the fact that they represent the most promising technique also for objective at home assessment. They include inertial sensors (accelerometers, gyroscopes, magnetometers), pressure and force sensor, piezoelectric sensors. Table 1 shows a summary of advantages

and disadvantages of wearable with respect to non-wearable sensors. This kind of systems are commonly coupled with ad hoc algorithms, often involving either shallow or deep machine learning techniques, using features extracted from their output signals to classify or predict motor activities [85].

**Smartphones** are receiving researchers' attention because they already embed inertial sensors like accelerometers and gyroscopes able to study human motion. Moreover, patients are usually not comfortable while wearing complex and heavy instrumentation, but they are used to carrying around a smartphone in their pocket. Several activity recognition systems have been recently developed for smartphone [33] [19] [9]. Smartphone use in FoG detection was explored in several studies [18] [44] [67] [15] [44] achieving performances comparable to other type of sensors.

Table 1: Comparison between Non Wearable (NWS) and Wearable (WS) systems from [73]

System	Advantages	Disadvantages
NWS	<ul style="list-style-type: none"> <li>-Allows simultaneous analysis of multiple gait parameters captured from different approaches</li> <li>-Non restricted by power consumption</li> <li>-Some systems are totally non-intrusive in terms of attaching sensors to the body</li> <li>-Complex analysis systems allow more precision and have more measurement capacity</li> <li>-Better repeatability, reproducibility and less external factor interference due to controlled environment.</li> <li>-Measurement process controlled in real time by the specialist.</li> </ul>	<ul style="list-style-type: none"> <li>-Normal subject gait can be altered due to walking space restrictions required by the measurement system</li> <li>-Expensive equipment and tests</li> <li>-Impossible to monitor real life gait outside the instrumented environment</li> </ul>
WS	<ul style="list-style-type: none"> <li>-Transparent analysis and monitoring of gait during daily activities and on the long term</li> <li>-Cheaper systems</li> <li>-Allows the possibility of deployment in any place, not needing controlled environments</li> <li>-Increasing availability of varied miniaturized sensors</li> <li>-Wireless systems enhance usability</li> <li>-In clinical gait analysis, promotes autonomy and active role of patients</li> </ul>	<ul style="list-style-type: none"> <li>-Power consumption restrictions due to limited battery duration</li> <li>-Complex algorithms needed to estimate parameters from inertial sensors</li> <li>-Allows analysis of limited number of gait parameters</li> <li>-Susceptible to noise and interference of external factors not controlled by specialist</li> </ul>

## 2.5 Inertial methodologies for FoG recognition

Inertial Measurement Units (IMUs) are one of the most widely used type of sensors in gait and FoG analysis, being low-cost, low-consumption, often wireless and easily miniaturizable [73]. Optimal type, number and placement are still argument of debate and different study propose different solutions [94]. Tri-axial accelerometers are the most used, either as a single sensor [3] [83] [112], or combined with gyroscopes [44] [27], or magnetometers [23]. Both one location and a combination of two or more locations have been used. The shin [23] [27] and waist [112] [18] were the most common choice, and could be used as single location. When two or more locations are considered, sensors were applied also on feet [70], knee [55], thigh [6], chest [105] or wrist [64]. Even if different approaches are used, a com-

mon pipeline can be identified among the different studies using inertial sensors for automatic FoG recognition. First of all, data are acquired, in lab or at home, through different protocols, either standard like TUG [70] [71] [62], 6MWT [15] or fog-elicitation-designed [106] [3]. Raw acceleration data (fig. 2) are usually filtered below 20-15 Hz using bandpass or lowpass filters, in order to remove noise and focus on the relevant components of the human motion spectrum [15] [17] [90] [85].

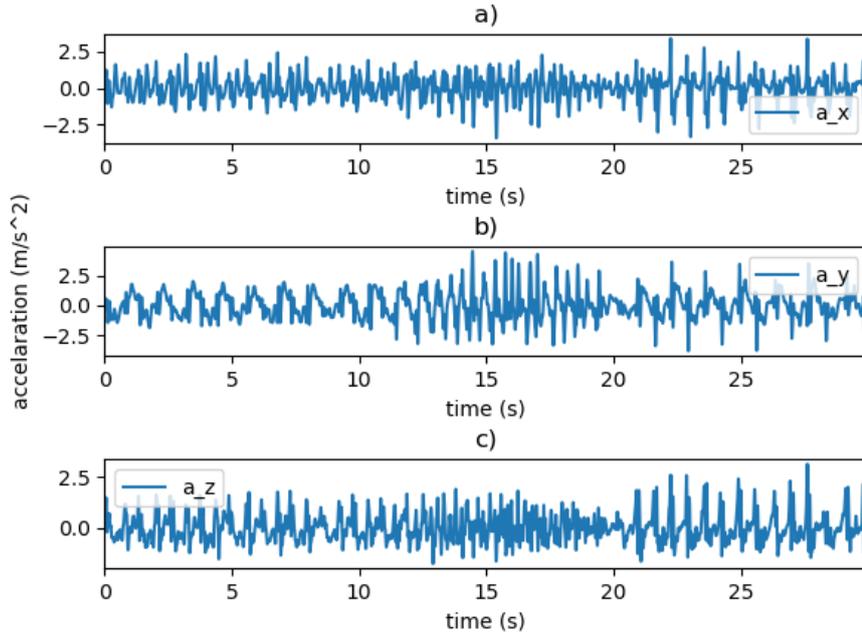


Figure 2: a)anterior, b)vertical, c)lateral acceleration signals from tri-axial accelerometer in a smartphone.

Further preprocessing to fix errors due to instrumentation can be applied [105], [15]. At this point, the signals are segmented using a sliding window with a given size and stride. Window size is a relevant parameter because the temporal resolution of the algorithm, i.e. the capability of identifying short-duration FoG episodes, is inversely proportional to the window duration [15]. Common windows in the literature range from 1 to 4s [85] [90] [69] [17] [58], usually with a half window length stride (50% overlap between consecutive segments). To knowledge of the writer, no approach for further reducing the input signals to relevant components prior segmentation has been tested. From each window, a set of temporal and spectral features are extracted, exploiting the peculiarity of inertial signals containing FoG, such as increased frequency content in the freezing band (3-8Hz) vs the walking band (0-3Hz) (fig. 3), abnormalities in gait (fig. 1) before and after the event (e.g. increase of peaks in acceleration due to shuffling steps, variations in the standard deviation of the signals). The choice of optimal features can be performed using different feature selection strategies, some exploit Pearson correlation [15], or mutual information to target [111]; Principal Component analysis is sometimes used for dimensionality reduction [85]. Eventually, the extracted features are employed to identify if a certain window contains FoG. The most

straightforward and simpler methods define a threshold on the extracted features (or on a single so called spectral “freezing index” [69] [83]), whereas more complex and performing solutions rely on machine learning models like Support Vector Machine [3], Naive Bayes [105], K-Nearest Neighbours [15], Decision Trees [105], Hidden Markov Models [90] or Neural Networks [90] [15] [17]. Threshold methods are less computational expensive, but overall achieve lower performances because it is complex to identify a threshold value valid for all the cases, due to FoG intrinsic variability in manifestation. Learning models reach better results in terms of specificity and sensitivity and can adapt better to variability, but a balance between computational burden and performance should be reached, in particular when designing real-time recognition systems [85].

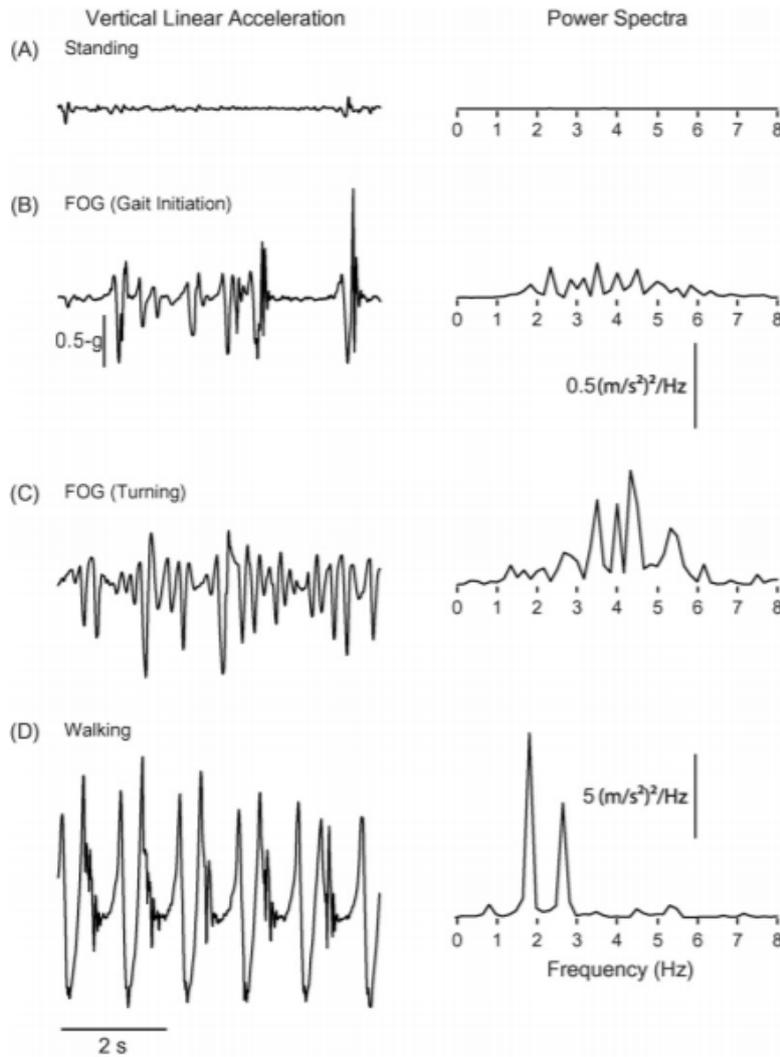


Figure 3: Spectrum of different activities compared to FoG. A clear shift to the 3-8Hz range can be appreciated. Ref: [69]

## 3 Proposed detection algorithm

### 3.1 Introduction

As described in section 2.5, many approaches for FoG assessment have been proposed which combine inertial sensors, in dedicated hardware or smartphones, and machine learning. As highlighted, all of them share a similar processing pipeline for the data, in particular regarding the segmentation of the inertial signals in consecutive windows. The aim of this study is to propose a new approach to the problem, which involves a further preprocessing step, to identify regions of interest where there is a higher probability to identify a FoG event. This additional operation was deemed important for two main reasons:

- **time efficiency**, because recorded signals can be quite long and reducing the analysis only to relevant portions can speed up significantly the computations;
- **precision**, because eliminating activities that could be misunderstood for FoG (false positive) could improve this metric.

About the second point, many studies report high values of accuracy and sensitivity in FoG classification, but they often do not include information about either precision or specificity, which could hinder that relevant results with respect to these two complementary metrics are not easily achieved [15]. However, the importance of false positive depends also on the type of application: if the monitoring is performed to produce a diary recording FoG episodes, to assess their frequency and duration, a lot of false positives have a great impact and could mislead physicians when designing the therapy; on the other hand, for a real time cueing application, a false positive would produce a non requested but innocuous stimulation, which could be accepted in order to catch as many episodes as possible (higher sensitivity). Nevertheless, whatever the application is, it is good practice to aim at both high precision and sensitivity, hence the importance that the proposed approach could have on the problem. In the following sections, the data employed in this study and the proposed algorithm pipeline will be described in details.

### 3.2 Data description

The data were collected in two different previous experiments [15], here defined for simplicity Phase 1 and Phase 2, performed in laboratory at University Hospital “Città della Salute e della Scienza”, Turin, Italy. In both, a commercial smartphone (Samsung Galaxy Mini s5) was mounted at the waist (around the third lumbar vertebra) of the subject as shown in Figure 4 and inertial signals from the internal accelerometer and gyroscope were recorded. Table 2 shows a summary of the characteristics of the two sensors.

Table 2: Characteristics of inertial sensors in Samsung Galaxy s5 mini. Ref: [15]

Sensor type	Range	Resolution	Sample frequency
Accelerometer	$\pm 2g$	40 mg	200 Hz
Gyroscope	$\pm 2000$ dps	60 dps	200 Hz



Figure 4: Positioning of the smartphone employed for data recordings. Ref: [15]

Using Matlab, the raw outputs of the sensors were recalibrated to remove contributions from gravity, filtered by a fourth-order zero-lag bandpass filter to the band 0.5-15Hz and the different activities they contained were recognised (and the corresponding time samples labeled) according to video recordings performed during the tests [15]. Finally, for each patient a Matlab matrix with a row for each time sample and the following columns was produced:

- x-axis acceleration
- y-axis acceleration
- z-axis acceleration
- x-axis angular velocity
- y-axis angular velocity
- z-axis angular velocity
- activity label

These output matrices were the starting point for this study.

### 3.2.1 Data Phase 1

Phase 1 contains data from 38 PD patients. The subjects were asked to perform a 6 Minutes Walking Test 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 was repeated for 6 minutes to increase the probability of eliciting a freezing event. No pause was planned during the test execution; nevertheless, the participants were free to quit or take breaks and later resume the exercise. To ensure safety, the tests were all performed under the supervision of clinical personnel. In addition to this, neither dual tasking or obstacle negotiation were included in the protocol [15]. Patients performed the test when different time has elapsed from their last levodopa in-take, so no neat distinction between on and off FoG was considered [15]. Figure 5 reports 1 minute of the recorded signals (after the previously mentioned preprocessing) from a patient with FoG.

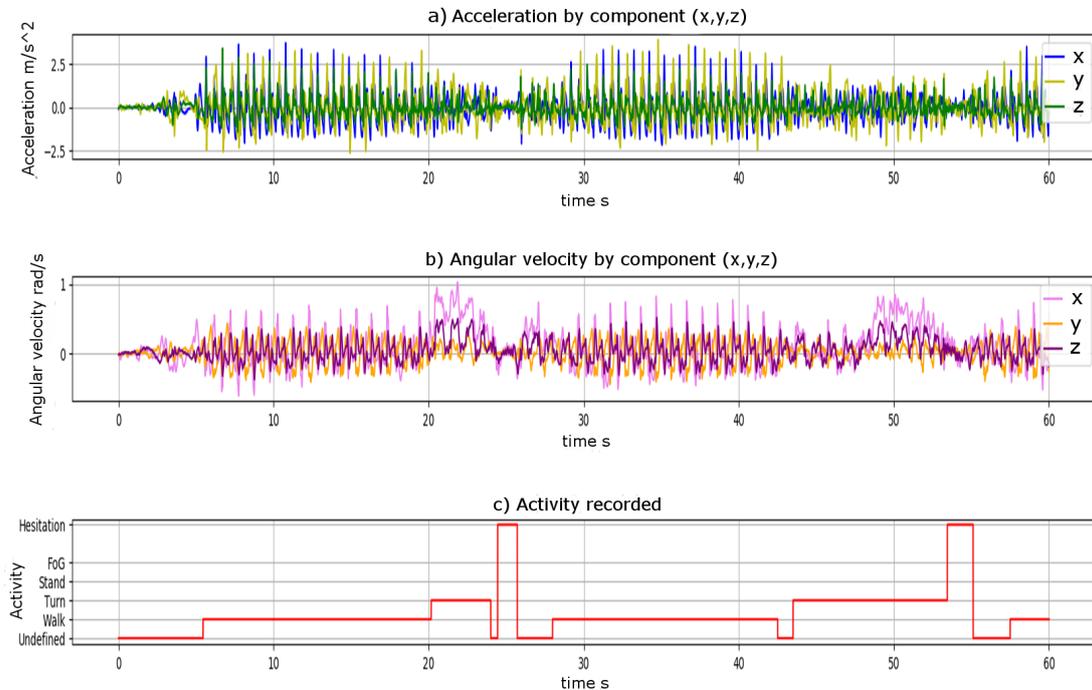


Figure 5: 1-minute extract of recorded inertial signals from freezing patient. a) Acceleration from accelerometer by components versus time; b) angular velocity from gyroscope by components versus time; c) activities recognised from video recordings versus time

From these data, four main activities can be identified: walk (label 1), turn (label 2), stand (label 3), FoG (label 4). In addition to these, label 6 represents short hesitations or brief FoG episodes: these events, as further discussed in 3.3.3, will be considered FoG as well. Finally, label 0 represents a mixture of pieces from different activities (walking, standing, smartphone positioning) that were not considered relevant; this label was maintained during selection of regions of interest, but discarded when training the final classifier. Figure 6 shows an

example of the most significant signals during the four main activities. Some relevant properties can be directly observed from the signals [15]:

- peaks in x-axis acceleration correspond to steps;
- peaks in z-axis acceleration correspond to steps as well and for this reason are often synchronized with peaks on x-axis component;
- turning is associated to increase of x-axis angular velocity;
- during standing all signals are practically 0;
- during FoG acceleration signals appear distorted and in general freezers have more irregular gait cycles.

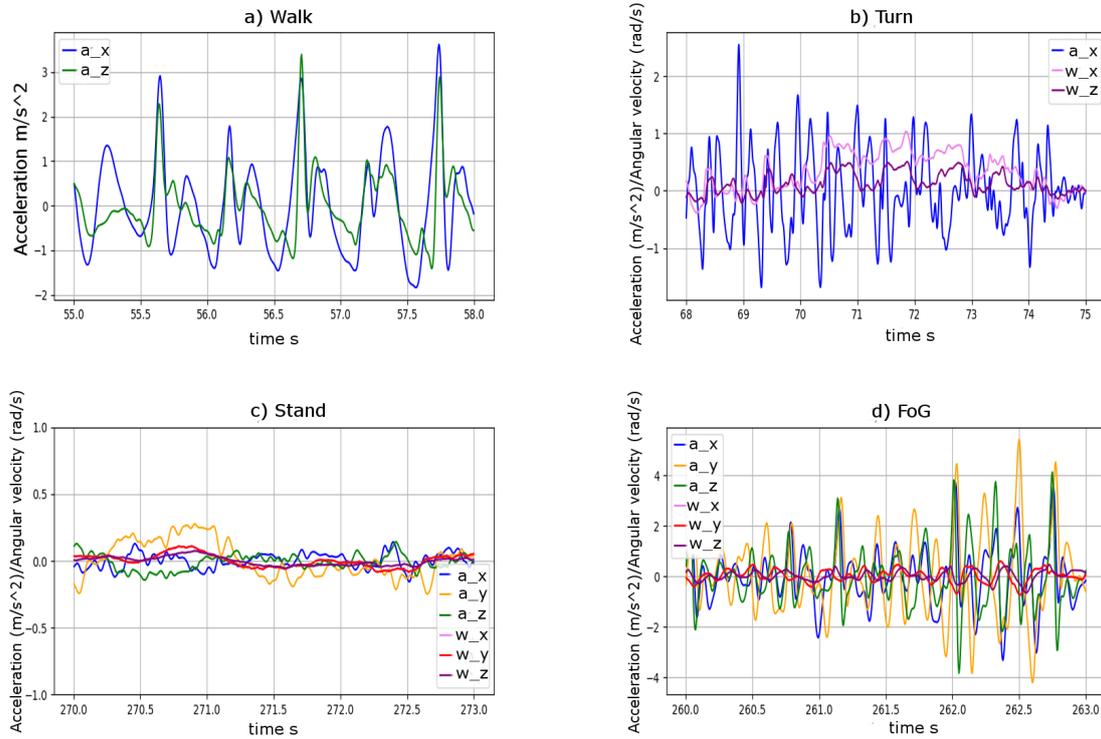


Figure 6: Most relevant inertial signals during a)walk, b)turn, c)stand, d)FoG. During walk, peaks of acceleration on x and z are synchronized; during turn angular velocity on x and z increases, acceleration on x decreases; in stand all signals are almost 0; during FoG altered signals are registered

Overall, only 5 patients showed FoG, in particular of the “trembling of legs” type, for a total duration of around 2 minutes. Table 3 contains some statistics about the length of the episodes, whereas Figure 8 shows histograms about the distribution of episodes length. As it can be appreciated most of the them are shorter than 4 s. Phase 1 data were used to design the proposed recognition algorithm and provide a first rough evaluation of its performances.

Table 3: Statistics about FoG episodes duration

N. of episodes	Max length [s]	Min length [s]	Avg length [s]	STD [s]	Total recorded length [s]
30	19.1	0.465	3.85	4.58	115.68

### 3.2.2 Data Phase 2

Phase 2 involved 47 patients. The experiment was conducted in a manner similar to Phase 1, but the recorded data include some additional activities and are generally longer. Figure 7 shows an example of about 1 minute of signals recorded for a freezer patient.

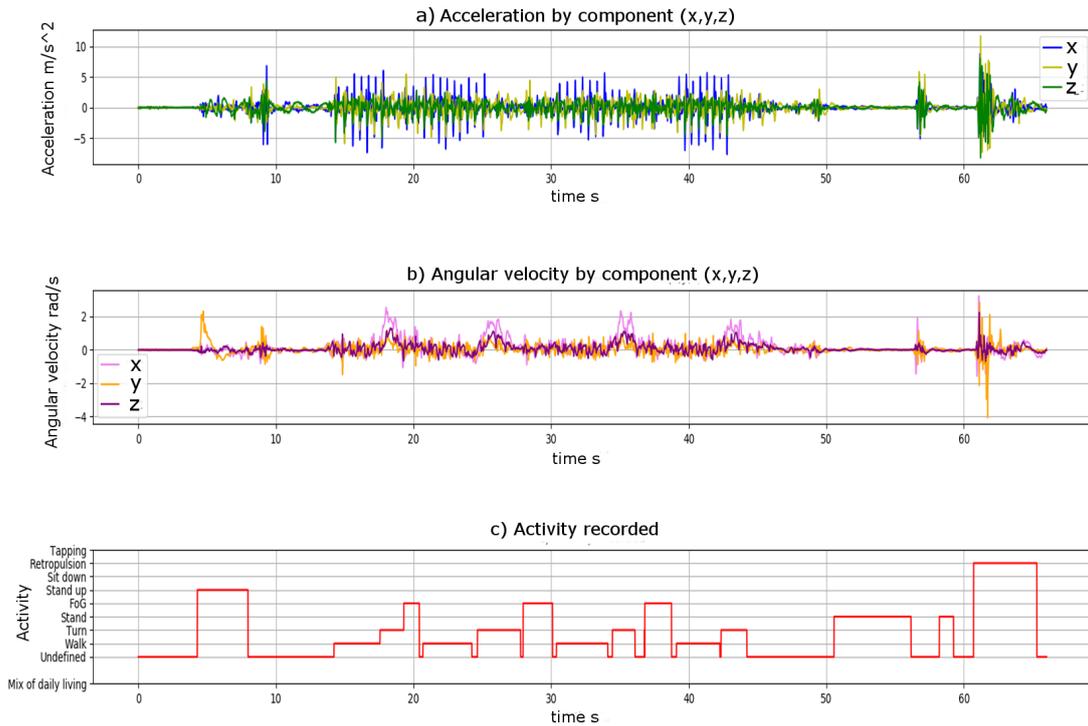


Figure 7: 1-minute extract of recorded inertial signals from freezing patient in Phase 2. a) Acceleration from accelerometer by components vs time; b) angular velocity from gyroscope by components vs time; c) activities recognised from video recordings vs time

The additional activities performed include:

- Stand up (label 5);
- Sit down (label 6);
- Retropulsion test (label 7);
- Tapping (label 8);
- Mix of daily living activities (e.g. taking a book, washing hands) (label -2);

The signals recorded in this phase are therefore more complex, but closer to the ones that could be recorded in actual daily living conditions, but for the physical tests (7-8) that indeed were not considered by the proposed FoG classifiers. Also in this case, only 5 patients manifested FoG during the test, but around 4 minutes of freezing events were identified. Table 4 contains some statistics about episodes duration, whereas Figure 9 shows histograms about the distribution of their length.

Table 4: Statistics about FoG episodes and their duration in Phase 2

N. of episodes	Max length [s]	Min length [s]	Avg length [s]	STD [s]	Total recorded length [s]
38	21	0.9	6.39	5.5	243

As it can be appreciated most of them are shorter than 8 s. However, with respect to the previous experiment longer events were overall observed. Phase 2 data were employed to improve the robustness of the classification models trained with Phase 1 data and provide a less variance-affected evaluation of their performances in validation and test. Moreover, non-freezers were employed for a false positive test with the aim of assessing the improvement in classification provided by the proposed pipeline.

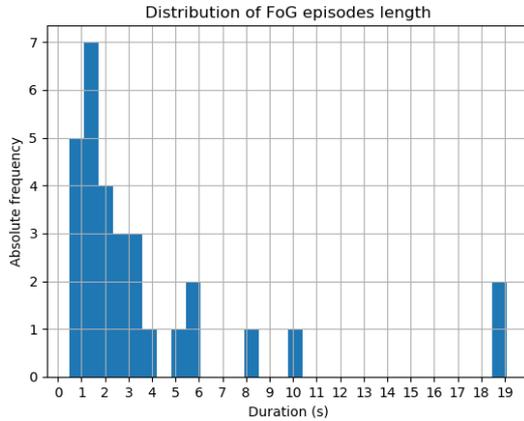


Figure 8: Distribution of duration of FoG episodes in Phase 1

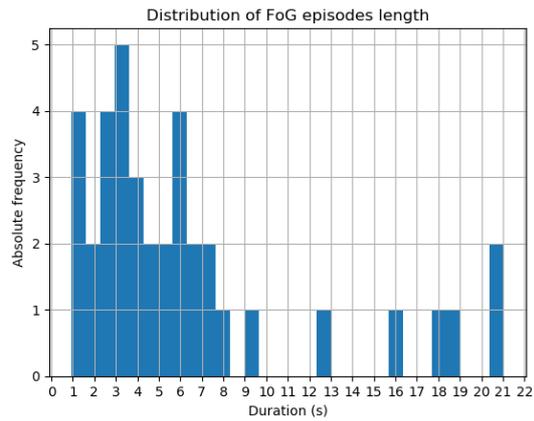


Figure 9: Distribution of duration of FoG episodes in Phase 2

### 3.3 Proposed algorithm pipeline

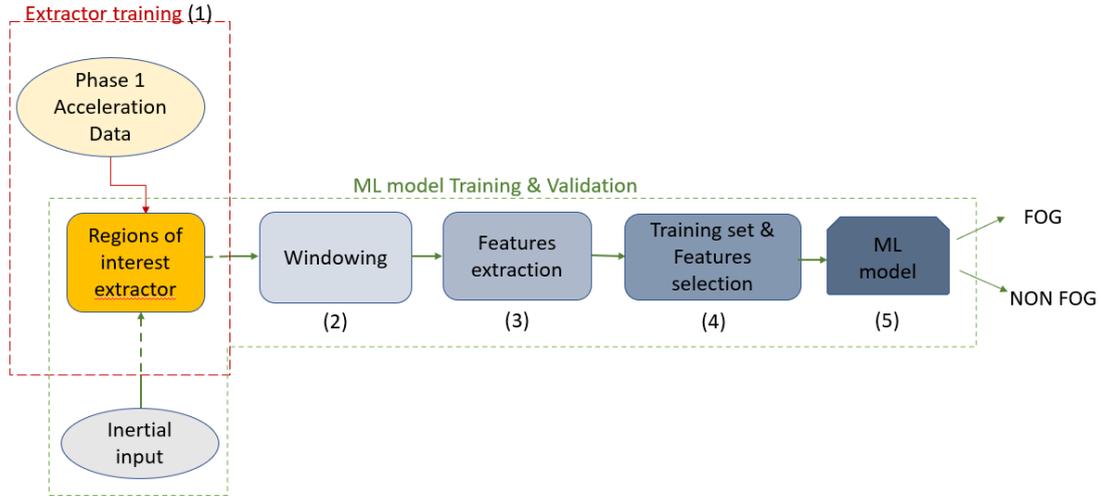


Figure 10: Pipeline used for training the proposed FoG recognition algorithm

Figure 10 reports the training pipeline for the proposed FoG recognition algorithm, which exploits only the **tri-axial acceleration signals** of the patients. First of all, acceleration data from Phase 1 were used to design the Regions of Interest Extractor (1). This was then applied to Phase 1 patients themselves and its output, for each subject, was segmented in windows of 2s (2). A set of FoG-related features was defined to characterize each window, as well as a class label representing the activity involved, needed for training of the final classifiers (3). A reduced dataset was defined starting from the data extracted from all patients and feature selection through mutual information was used to highlight the most significant features among the ones proposed in step 3 (4). The reduced dataset with the selected relevant features was eventually used to train, validate and test three different machine learning classifiers, whose performances in FoG recognition were compared (5). Once the whole algorithm was tuned and the best classifier was selected, the pipeline for processing and classifying new data reduced to the one in Figure 11.

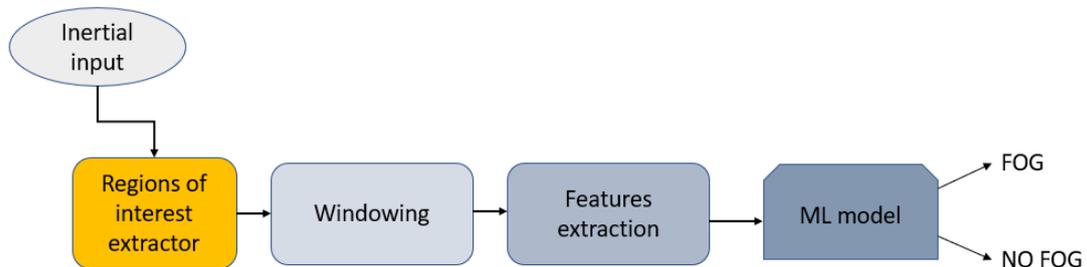


Figure 11: Testing pipeline for the proposed FoG recognition algorithm

### 3.3.1 Regions of interest extractor (RIE)

As discussed in Chapter 2, FoG occurs when a patient is walking (or starting to) and in particular when he/she is turning, hence an episode will be very likely identified close to these kind of activities. For the purpose of this study, a **Region of Interest (RI)** for FoG recognition was defined as a region of body acceleration containing either *walk* or *turn*. Several approaches exist in literature to identify gait or turning [102] [104] [108], but they are often rather complex [80] or require signals coming from more than one sensor [43] [35]. In this study, a continuous wavelet transform approach has been used for RI detection.

**Continuous Wavelet Transform (CWT)** is a mathematical transformation which allows to obtain a time-frequency representation of a signal, using a local wavelike function [2]. Such function, also known as the mother wavelet, can be scaled and shifted: each scaling represents a specific frequency, proportional to the centre frequency of the wavelet. In CWT, different scaled versions of the wavelet are shifted and convolved with the input signal, producing a map of coefficients, representing at each time instant which frequencies are more relevant. In a more formal definition, the CWT of a time *window* of acceleration signal  $a(t)$  is defined as a function of both scale  $s$  and translation time  $\tau$

$$C(s, \tau) = \int_0^{window} a(t) \Psi^* \left( \frac{t - \tau}{s} \right) dt \quad (1)$$

in which  $C(s, \tau)$  are the CWT coefficients of  $a(t)$  in *window* and  $\Psi$  is the chosen mother wavelet [83]. Such wavelike signal can be either real or complex, hence the need of the symbol of complex conjugation  $*$  in the formula. The frequency  $F_s$  associated to a certain scale  $s$  can be easily computed as

$$F_s = \frac{F_c}{s\Delta} \quad (2)$$

in which  $F_c$  is the centre frequency of the wavelet,  $\Delta$  is the sampling time of the input signal ( $\frac{1}{200}$  s in this study). It can be trivially observed that the scale is inversely proportional to the frequency, hence bigger scales will correspond to smaller frequencies. Figure 12 shows how equation 1 works in practice. CWT has recently found a number of applications for processing signals coming from different engineering and medical fields, among which gait analysis [11] [80] and FoG recognition [83], due also to its ability to detect abrupt discontinuities [2].

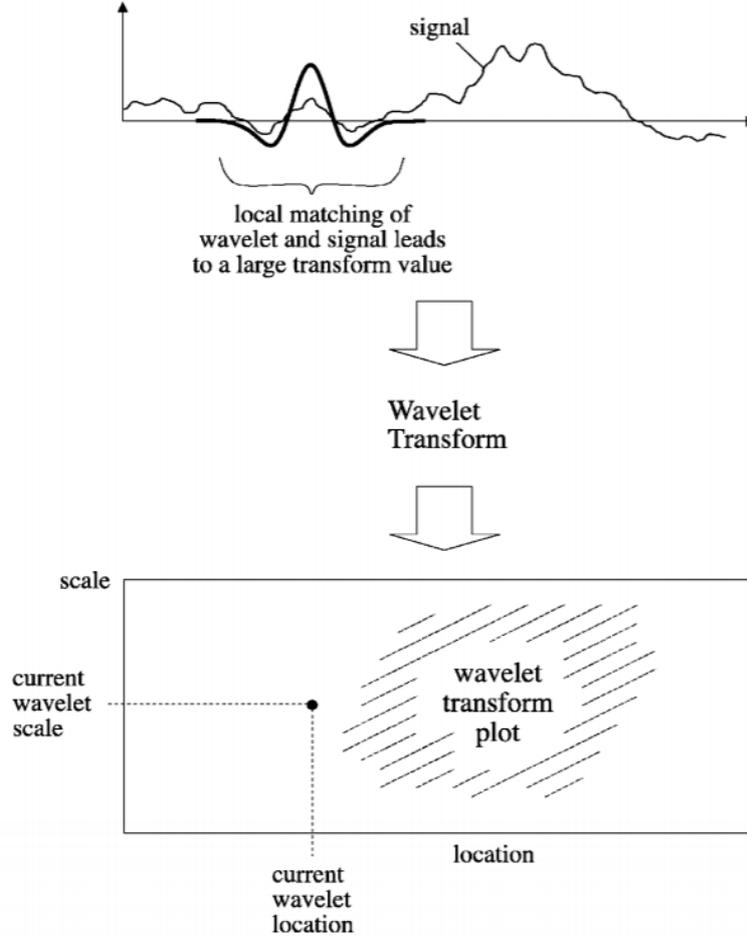


Figure 12: Schematic representation of CWT. A wavelet is scaled, shifted and convolved with a signal at a certain time. The resulting coefficient is mapped on a location(time)-scale(frequency) plane. Ref: [2]

In the proposed **Regions of Interest Extractor**, CWT was used to identify RIs as time portions of acceleration of patients having an high frequency content in the band 0.5-8 Hz, which is the band related both to walking (0.5-3Hz) and freezing. Actually only the x component of acceleration ( $a_x(t)$ ) was employed because it is the component which provides the most relevant information about gait and steps. A combination with z-axis component was considered but not implemented, after observing that also in terms of CWT, the two components provide almost identical results, only differently scaled. As mother wavelet, “Mexican Hat” (fig. 13) function was chosen

$$\Psi(t) = (1 - t^2)e^{-\frac{t^2}{2}} \quad (3)$$

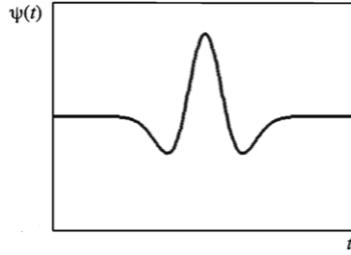


Figure 13: Mexican hat wavelet. Ref: [2]

This kind of wavelet is a fairly standard choice for CWT and its shape is similar to the spiky behaviour observed in  $a_x(t)$  while walking; also other continuous wavelets (e.g. “Morlet” wavelet), were evaluated, but “Mexican Hat” seemed to provide the best results. The set of scales to consider was defined from equation 2 in order to span all frequencies between 0.5-8Hz with a 0.5 step (e.g. 0.5, 1, 1.5, 2...), a choice which provided good results in [83]. Figure 14 reports the scalogram obtained when applying CWT to  $a_x(t)$  of a freezing patients from Phase 1.

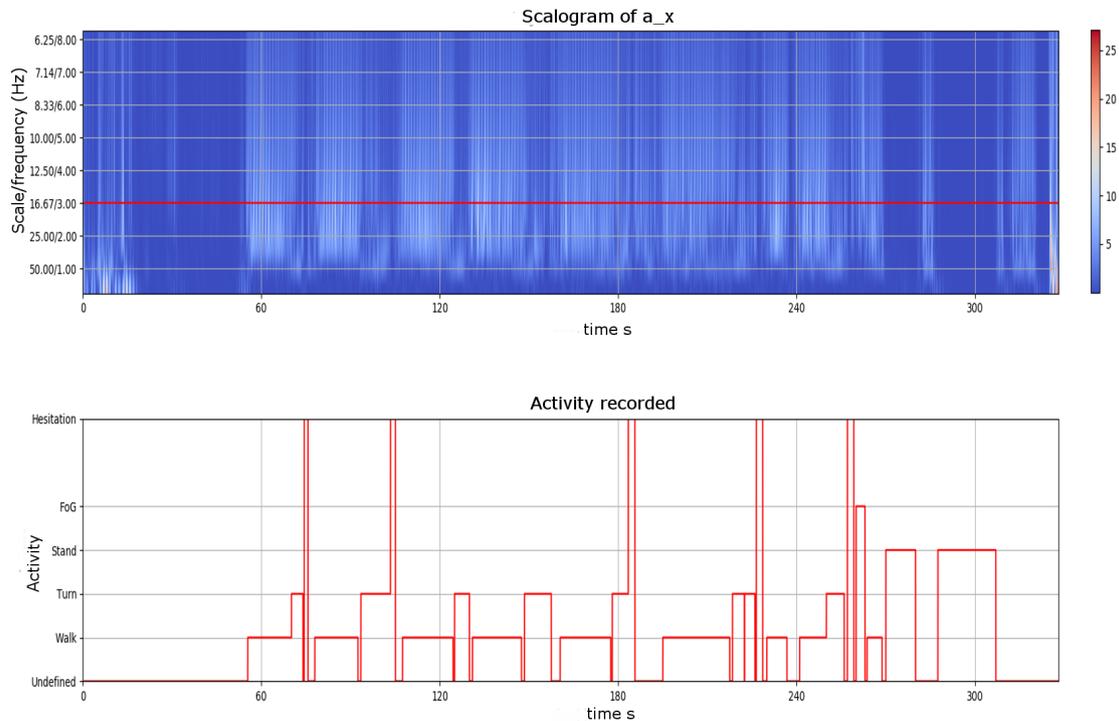


Figure 14: Scalogram of  $a_x(t)$  of a freezing patient. On y-axis both scale and the corresponding frequency are reported. As it can be appreciated, white pieces correspond to walking and turning. The red line corresponds to separation between walking and freezing band.

A scalogram is a plot having on x-axis the time reference and on y-axis the value of scale  $s$  considered, which can be easily converted in a corresponding frequency value from equation 2. The colour of the points in the plane is shaded according to the absolute value of the corresponding CWT coefficients. By visual

inspection of the scalogram, it is trivial to identify regions where the patient is “active”; more complex is the automatic identification of such time intervals from CWT coefficients. The following procedure, represented in Figure 15, was used:

1. for each time instant  $t$ , an average coefficient is computed as

$$C_{avg}(t) = \frac{1}{|S|} \sum_{s \in S} C(t, s) \quad (4)$$

in which  $S$  is the set of considered scales and  $|S|$  its cardinality,  $C(s, t)$  the CWT coefficient of  $a_x(t)$  at time  $t$  for scale  $s$ ;

2. on the signal made by average coefficients  $C_{avg}(t)$ , a **moving average** of window  $w$  is used to compute signal  $C_{ma}(t)$ , in which short-term fluctuations are smoothed out. Specifically, a window  $w$  of 1 s was selected, which seemed a reasonable value, considering that a patient will likely perform a certain activity for a much longer period of time;
3. on  $C_{ma}(t)$ , **thresholding** is applied to distinguish between time intervals corresponding to RI and non-RI;
4. RIs are extended of a time length  $l$  in order to include FoG that verifies at their beginning or end.
5. from all the three components of acceleration  $a_x(t)$ ,  $a_y(t)$ ,  $a_z(t)$ , the extended RIs are extracted to be provided in input to the subsequent steps of the overall pipeline.

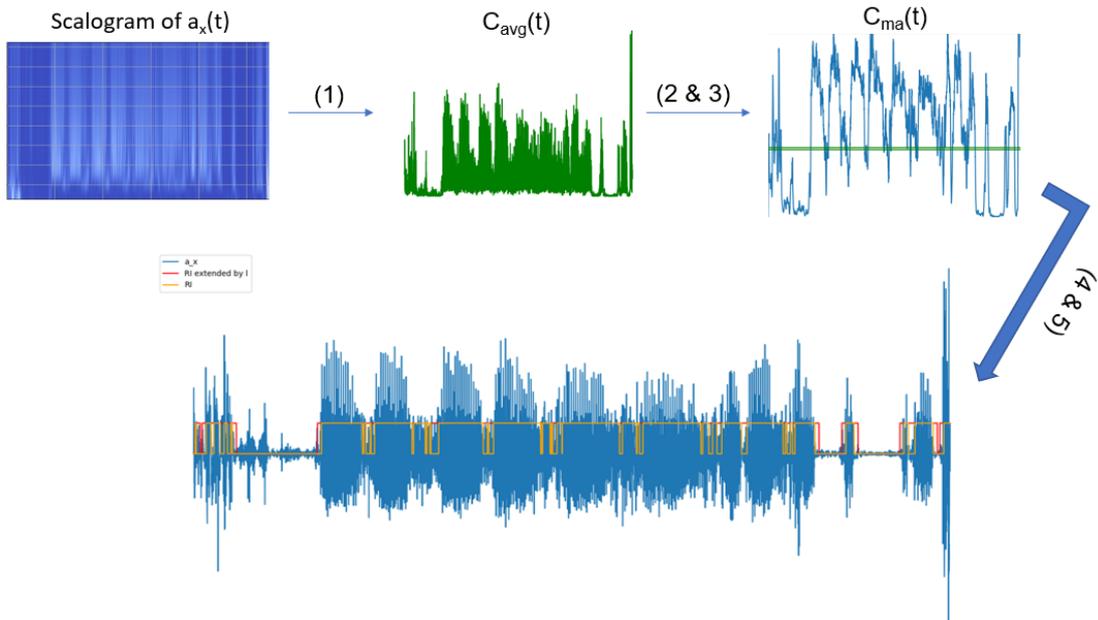


Figure 15: Internal pipeline of RIE. The numbers on the arrows correspond to steps listed in the text.

For thresholding, three alternative approaches were compared:

- a single fixed threshold
- a double fixed threshold (hysteresis)
- a patient-tailored threshold

**Single and double fixed threshold** were empirically estimated from the data. The moving averages of all the patients in Phase 1 were concatenated and optimal thresholds were searched through an exhaustive search, considering values between the minimum and maximum of the signal, with a 0.1 step. The optimal selected threshold, or pair of thresholds, was the one that included in the extracted RIs as much as possible walk, turn and FoG, while leaving out unwanted activities (stand, label 0). For single threshold, the identified value was 1.9, whereas for hysteresis the best pair was (1.5, 1.54). Hysteresis was explored to make the system more robust to residual fluctuations in moving average that could interrupt a continuous RI.

**A patient-tailored threshold** was considered because the intrinsic variability of gait parameters among individuals, especially for PD patients, limits the possibility of identifying a threshold that would be optimal in every case. Therefore, for each subject, a random window  $W$  containing walk was extracted from  $a_x(t)$ ; the moving average of the CWT coefficients of this window, defined as  $C_{Wma}(t)$ , was estimated. The tailored threshold for the subject was defined as

$$T_{tailored} = \bar{C} - \sigma(C) \quad (5)$$

in which  $\bar{C}$  is the average and  $\sigma(C)$  the standard deviation over time of  $C_{Wma}(t)$ . Table 5 compares the three approaches in terms of walking samples included in RIs with respect to the total walking samples (*walk percentage*) and an error term, defined as the portion of non relevant activities included in RIs (label 0, stand) divided by the total duration of RIs (*RIE error*). As expected the tailored-threshold allows to recognize all walking segments, even though producing a slightly larger error. However, in order to generalize as much as possible the algorithm, the hysteresis approach was selected.

Table 5: Comparison of thresholding methods for RIE

Method	Walk percentage [%]	RIE Error [%]
Single fixed threshold	93	20
Hysteresis	98	23
Patient-tailored threshold	100	25

Finally, the RIs are enlarged at their borders of a length  $l$ : this is done because a FoG event could verify not only while walking or turning, but also at the beginning or end of such activities, hence at the limits of a RI. For  $l$  an empirical “supervised” approach was used: considering for all Phase 1 subjects, all the instants of the signals labeled as FoG but not included in a RI,  $l$  was defined as

the average distance between these time instants and the closest RI. This value was found to be  $l = 1.75$  s. Figure 16 shows the RIs on  $a_x(t)$ , extracted with the proposed technique, for four different patients.

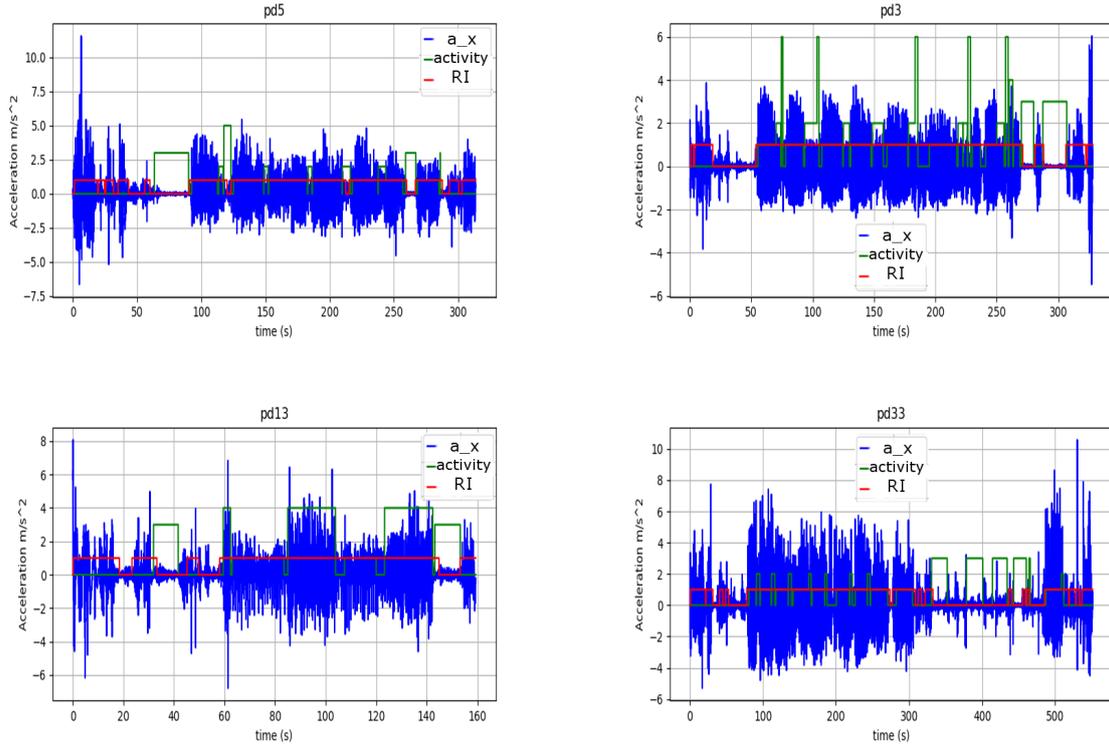


Figure 16: Regions of interest extracted from four different patients

### 3.3.2 Windowing

Only on the extracted RIs, windowing was applied. Windowing consists in sliding a time window of fixed size  $w$  over a continuous signal, to partition it into smaller segments. This process is done in order to obtain local information (features) from each segment, to evaluate where in time a FoG event is verifying. The window slides over the signal with a stride  $s$ , which can be chosen in the range  $0 < s \leq w$ : larger its value, smaller will be the overlap between consecutive windows. Some overlap is necessary to avoid loss of information [85]. As already mentioned in section 2.5, several window sizes and strides have been explored in the literature, with values normally ranging from 1 to 5 s and 50% or lower overlap. Window size is inversely proportional to the time-resolution of the recognition system [15], meaning that a smaller window will better identify short episodes, even though if too small, it could badly affect specificity [70]. Moreover, if the application is a real time cueing system, its latency<sup>6</sup> depends on the window size, thus this should be as short as possible. The final choice in this study was  $w = 2$  s and  $s = 1$  s, which provided good results when applied in [15] on the same data.

<sup>6</sup>time delay between an event and its recognition

### 3.3.3 Features extraction

From each time window, a set of temporal and spectral numerical features, related to FoG recognition, were extracted. A number of options are available in the literature, Table 6 and Table 7 report respectively the temporal and spectral features that were selected. Freezing indexes are reported in spectral features because they are defined from spectral properties of the signals. They were extracted for each acceleration component separately, for a total of 48 distinct values: at this point of the pipeline, each patient window can be regarded as a point in 48-dimensional space.

Table 6: List of temporal features extracted from each window.

Feature Name	Source
Mean	[85]
Standard Deviation	[85]
Variance	[81]
Root Mean Square	[111]
Zero crossing rate	[15]
Number of peaks	[15]
Average Absolute Variation	[22]
Correlation between axes	[85]

Table 7: List of spectral features extracted from each window.

Feature Name	Source
Freezing index	[66]
Kurtosis	[85]
Skewness	[85]
Dominant frequency	[15]
Freezing ratio	[15]
CWT Freezing ratio	[83]
0.5Hz average CWT	*
Total Power	[90]

As it can be observed, the list includes both standard temporal and spectral features of signal processing (e.g. mean, variance, kurtosis, dominant frequency, total power), but also ad hoc FoG features, as the freezing indexes coming from studies in which a thresholding solution for FoG recognition was employed (*freezing index* and *CWT freezing ratio*). About *CWT freezing ratio*, it was defined in [83] and was included also to exploit the CWT coefficients already computed for the RIE. The feature denoted by \* in the column “Source” was added in the current study, observing that the average value of the coefficient corresponding to frequency 0.5 Hz is often small during FoG.

In order to train and test supervised machine learning models, a label for each window had also to be defined together with the features. Each window was classified as the most frequent activity it contains, but for the case of FoG: a window is labeled as FoG only if it contains at least 60% of time samples that were classified so by clinicians. This is done to avoid windows with half FoG and half walking or turning, whose mixed characteristics would affect in a negative manner both the subsequent feature selection and classifier training. Table 8 reports the mapping between class labels and activities. Activity marked as 6 in Phase 1 data (“short hesitation or FoG”), was classified as FoG as well, to avoid the loss of brief episodes due to uncertainty, considering also that few freezing windows were available. Windows containing activity “0” were discarded because, as previously mentioned, they actually contain a mix of activities, hence are not considered relevant for the analysis and could even confuse the classifiers during training.

Table 8: Mapping between activities and class labels

Activity	Class Label
Walk	1
Turn	2
Stand	3
FoG	4

### 3.3.4 Training set creation & feature selection

Table 9 illustrates the number of windows per activity that were obtained processing all patients from Phase 1. As it can be observed, FoG recognition is an unbalanced problem: the number of FoG windows with respect to the most frequent windows (*walk* and *turn*) is almost two order of magnitude smaller. This condition could badly affect the classifiers, making their predictions biased toward the most frequent classes and, in addition to this, using all these data for several models would require a lot of computational effort. To cope with this issue, **undersampling** of the two most frequent classes was performed to obtain a reduced training set, a procedure that is quite common and well established [20] [15]. To preserve the statistical variability of the data, undersampling was applied selecting randomly 9 windows (or the maximum available, if smaller than 9) for each of *walk* or *turn*, for each patient. All FoG and stand windows, instead, were included. Table 10 reports the obtained reduced dataset. About stand, the small number of windows was not considered an issue for two reasons:

- the goal of the final classifiers was the correct recognition of FoG, not of the other activities involved;
- the reduction of stand is an effect of RIE, hence stand was removed because considered irrelevant (or even harmful) for FoG recognition.

Once the reduced dataset was defined, it was employed to perform feature selection.

Table 9: Windows extracted from Phase 1

Activity	N. of windows
Walk	4686
Turn	1267
Stand	29
FoG	88

Table 10: Windows in Phase 1 reduced dataset

Activity	N. of windows
Walk	342
Turn	311
Stand	29
FoG	88

**Feature selection** consists in identifying among the extracted features the more relevant to the problem [41], in order to:

- reduce the dimensionality of the input data;

- decrease the computational effort in training;
- remove attributes of the inputs that could be irrelevant or even misleading;
- reduce the probability of overfitting the training data.

For FoG recognition, different approaches are used; they are often filter methods<sup>7</sup> based either on Pearson’s correlation [15] [85] or Mutual Information (MI) [111] of each feature  $f_i$  to the target class label  $l$ . In this study, Mutual Information was selected. Considering both  $f_i$  and  $l$  as two random variables, their MI is defined as

$$MI(f_i; l) = H(f_i) - H(f_i|l) \quad (6)$$

in which  $H(f_i)$  and  $H(f_i|l)$  are respectively entropy of feature  $f_i$  and its conditional entropy when observing class label  $l$ . This metric was chosen because it is capable of catching any kind of statistical dependence, not only linear, even though it requires as much sample as possible to be accurate (it has to estimate probability density functions) [93]. About this aspect, a relevant issue is related to when perform feature selection with respect to train&test split<sup>8</sup> of the dataset. On the one hand, performing feature selection on the whole data would produce more accurate results, considering also that MI works better with additional samples (and FoG samples are limited). Moreover, the set of the most significant features would very likely change considering different splits. On the other hand, this approach is generally not recommended, because it would exploit also test data, that should be completely unseen by the classifier prior testing, to provide a statistically relevant evaluation. To reach a compromise between these two sides, a “crossvalidation-inspired” method was defined as follows:

1. perform  $k = 10$  different train&test splits (random, stratified, without replacement) of the reduced dataset with a 7:3 proportion;
2. for each obtained training set  $T_k$ , compute  $MI(f_i; l)$  for  $f_i \in F$  ( $F$  the set of all features) and add the 15-top-scoring features to set  $S$ ;
3. for each the feature  $f_j \in S$ , compute a relevance metric  $R$  defined as

$$R = n_j MI_{avg}(f_j; l) \quad (7)$$

in which  $n_j$  is the number of time the feature was included in  $S$  among the different splits and  $MI_{avg}(f_j; l)$  is the average mutual information that  $f_j$  scored in the splits where it was included in  $S$ ;

4. select as optimal, features with  $R \geq 1.5$

Figure 17 shows a plot of metric  $R$  versus the features that were included in  $S$ . The 1.5 threshold was chosen because it is more or less the mean value between the maximum and the minimum assumed by  $R$  in the experiment.

---

<sup>7</sup>selection methods that do not consider the final classifier [41]

<sup>8</sup>procedure of preprocessing in which the available data are divided in two sets, one for training a model and the other one for testing it

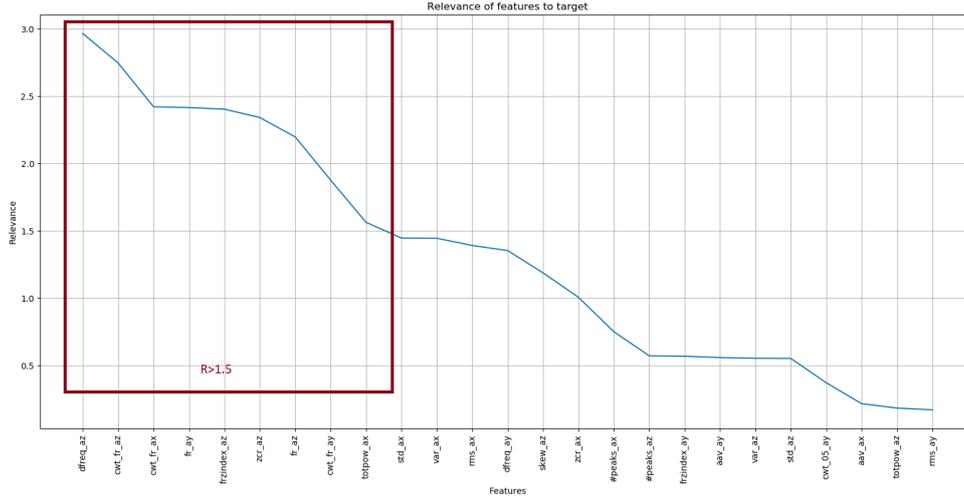


Figure 17: Relevance  $R$  vs features in optimal set  $S$ . In the red box, the features scoring more than half the maximum value observed.

Table 11: Feature set after selection

Feature	Relevance $R$
Dominant frequency (z)	2.96
CWT freezing ratio (z)	2.74
CWT freezing ratio (x)	2.45
Freezing ratio (y)	2.43
Freezing index (z)	2.40
Zero crossing rate (z)	2.38
Freezing ratio (z)	2.23
CWT freezing ratio (y)	1.82
Total Power (x)	1.58

Table 11 contains the final 9 features that were selected. The selection seems to highlight the importance of spectral features and freezing indices with respect to temporal attributes, in particular from  $a_x(t)$  and  $a_z(t)$ ; it is also worth noticing that the introduced *CWT freezing ratio* seems to be a quite relevant discriminating parameter for FoG recognition.

At these points of the algorithm, each window can be represented as a vector  $\mathbf{x} \in R^F$ , with  $F = 9$  and a label  $y \in \{1, 2, 3, 4\}$ ; the whole reduced dataset can be written as a matrix  $\mathbf{X} \in R^{P,F}$ , with  $P = 770$  (total number of windows) and a vector of labels  $\mathbf{y} \in R^P$ . Hence, Matrix  $\mathbf{X}$  was divided into a training matrix  $\mathbf{X}_{tr}$  and a test matrix  $\mathbf{X}_{te}$ , using a stratified train&test split with 70% of the data in training and the remaining 30% left out for test. The two matrices were standardized, meaning that each row vector  $\mathbf{x} \in X_{tr}, X_{te}$  was transformed in a vector  $\mathbf{z}$  such that

$$\mathbf{z} = \frac{\mathbf{x} - \bar{\mathbf{x}}_{tr}}{\sigma_{tr}} \quad (8)$$

in which  $\bar{\mathbf{x}}_{tr}$  and  $\sigma_{tr}$  are respectively the vectors containing the means and the standard deviations of the features, evaluated from  $\mathbf{X}_{tr}$  only. This operation, also known as “Z-Score” standardization, was done because it ensures that all features

are in the same scale. This aspect is relevant when evaluating distance metrics between data points, for example in Principal Component Analysis and in classifiers like K-NN and SVM, that will be later employed. Moreover, standardization is considered a good practice in machine learning, which often improves the training phase [51].

**Principal Component Analysis (PCA)** was also considered to explore if it was possible to further condense the information coming from the selected features. PCA is an unsupervised<sup>9</sup> technique that is used to perform dimensionality reduction and features decorrelation. [51] It consists in finding a projection of the data in a lower dimensional space, preserving most of their variance (which is related to information content): indeed the data are projected with respect to the directions associated to the maximum variation. These directions are called the **principal components**. To apply PCA, the following operations over  $\mathbf{X}_{tr}$  (after standardization, which is a requirement) must be performed:

- compute the sampled covariance matrix

$$\Sigma = \mathbf{X}_{tr}^T \mathbf{X}_{tr} \quad (9)$$

$\Sigma$  is a symmetric  $F \times F$  matrix hence it can be eigendecomposed;

- compute the eigenvalues  $\lambda_i$  and eigenvectors  $\mathbf{u}_i$  of the sampled covariance matrix, sort them in descending order according to the eigenvalues magnitude. The eigenvalues represents a measure of the variance explained by the associated direction;
- select the first K eigenvectors such that

$$\sum_{i=1}^K \lambda_i \geq \beta \sum_{k=1}^F \lambda_k \quad 0 < \beta < 1 \quad (10)$$

meaning that only a fraction of the total variance, decided by parameter  $\beta$ , will be maintained;

- the selected eigenvectors form a basis of the space where the data will be projected. Collect them in a matrix  $\mathbf{W}$  and compute the projection of the original data with respect to this new space as  $\tilde{\mathbf{X}}_{tr} = \mathbf{X}_{tr} \mathbf{W}$ .

Figure 18 shows how the variance is distributed among the identified principal components. As it can be appreciated, the first component already accounts for most of the variability (almost 60%), which could be explained by the fact that several of the selected metrics are correlated to similar spectral properties of the data; however, to preserve around 90% of the original variance, the first 4 components were selected.

---

<sup>9</sup>statistical method that does not exploit the information about class label of the data

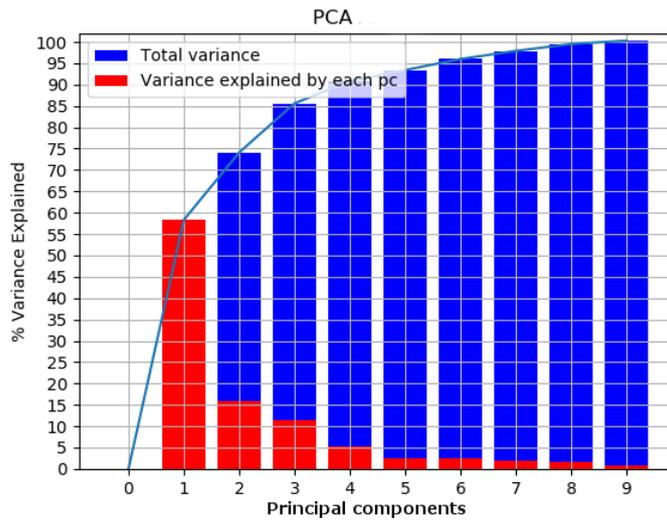
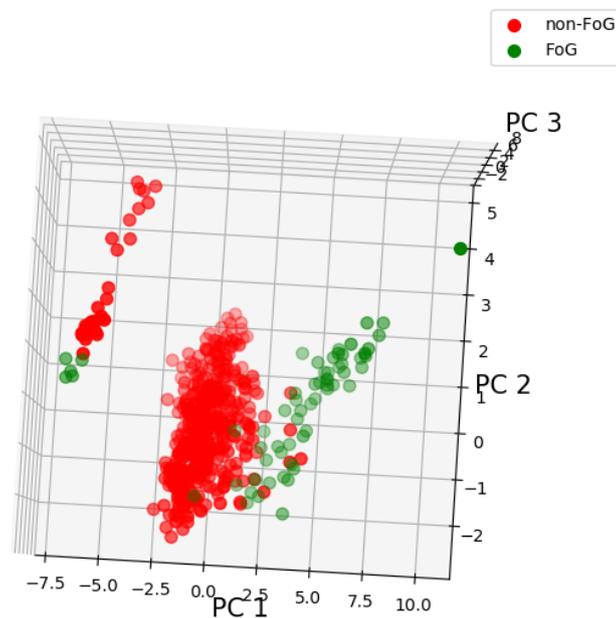


Figure 18: Distribution of training data variance among the computed principal components. In blue the cumulative variance, in red the variance explained by each component.

Figure 19 represents how the training windows, divided in FoG vs non-FoG, are scattered with respect to the first 3 principal components, the maximum dimension that can be used to observe them in a plot. It is evident that after the remapping some separation between the two classes is possible.



Data with respect to first 3 PC

Figure 19: Scatter plot of training data with respect to principal components

### 3.3.5 Machine learning models

Three classification models were compared:

- K-Nearest Neighbours (KNN)
- Support vector machine (SVM)
- Random Forest (RF)

KNN and SVM are two shallow supervised approaches that are common in the literature of FoG recognition, the latter in particular [15] [3] [85]. RF is instead an ensemble method based on a set of simpler decision trees, which are often employed for classification problems in the medical field [65] [15]. For all the models, both a binary and a multiclass version was implemented: in the binary case, windows were labeled either as FoG (label 1) or non-FoG (label 0), which includes *walk*, *turn* and *stand* windows. In the multiclass case, the classifiers assigned the labels in Table 8. While KNN and RF are suitable for both types of classification, in the case of SVM, which is designed for binary classification only, a One-vs-One approach was used: as many SVMs as all possible pairing of classes were trained and the final classification results were obtained by combining the outcomes of each single model. Moreover, both binary and multiclass models were trained once on the 9 features selected using MI and once on the projection of the data with respect to their principal components, to perform a comparison between these two feature sets.

**KNN** is the easiest machine learning algorithm for classification. It is based on assigning to an unlabelled data object the same label of the closest  $K$  objects.  $K$  is an hyperparameter<sup>10</sup> to tune; it is good practice to choose an odd value: when  $K > 1$ , the most common strategy to select the correct label is majority voting, hence such choice prevents ambiguities. To define the K-closest objects, different concepts of distance can be employed, in this study the following were considered ( $\mathbf{x}, \mathbf{y}$  being two vectors in  $R^{F,1}$ )

- Euclidean:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i^F (x_i - y_i)^2} \quad (11)$$

- Manhattan:

$$d(\mathbf{x}, \mathbf{y}) = \sum_i^F |x_i - y_i| \quad (12)$$

- Chebyshev:

$$d(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i| \quad (13)$$

---

<sup>10</sup>parameter of a classifier that cannot be learned but must be defined by the user

KNN can identify non linear decision boundaries, as shown in Figure 20, and it actually does not require any proper training phase, because as soon as a new unlabeled point is available, a prediction can be performed. However, this requires to store all the training points (memory occupation could be a problem for limited memory devices) and test time can be very long if too many distances have to be computed. Moreover, it may sometimes overfit<sup>11</sup> the training data.

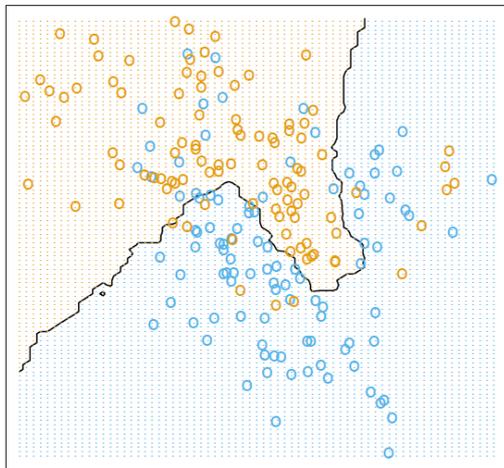


Figure 20: Example of KNN algorithm. Ref: [51]

**SVM** is a common choice for FoG recognition, achieving often accuracy, precision and sensitivity above 90% [15] [3]. In its most strict definition, known as “hard margin”, SVM consists in finding the **hyperplane** that separates two classes of points (labels -1, +1) with the largest **margin**, where the margin is defined as the distance between such hyperplane and the closest points from the two classes (figure 21). Such points are known as the **support vectors** and they are the only training points needed to define the decision boundary. Indeed, solving the optimization problem associated to SVM formulation, the separating hyperplane can be written as

$$f(\mathbf{x}) = \sum_{i=1}^P \alpha_i \mathbf{x}^T \mathbf{x}_i \quad (14)$$

in which  $P$  is the number of training samples,  $\mathbf{x}_i$  is a training sample and  $\alpha_i$  its coefficient obtained from the optimization problem. Support vectors are the training samples for which  $\alpha_i \neq 0$ . Classification consists simply in replacing in equation 14 an unlabeled point: when the result is  $> 0$  the point belongs to the positive class, otherwise to the negative one. If the classes are not perfectly linearly separable, however, the algorithm does not reach convergence and also when converging it may easily overfit. To relax the condition on margin, a “soft definition” exists, which includes in the optimization problem “slack” variables  $\xi$  weighted by an hyperparameter  $C$ , sometimes called the “cost of misclassification”: smaller

<sup>11</sup>condition in which a classifier has a very low bias, but an high variance, hence it is not general enough to adapt to test data

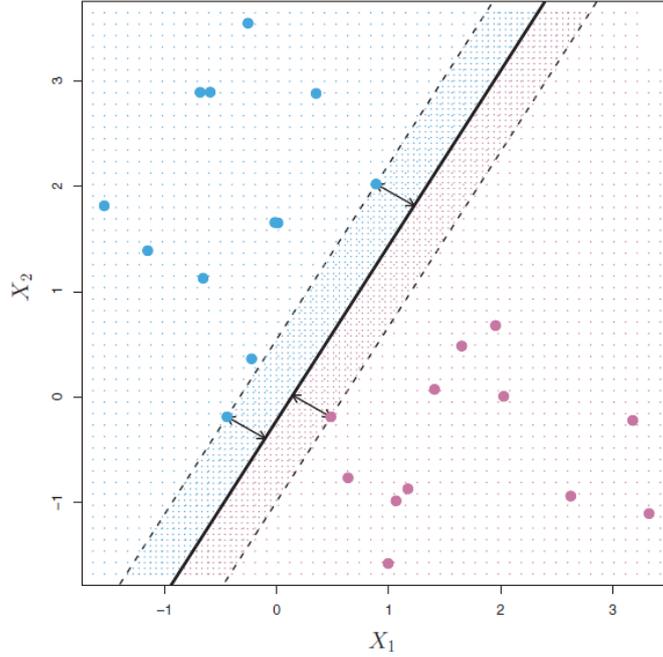


Figure 21: A representation of hard margins SVM. Ref: [51]

its value, more samples will be allowed to fall in the margin or even on the wrong side of the hyperplane. Further improvement to the performances of the classifier can be achieved using a **kernel trick**, to transform the linear decision boundary in a nonlinear one. The idea is to map training points in an higher dimensional space where they are more linearly separable, through a mapping function  $\phi(x)$ . Such function is theoretically complex to identify, however in practice it is sufficient to identify a kernel function  $K(x, x)$  such that

$$K(x, x) = \phi(\mathbf{x})^T \phi(\mathbf{x}) \quad (15)$$

and replace with it the scalar product of equation 14, to get

$$f(\mathbf{x}) = \sum_{i=1}^P \alpha_i K(\mathbf{x}, \mathbf{x}_i) \quad (16)$$

which corresponds to a nonlinear decision boundary in the original space of training data. Common kernels are the polynomial kernel and the Radial Basis Function (RBF) kernel; both were considered in this study.

**Random Forest** is an **ensemble technique** that combines predictions made by several **decision trees**. A decision tree is a machine learning model that is constructed partitioning the feature space  $R^F$  in  $J$  decision regions, which are usually hyperdimensional rectangles. When a data object falls in one of the regions, it is classified according to which class the majority of the samples in that region belong to. The partition is performed considering once at a time each feature and performing a binary (also multinomial is possible) split over the range of its values. Once all the attributes have been considered or another stopping

condition is met, the algorithm stops and the tree is complete. The attributes for splitting are selected using a greedy approach based on some “relevance” criterion: a common choice is considering first the features that provide the largest decrease in “impurity” in the resulting regions, being impurity defined as a measure of the heterogeneity of the class labels. A region with samples belonging all to the same class is completely pure, hence no more splits over that region are required. RF improves simple decision tree through **boosting** and **feature sampling**. In boosting, many smaller training subsets are defined and each one is used to build a simple decision tree. A data object to be classified is analysed by each tree and the final outcome is chosen through majority voting. This mechanism is further improved using feature sampling: each time a tree is grown, for each split to perform, only a subset of  $m$  features, among all the available, is considered. Generally  $m = \sqrt{F}$ . The main hyperparameter to tune for RF is the number of basic learner to train.

### 3.4 Validation & Test

Considering binary classification and FoG as *positive* class, the prediction provided by a classifier on a single window can be:

- **True Positive (TP)**, the data object was correctly recognized as FoG, this is also known as a *hit*;
- **False Positive (FP)**, the data object was wrongly recognized as FoG, this is also known as *type I error*;
- **True Negative (TN)**, the data object was correctly recognized as non-FoG, this is also known as *correct rejection*;
- **False Negative (FN)**, the data object was wrongly recognized as non-FoG, this is also known as *type II error*;

The distribution of such results is often reported in the form of a matrix, called the *confusion matrix*. From it, a number of metrics for evaluating the quality of a classifier can be defined. The most common is **accuracy**, which is defined as

$$accuracy = \frac{TP + TN}{FP + FN + TP + TN} \quad (17)$$

and provides an overall evaluation of the classifier, without an insight of the performances in the single classes to recognize. Because of this lack of detailed information, it is not meaningful when dealing with an unbalanced dataset: if the model simply classifies all points with the most frequent class label, the value of accuracy will be high, even though the less frequent class samples are all wrongly labeled. Being the problem under study an unbalanced problem, other per-class-specific metrics were employed together with accuracy:

- **Precision** or *Positive Predicted Value*, which is the number of correctly classified positive examples divided by the number of examples labeled by the system as positive [97]

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

- **Recall** or *sensitivity*, which is the number of correctly classified positive examples divided by the number of positive examples in the data [97]

$$Recall = \frac{TP}{TP + FN} \quad (19)$$

- **F<sub>1</sub> – score** or *F-measure*, which is the harmonic mean between precision and recall [103]

$$F_1 - score = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (20)$$

- **Specificity** or *True Negative Rate*, which is a measure complementary to precision and represents the proportion of the negative samples that were correctly classified [103]

$$Specificity = \frac{TN}{TN + FP} \quad (21)$$

As discussed in the previous section, during training, the considered machine learning models require hyperparameters tuning. This procedure was performed through grid search.

**Grid search** is common practice when dealing with few hyperparameters because it is performed training a model for every joint values in the Cartesian product of the set of values for each individual hyperparameter. The experiment that yields the best validation score provides the optimal hyperparameter set [49]. The values to explore, provided for each hyperparameter, should lie in a reasonable range, whose cardinality will affect the duration of the search. Too many hyperparameters or too many values to explore would make such kind of search unfeasible in time [49]. In this study, grid search was combined with 5-fold cross-validation, so that the validation score associated to each experiment was as less as possible data-dependent. Moreover, the *F<sub>1</sub> – Score* of class FoG was employed as optimization metric, as it is more meaningful with respect to accuracy when dealing with unbalanced datasets.

For KNN the following hyperparameters were considered:

- distance (euclidean, manhattan, chebyshev)
- value of *k* (range [1,23] with a step of 2)

For SVM:

- kernel (RBF, polynomial)
- Cost of misclassification *C* (logarithmic range between 10<sup>-3</sup> to 10<sup>3</sup>)
- kernel related parameters
  - for RBF kernel, parameter  $\gamma$  (logarithmic range between 10<sup>-3</sup> to 10<sup>3</sup>)

- for polynomial kernel, degree of the polynomial (range [2,10] with a step of 1)

For RF:

- number of estimators to train (range [1,200] with a step of 20)
- maximum number of features to consider during a split ( $\sqrt{F}$  or  $\log_2 F$ )

As previously described, the reduced data set was divided into a training matrix  $\mathbf{X}_{tr}$  and a test matrix  $\mathbf{X}_{te}$ . To validate the optimal models found through grid search, stratified k-fold crossvalidation was performed on  $\mathbf{X}_{tr}$ .

**K-fold Crossvalidation** consists in randomly dividing the set of observations into  $K$  groups, or folds, of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining  $K - 1$  folds [51]. The “stratified” attribute implies that the different folds contain the same distribution of classes present in the complete dataset. The chosen metrics are evaluated on each trained model and they are averaged to provide an estimate of the performances that should be less affected by the variance of the data. In this study  $K$  was chosen equal to 10, which is a recommended value in the literature [51].

The results coming from validation were also compared to the the one obtained by applying the optimal models to the samples in  $\mathbf{X}_{te}$ . Due to the reduced dimension of the dataset, testing outcomes could be influenced by a large variability (reason why validation was performed), nevertheless it was considered relevant to provide also the results obtained from data that were never seen by the models. In addition to this, Phase 2 data were employed for testing. First of all, the initial and the final 40 s of each the acceleration record were removed, to discard recordings related to the montage of the smartphone or to the explanation of the experiment to patients. Then, they were used as follows:

1. all the Phase 2 patients were used to tune the RIE for this new type of data (new double thresholds were selected);
2. only data from freezer patients were segmented and added to Phase 1 data, discarding windows related to physical tests (label 7 and 8), which are not considered common of daily living. Activities labeled as -2, 5, 6 and some piece of the “undefined” activity 0, which could occur in daily living, were maintained to perform an analysis closer to an “at home” evaluation. However, they were all mapped, for the multiclass recognition task, to a common class label 0, because they are in a small number after segmentation and are absent in Phase 1 data;
3. a new reduced dataset from this mixture of Phase 1 and 2 was defined;
4. feature selection was performed again exploiting the additional data;
5. models were trained and validated, as previously discussed, on 70% of the new reduced dataset, whereas again a 30% was left out for testing;

6. a false positive test was performed on non-freezer patients of Phase 2 using the retrained models.

The new reduced dataset is described in Table 12.

Table 12: Reduced dataset from Phase 1 and Phase 2

<b>Activity</b>	<b>N. of windows</b>
Mix of -2,0,5,6 (0)	150
Walk (1)	504
Turn (2)	451
Stand (3)	42
FoG (4)	277

As it can be observed, mixing the data from the two phases, the number of FoG windows triplicates. This provides a better insight in the characteristics of freezing event; for this reason on the reduced dataset, the same previously described procedure for feature selection was performed and 10 optimal features were identified (Table 13). Instead, always the first four principal components were retained, because they still contained 90% of the explained variance.

Table 13: Optimal features from feature selection

<b>Feature</b>	<b>Relevance R</b>
Freezing index ( $z$ )	3.38
Dominant frequency ( $z$ )	3.26
CWT freezing ratio ( $z$ )	2.81
Freezing ratio ( $z$ )	2.68
CWT freezing ratio ( $x$ )	2.23
Root mean square ( $x$ )	2.17
Total power ( $x$ )	2.16
Variance ( $x$ )	2.16
Standard deviation ( $x$ )	2.15
Zero crossing rate ( $z$ )	1.64

From Table 13 it can be observed that most of the features were confirmed as relevant also on the increased dataset (*Freezing index ( $z$ )*, *Dominant frequency ( $z$ )*, *CWT freezing ratio ( $z$ )*, *Freezing ratio ( $z$ )*, *CWT freezing ratio ( $x$ )*, *Total power ( $x$ )*, *Zero crossing rate ( $z$ )*) whereas few temporal features were added (*Root mean square ( $x$ )*, *Variance ( $x$ )*, *Standard Deviation ( $x$ )*). Still spectral features and freezing indices/ratios seem to be the most relevant characteristics to distinguish FoG.

About the false positive test on Phase 2 non-freezers, it was conducted in this way:

1. for each non freezer patient, the testing pipeline in Figure 11 was applied, once with the RIE on and once with the RIE off;
2. a comparison between the performances in terms of number of False Positive and specificity with and without extraction of RIs was performed.

## 4 Results and discussion

The pipeline described in Chapter 3 was implemented in Python 3.7 using Scipy library for signal processing and linear algebra, Pywavelet library for CWT and Scikit-Learn library for data processing and machine learning models training and test. The code was executed on a commercial HP laptop with Intel i7 7<sup>th</sup> generation core and 8 GB of RAM memory while no other significant resource-consuming application was running.

First, general results about the Regions of Interest Extractor and its effects on performances in terms of execution time and included/excluded activities are discussed. Then, validation and test results are organised per phase.

### 4.1 Statistics on RIE

As stated in Chapter 3, one of the goal of introducing a RIE in the traditional pipeline for FoG recognition was to improve time efficiency of the system: in the perspective of a daily living usage, it would be required to process very long signals, with likely only few portions of the recordings that are relevant. Table 14 reports some statistics related to average execution time of RIE for one patient and its effects on the subsequent steps of windowing and feature extraction on data from the two phases.

Table 14: Execution time statistics.

Execution time	Phase 1		Phase 2	
Regions of Interest Extractor	65 ms		84 ms	
	<b>RIE off</b>	<b>RIE on</b>	<b>RIE off</b>	<b>RIE on</b>
Windowing & feature extraction	8.05 s	5.65 s	12.27 s	2.96 s

As it can be appreciated, the average execution time of the RIE is in the order of ms for processing the whole 6 minutes coming from 6MWT in Phase 1. For Phase 2, the slight increase is justified by the fact that, as already mentioned, this data are generally longer (some recordings are over 10 minutes). In Phase 1, the time improvement in windowing and feature extraction is of about 2 s and a half; much bigger is the improvement in Phase 2, with a reduction of almost 10 s. This is explained by the fact that Phase 1 data have an higher “concentration” in terms of significant activities performed (*walk* or *turn*) hence more pieces have to be maintained and processed. In Phase 2 recordings, there are longer intervals of inactivity or non-relevant activities which are removed by the RIE, reducing the load for windowing; this characteristic of Phase 2 data make them more similar to daily living recordings, where it is expected that a patient will not walk or turn all the time. Thus, it is reasonable to expect that also on these kind of data similar improvements could be achieved. Tables 15 and 16 contain a summary of the activities that were removed and maintained in the two phases using the RIE.

Table 15: Percentages of activities selected by RIE in Phase 1

activity	tot (s)	removed (s)	removed %	maintained %
undefined (0)	3934.79	1301.07	33.07%	66.93%
walk (1)	4624.68	9.86	0.21%	99.79%
turn (2)	1225.91	19.47	1.59%	98.41%
stand (3)	1041.77	972.99	93.40%	6.60%
fog (4)	77.83	2.93	3.76%	96.24%
hesitation (6)	37.86	11.85	31.30%	68.70%

Table 16: Percentages of activities selected by RIE in Phase 2

activity	tot [s]	removed [s]	removed [%]	maintained [%]
daily living mix (-2)	1221.38	588.50	48.18%	51.82%
undefined (0)	15660.28	12544.29	80.10%	19.90%
walk (1)	997.32	4.11	0.41%	99.59%
turn (2)	571.98	21.75	3.80%	96.20%
stand (3)	1474.00	1411.51	95.76%	4.24%
fog (4)	215.61	7.41	3.44%	96.56%
stand up (5)	250.25	51.41	20.54%	79.46%
sit down (6)	461.76	135.01	29.24%	70.76%
retropulsion test (7)	456.93	24.12	5.28%	94.72%
tapping (8)	595.93	528.74	88.73%	11.27%

In both phases, almost all FoG episodes are retained by the RIE, only less than 4% of seconds classified as proper FoG by clinicians are discarded. In Phase 1, 30% of activity 6 is lost but, as already mentioned, it was not possible to understand if this activity corresponded to FoG or to a simple hesitation, hence the removed part could also be irrelevant or even confusing for FoG recognition. About *walk* and *turn*, they are maintained with a very high percentage, as desired when designing the RIE. About the other activities, more than 90% of *stand* in both phases is removed, again as desired. In Phase 1 some pieces marked as 0 are removed, but a 66.90% is kept: this could be explained by the fact that this undefined activity often contains inside pieces where the patient was walking or performing an active task. In Phase 2, instead, it was much largely reduced (80.10%). Tapping (8) was also largely reduced, with respect to the retropulsion test (7) which was instead retained for the most; however these two activities are not relevant because they do not correspond to actions that a patient would normally perform in daily living. About daily living, the mix of activities marked as -2 was halved by the RIE; less effective is the reduction of *stand up* and *sit down* which are both maintained in more than 70% of the cases. Overall it is evident that selection of RIs reduces a lot the data to process in Phase 2 with respect to Phase 1, thus the great improvement in execution time of windowing for this phase.

## 4.2 Results on Phase 1 data

The results of validation and test of models trained on Phase 1 data only, as described in section 3.4, are organised in four tables: Table 17 with results from multiclass models trained on the 9 features from feature selection; Table 18 with results from multiclass models trained on the first 4 principal components of the training data; Table 19 with results from binary models trained on the 9 features from feature selection; Table 20 with results from binary models trained on the first 4 principal components of the training data. Precision, recall and F-score refers to class FoG only, which is the relevant to our analysis.

Table 17: Results of validation & test of multiclass classifiers trained on selected features

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	70.0	94.0	92.0	92.0	69.0	82.1	88.5	85.2
SVM (RBF)	75.0	93.0	93.0	93.0	73.0	88.5	88.5	88.5
random forest	74.0	89.0	85.0	86.0	75.0	77.4	92.3	84.2

Table 18: Results of validation & test of multiclass classifiers trained on principal components

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	70.0	94.0	92.0	92.0	69.0	81.5	84.6	83.0
SVM (RBF)	73.0	93.0	92.0	92.0	74.0	79.3	88.5	83.6
random forest	68.0	89.0	89.0	88.0	70.0	84.6	84.6	84.6

Table 19: Results of validation & test of binary classifiers trained on selected features

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	98.0	92.0	90.0	91.0	99.0	100	88.5	93.9
SVM (RBF)	98.0	94.0	87.0	90.0	97.0	95.4	80.7	87.5
random forest	97.0	90.0	84.0	86.0	98.0	95.6	84.6	89.7

Table 20: Results of validation & test of binary classifiers trained on principal components

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	97.0	89.0	89.0	88.0	98.0	95.6	84.6	89.8
SVM (RBF)	97.0	92.0	85.0	88.0	98.0	95.6	84.6	89.8
random forest	97.0	90.0	82.0	85.0	97.0	91.6	84.6	88.0

First of all, a brief comment on accuracy. In multiclass, accuracy is very low: this is due to the fact that the selected features are relevant for FoG recognition, which is proved by the good results in terms of precision and recall, but not for distinguishing among the other classes. However, the goal was to classified correctly class FoG, not all the activities, hence it is not considered an issues. Indeed in binary classification, where only a distinction between FoG and non-FoG is required, also values of accuracy are high. Overall, multiclass models seem to perform slightly better in validation, but during test their performances exhibit a drop of 5 percentage points, which suggest some degree of overfitting. This effect

is less evident in binary models, even though it should always be considered that the test set is quite small so test results could be affected by an high variability. Some test results are better than validation: this is explained by the fact that a single train&test split was considered, hence this could produce an optimal test subset for that specific model and that metric. However, as already mentioned, validation is the more statistically significant evaluation and should be considered first with respect to test. Precision, which was one of the goal to achieve, is quite high in all models, with acceptable values of recall as well. The usage of principal components provide slightly worse results, but it should also be considered that the whole problem is reduced to 4 features only, from the original 48 extracted. Both KNN and SVM with RBF kernel provide comparable results and outperform Random Forest.

## 4.3 Results on Phase 2 data

### 4.3.1 Validation and test results of retrained models

The results of validation and test of models trained on Phase 1 and 2 are organised in four tables: Table 21 with results from multiclass models trained on the 9 features from feature selection; Table 22 with results from multiclass models trained on the first 4 principal components of the training data; Table 23 with results from binary models trained on the 9 features from feature selection; Table 24 with results from binary models trained on the first 4 principal components of the training data. Precision, recall and F-score refer again only to class FoG.

Table 21: Results of validation & test of multiclass classifiers (Phase 1+2) trained on selected features

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	68.0	85.0	86.0	85.0	62.0	82.0	87.9	84.9
SVM (RBF)	70.0	87.0	89.0	87.0	66.0	88.5	92.8	90.6
random forest	67.0	82.0	88.0	84.0	67.0	84.0	95.2	89.3

Table 22: Results of validation & test of multiclass classifiers (Phase 1+2) trained on principal components

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	67.0	84.0	83.0	83.0	58.0	83.9	81.9	82.9
SVM (RBF)	67.0	84.0	87.0	85.0	64.0	85.8	87.9	86.9
random forest	66.0	84.0	86.0	84.0	60.0	82.9	87.9	85.3

Table 23: Results of validation & test of binary classifiers (Phase 1+2) trained on selected features

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	94.0	88.0	83.0	85.0	94.0	86.2	83.1	84.6
SVM (RBF)	95.0	90.0	84.0	86.0	94.0	86.4	84.3	85.3
random forest	93.0	86.0	80.0	82.0	95.0	88.9	86.7	87.8

Table 24: Results of validation & test of binary classifiers (Phase 1+2) trained on principal components

Classifier	Validation				Test			
	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]	Accuracy [%]	Precision [%]	Recall [%]	F-score [%]
KNN	94.0	88.0	80.0	83.0	93.0	83.9	81.9	82.9
SVM (rbf)	94.0	89.0	79.0	83.0	93.0	84.4	78.3	81.2
random forest	93.0	85.0	79.0	81.0	95.0	87.0	89.1	88.1

Also in this case, the same discussion done in the previous section about accuracy in the multiclass models is valid. Results in validation are slightly worse with respect to Phase 1 only, but two important considerations must be taken into account:

- A bigger dataset was used for training, hence these results are more statistically relevant because they involve more variability; indeed the gap between the results in validation and in test is reduced with respect to Phase 1 only;
- The additional activities (-2,0,5,6) that are provided as a single label 0 to models can be mistaken for FoG more easily with respect to the activities provided in Phase 1; however they are deemed important because they would certainly appear in a usage of the algorithm in daily living conditions.

Nevertheless, all the metrics are above 80% and often closer to 90%. Overall, binary models perform better in terms of validation precision, which was considered as a relevant metric in this study. Binary SVM with RBF kernel is the model that performs better in terms of precision, with 90% score. In this case the gap of KNN and SVM versus RF is reduced, more or less the three classifiers have similar performances. Models trained on principal components provide slightly smaller figures, but again with the advantage of considering only 4 features.

### 4.3.2 False positive test results

False positive test was carried out on non freezing patients from Phase 2, once processing their data with a pipeline without RIE and once with a pipeline including RIE. This double procedure was done to evaluate the effect of the RIE on false positives and specificity (larger specificity implies also that the classifier was more precise). In the ideal case, the algorithm should identify 0 FPs and specificity=1, considering that no freezing events are recorded in these patients. The results were computed patient-wise and so are reported in Table 25. For the sake of brevity, only the results obtained using binary SVM trained on selected features from Phase 1 and 2 are reported in the table. This model was chosen because it is the best performing in terms of precision in validation.

As it can be observed, the introduction of RIE produces for all patients a drastic reduction of the false positives, hence an increase of specificity, which is above 90% in all the cases but 3 and often with value 100%. Without segmentation in RIs, the average specificity is quite low (61%), possibly because a lot of activities that confuse the classifier or that were not present enough in the training data are observed. When the RIE is introduced, most of these activities are removed and the classifier is capable of understanding that none of the remaining windows

contains FoG, with an increase of specificity to an average value of 97%. Figure 22 contains pie charts of which activities were mistaken as FoG with and without RIE.

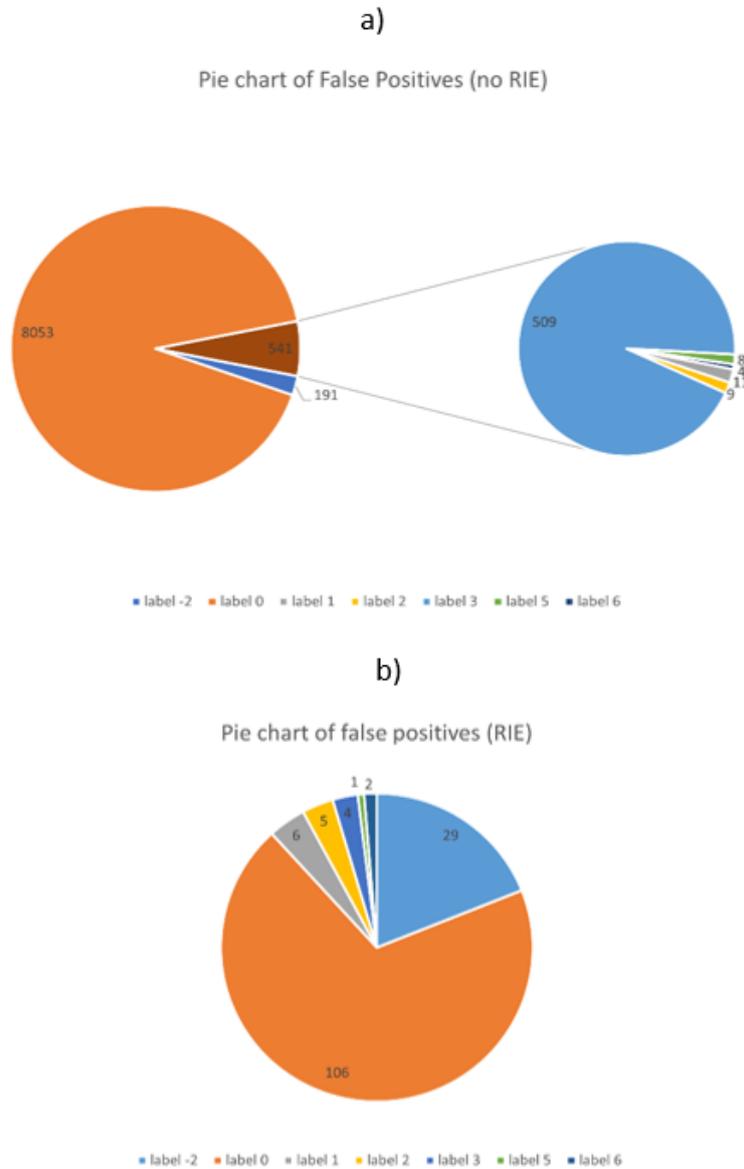


Figure 22: Type of windows that produced false positives in false positive test: a) the FP distribution in test without RIE; b) the FP distribution in test with RIE. Undefined activity 0, daily living mix -2 and stand 3 are responsible for most of the false positives.

As it can be observed from the pie charts, three are the types of window that mainly produce false positives: the undefined activity (0), the mix of daily-living-like activities (-2) and stand (3). When the Regions of Interest Extractor removes most of them, the specificity is boosted; indeed, with RIE on, most of the False Positives come from the the residual of those three activities, but for few exceptions that are negligible considering the large number of classified windows.

Table 25: Results of false positive test on non freezers from Phase 2

Patient ID	SVM (RBF)			
	NO RIE		RIE	
	FP	Specificity [%]	FP	Specificity [%]
1	548	39.8	17	86.0
2	325	71.4	4	95.3
3	15	97.0	0	100.0
5	40	73.3	0	100.0
6	1	99.5	0	100.0
7	35	74.2	0	100.0
8	372	29.0	8	91.6
10	184	49.0	1	99.0
11	210	26.6	1	98.3
12	467	28.4	4	94.2
13	324	50.2	19	86.8
14	227	55.1	3	96.5
15	250	44.0	0	100.0
16	5	96.5	1	98.1
17	725	45.9	0	100.0
18	225	42.2	2	98.2
19	267	37.2	38	76.1
20	211	48.0	0	100.0
21	286	36.0	4	95.9
22	74	84.6	8	93.1
23	93	79.2	2	98.9
24	510	42.9	14	92.1
25	485	36.9	1	98.9
26	267	68.6	0	100.0
27	360	29.7	0	100.0
28	246	40.4	6	95.8
29	54	75.4	1	99.0
30	40	83.1	1	98.9
31	1	98.6	0	100.0
32	13	94.6	8	95.4
33	47	85.7	0	100.0
35	23	77.7	0	100.0
36	87	58.3	0	100.0
37	5	96.9	2	97.9
38	153	57.5	5	93.7
39	931	23.5	0	100.0
40	305	39.1	3	95.7
41	259	50.3	1	98.9
42	32	76.8	0	100.0
43	64	58.1	0	100.0
44	16	86.9	0	100.0
47	6	91.2	0	100.0

## 5 Conclusions

### 5.1 Achieved goals

The proposed processing pipeline was designed for two goals: increasing time efficiency and precision in FoG recognition, reducing as much as possible false positives. From the results described in Chapter 4, it is reasonable to affirm that both were achieved. In terms of time efficiency, even adding the time required by RIE to run, the time for extracting the windows and their features for a patient is still smaller than processing the whole signals. About precision values, good results were achieved in validation of Phase 1 and 2 together, with the best results for binary SVM with RBF kernel. Also recall, which in general has an opposite behaviour with respect to precision, remained largely above 80%. Moreover, a neat reduction of FPs was observed when employing the RIE, which is relevant for an at home monitoring where many activities (as observed in Phase 2 data) could provide false alarms. Time efficiency and precision were achieved without losing too many FoG episodes: indeed, in both phases, selection of RIs maintained 96% of seconds classified as FoG. Furthermore, it must be considered that the following system was implemented using only acceleration signals coming from one single sensor; using data coming from multiple sensors in different locations could allow to design a more sophisticated RIE with even better performances.

Different models, both binary and multiclass were trained and compared. SVM was confirmed as a very promising shallow method for FoG recognition, but also KNN proved to be effective, either in the multiclass or the binary case. About models trained on principal components, it was observed that the features needed for FoG recognition can be condensed in a small set, at the expense of some reduction in performances. However, this could be exploited when a trade off between performances and dimension of feature set is required, for example for systems with limited memory.

Some limitations, however, must be taken into account: the pipeline was designed for an offline application and some effort would be required for its translation into an online system. In addition, it should be considered that the results were obtained from a dataset which contains overall few FoG episodes and only of the “trembling of leg” type, hence their statistical validity is reduced. Additional data would be required to provide a more thorough training and evaluation of the proposed pipeline.

### 5.2 Future improvements

About possible future improvements to this study, a number of options are available. First of all, about the data, it would be interesting to leverage also angular velocity signals provided by gyroscope, both for improving the RIE and to define some additional features useful for FoG recognition, which are generally related to properties of acceleration signals only. Moreover, collecting more data containing FoG events would be crucial to generalize the proposed algorithm as much as possible. A different approach would be to tailor the proposed pipeline for each single patient, as partially explored when studying the options for thresholding

in RIE. This could be done collecting data from a patient at his own home, with recordings spanning the whole day: in this way a number of freezing episodes could be observed, surrounded by daily living activities. On this data only, the whole pipeline could be trained and tested; it is reasonable to expect that the performances would be largely improved by this approach, considering the results that were obtained mixing data from different patients with different gait characteristics. About the pipeline itself, a step that offers room for improvements is feature selection. For example, a more structured approach could be implemented using a wrapper method, in combination with a final SVM, a classifier that proved efficient in the current study. In addition to this, it would be interesting to define additional features from CWT coefficients, considering that the feature *CWT Freezing ratio* was found quite relevant in the classification task. Finally, considering that the RIE worked in the order of ms for an offline extraction over signals with an average length of 6 minutes, it would be interesting to adapt it for an online system, working on real time windows of few seconds.

## References

- [1] Schrag A., Horsfall L., Walters K., and Petersen I. Presentations of Parkinson's disease before diagnosis - A large case-control study of the pre-diagnostic features of Parkinson's disease. *Movement Disorders*, 2014.
- [2] Paul S Addison. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance 1st Edition*. CRC Press, 2002.
- [3] Claas Ahlrichs, Albert Samà, Michael Lawo, Joan Cabestany, Daniel Rodríguez-Martín, Carlos Pérez-López, Dean Sweeney, Leo R Quinlan, Gearòid Laighin, Timothy Counihan, Patrick Browne, Lewy Hadas, Gabriel Vainstein, Alberto Costa, Roberta Annicchiarico, Sheila Alcaine, Berta Mestre, Paola Quispe, Àngels Bayes, and Alejandro Rodríguez-Molinero. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. *Medical and Biological Engineering and Computing*, 54(1):223–233, 2016.
- [4] Q. J. Almeida and C. A. Lebold. Freezing of gait in Parkinson's disease: A perceptual cause for a motor impairment? *Journal of Neurology, Neurosurgery and Psychiatry*, 2010.
- [5] Guido Alves, Elin Bjelland Forsaa, Kenn Freddy Pedersen, Michaela Dreetz Gjerstad, and Jan Petter Larsen. Epidemiology of Parkinson's disease. In *Journal of Neurology*, 2008.
- [6] M. Bächlin, M. Plotnik, D. Roggen, N. Giladi, J. M. Hausdorff, and G. Tröster. A wearable system to assist walking of parkinson's disease patients benefits and challenges of context-triggered acoustic cueing. *Methods of Information in Medicine*, 2010.
- [7] Anna L. Bartels and Klaus L. Leenders. Parkinson's disease: The syndrome, the pathogenesis and pathophysiology. *Cortex*, 2009.
- [8] Claudia Barthel, Elizabeth Mallia, Bettina Debû, Bastiaan R Bloem, and Murielle Ursulla Ferraye. The Practicalities of Assessing Freezing of Gait, 2016.
- [9] Akram Bayat, Marc Pomplun, and Duc A Tran. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, 34(C):450–457, 2014.
- [10] A Berardelli. Pathophysiology of bradykinesia in Parkinson's disease. *Brain*, 124(11):2131–2146, nov 2001.
- [11] Niranjana Bidargaddi, Antti Sarela, Lasse Klingbeil, and Mohanraj Karunanithi. Detecting walking activity in cardiac rehabilitation by using accelerometer. In *Proceedings of the 2007 International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP*, 2007.

- [12] Annett Blochberger and Shelley Jones. Parkinson’s disease clinical features and diagnosis. *Clinical Pharmacist*, 2011.
- [13] Bastiaan R Bloem, Jeffrey M Hausdorff, Jasper E Visser, and Nir Giladi. Falls and freezing of Gait in Parkinson’s disease: A review of two interconnected, episodic phenomena, 2004.
- [14] Emily Bomasang-Layno, Iris Fadlon, Andrea N. Murray, and Seth Himelhoch. Antidepressive treatments for Parkinson’s disease: A systematic review and meta-analysis, 2015.
- [15] Luigi Borzì, Gabriella Olmo, Carlo Alberto Artusi, and Leonardo Lopiano. Detection of freezing of gait in people with parkinson’s disease using smartphones. In *Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, volume 1, pages 625–635. IEEE, 2020.
- [16] Leonid Breydo, Jessica W Wu, and Vladimir N Uversky.  $\alpha$ -Synuclein misfolding and Parkinson’s disease, feb 2012.
- [17] Julià Camps, Albert Samà, Mario Martín, Daniel Rodríguez-Martín, Carlos Pérez-López, Joan M. Moreno Arostegui, Joan Cabestany, Andreu Català, Sheila Alcaine, Berta Mestre, Anna Prats, Maria C. Crespo-Maraver, Timothy J. Counihan, Patrick Browne, Leo R. Quinlan, Gearóid Laighin, Dean Sweeney, Hadas Lewy, Gabriel Vainstein, Alberto Costa, Roberta Annicchiarico, Àngels Bayés, and Alejandro Rodríguez-Molinero. 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*, 2018.
- [18] Marianna Capecci, Lucia Pepa, Federica Verdini, and Maria Gabriella Ceravolo. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson’s disease. *Gait and Posture*, 50:28–33, 2016.
- [19] Nicole A Capela, Edward D Lemaire, and Natalie Baddour. Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. *PLoS ONE*, 10(4):e0124414, 2015.
- [20] Nitesh V. Chawla, Nathalie Japkowicz, and Aleksander Kotcz. Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 2004.
- [21] Pei-Hao Chen, Rong-Long Wang, De-Jyun Liou, and Jin-Siang Shaw. Gait Disorders in Parkinson’s Disease: Assessment and Management. *International Journal of Gerontology*, 7(4):189–193, 2013.
- [22] Guglielmo Cola, Marco Avvenuti, Fabio Musso, and Alessio Vecchio. Personalized gait detection using a wrist-worn accelerometer. In *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2017*, 2017.

- [23] Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet, Maud Pasquier, Bernard Espiau, and Christian Geny. Detection of freezing of gait in Parkinson disease: Preliminary results. *Sensors (Switzerland)*, 14(4):6819–6827, apr 2014.
- [24] Nabila Dahodwala, Andrew Siderowf, Ming Xie, Elizabeth Noll, Matthew Stern, and David S Mandell. Racial differences in the diagnosis of Parkinson’s disease. *Movement Disorders*, 24(8):1200–1205, 2009.
- [25] William Dauer and Serge Przedborski. Parkinson’s disease: Mechanisms and models, 2003.
- [26] C. A. Davie. A review of Parkinson’s disease, 2008.
- [27] Milica D Djuric-Jovicic, Nenad S Jovicic, Sasa M Radovanovic, Iva D Stankovic, Mirjana B Popovic, and Vladimir S Kostic. Automatic identification and classification of freezing of gait episodes in Parkinson’s disease patients. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(3):685–694, 2014.
- [28] E. R. Dorsey, R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. Kieburtz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf, and C. M. Tanner. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 2007.
- [29] Richard L. Doty, Daniel A. Deems, and Stanley Stellar. Olfactory dysfunction in parkinsonism: A general deficit unrelated to neurologic signs, disease stage, or disease duration. *Neurology*, 1988.
- [30] European Parkinson’s Disease Association. What is Parkinson’s? <https://www.epda.eu.com/about-parkinsons/what-is-parkinsons>.
- [31] Stanley Fahn. Description of Parkinson’s disease as a clinical syndrome. In *Annals of the New York Academy of Sciences*, 2003.
- [32] Murielle U. Ferraye, Bettina Debû, and Pierre Pollak. Deep brain stimulation effect on freezing of gait, 2008.
- [33] Mohd Fikri, Ali Fahmi, Perwira Negara, Shohel Sayeed, and Deok-jai Choi. Classification Algorithms in Human Activity Recognition using Smartphones. *International Journal of Computer and Information Engineering*, 6(8):77–84, 2012.
- [34] Lysia S. Forno. Neuropathology of Parkinson’s disease, 1996.
- [35] Emma Fortune, Vipul A. Lugade, and Kenton R. Kaufman. Posture and movement classification: The comparison of tri-axial accelerometer numbers and anatomical placement. *Journal of Biomechanical Engineering*, 2014.
- [36] J G Francois, Vingerhoets, Michael Schulzer, Donald B Calne, and Barry J Snow. Which clinical sign of Parkinson’s disease best reflects the nigrostriatal lesion? *Annals of Neurology*, 41(1):58–64, 1997.

- [37] Nir Giladi and Jeffrey M Hausdorff. The role of mental function in the pathogenesis of freezing of gait in Parkinson’s disease. *Journal of the Neurological Sciences*, 248(1-2):173–176, 2006.
- [38] Nir Giladi, Joseph Tal, Tali Azulay, Oliver Rascol, David J Brooks, Eldad Melamed, Wolfgang Oertel, Werner H Poewe, Fabrizio Stocchi, and Eduardo Tolosa. Validation of the Freezing of Gait Questionnaire in patients with Parkinson’s disease. *Movement Disorders*, 24(5):655–661, apr 2009.
- [39] Pieter Ginis, Evelien Nackaerts, Alice Nieuwboer, and Elke Heremans. Cueing for people with Parkinson’s disease with freezing of gait: A narrative review of the state-of-the-art and novel perspectives, 2018.
- [40] Christopher G Goetz, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Glenn T Stebbins, Matthew B Stern, Barbara C Tilley, Richard Dodel, Bruno Dubois, Robert Holloway, Joseph Jankovic, Jaime Kulisevsky, Anthony E Lang, Andrew Lees, Sue Leurgans, Peter A LeWitt, David Nyenhuis, C Warren Olanow, Olivier Rascol, Anette Schrag, Jeanne A Teresi, Jacobus J Van Hilten, and Nancy LaPelle. Movement disorder society-sponsored revision of the unified Parkinson’s disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Movement Disorders*, 22(1):41–47, 2007.
- [41] Isabelle Guyon and André Elisseeff. An introduction of variable and feature selection. *J. Machine Learning Research Special Issue on Variable and Feature Selection*, 3:1157 – 1182, 01 2003.
- [42] Mark Hallett. Parkinsonism and Related Disorders Parkinson ’ s disease tremor : pathophysiology. *Parkinsonism and related Disorders*, 18(2012):S85—S86, 2016.
- [43] Jonghee Han, Hyo Seon Jeon, Won Jin Yi, Beom Seok Jeon, and Kwang Suk Park. Adaptive windowing for gait phase discrimination in Parkinsonian gait using 3-axis acceleration signals. *Medical and Biological Engineering and Computing*, 2009.
- [44] Hanbyul Kim, Hong Ji Lee, Woongwoo Lee, Sungjun Kwon, Sang Kyong Kim, Hyo Seon Jeon, Hyeyoung Park, Chae Won Shin, Won Jin Yi, Beom S Jeon, and Kwang S Park. Unconstrained detection of freezing of Gait in Parkinson’s disease patients using smartphone. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, volume 2015, pages 3751–3754. IEEE, 2015.
- [45] Rick C Helmich, Mark Hallett, Günther Deuschl, Ivan Toni, and Bastiaan R Bloem. Cerebral causes and consequences of parkinsonian resting tremor: A tale of two circuits?, nov 2012.
- [46] Mariese A Hely, John G L Morris, Wayne G J Reid, and Robert Trafficante. Sydney Multicenter Study of Parkinson’s disease: Non-L-dopa-responsive

- problems dominate at 15 years. *Movement Disorders*, 20(2):190–199, feb 2005.
- [47] Elke Heremans, Alice Nieuwboer, Joke Spildooren, Jochen Vandenbossche, Natacha Deroost, Eric Soetens, Eric Kerckhofs, and Sarah Vercruysse. Cognitive aspects of freezing of gait in parkinson’s disease: A challenge for rehabilitation. *Journal of neural transmission (Vienna, Austria : 1996)*, 120, 01 2013.
- [48] Elke Heremans, Alice Nieuwboer, and Sarah Vercruysse. Freezing of gait in Parkinson’s disease: where are we now?, 2013.
- [49] Aaron Courville Ian Goodfellow, Yoshua Bengio. Deep Learning Book. *Deep Learning*, 2015.
- [50] Natalie J Ives, Rebecca L Stowe, Joanna Marro, Carl Counsell, Angus Macleod, Carl E Clarke, Richard Gray, and Keith Wheatley. Monoamine oxidase type B inhibitors in early Parkinson’s disease: Meta-analysis of 17 randomised trials involving 3525 patients. *British Medical Journal*, 329(7466):593–596, 2004.
- [51] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibishirani. *An Introduction to Statistical Learning with Applications in R (older version)*. Springer, 2013.
- [52] J Jankovic. Parkinson’s disease: Clinical features and diagnosis, 2008.
- [53] Joseph Jankovic. Pathophysiology and clinical assessment of parkinsonian symptoms and signs. In *Handbook of Parkinson’s Disease, Third Edition*. 2003.
- [54] Wolfgang H. Jost. Gastrointestinal dysfunction in Parkinson’s Disease. *Journal of the Neurological Sciences*, 2010.
- [55] E Jovanov, E Wang, L Verhagen, M Fredrickson, and R Fratangelo. deFOG — A real time system for detection and unfreezing of gait of Parkinson’s patients. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2009, pages 5151–5154. IEEE, 2009.
- [56] Naheed L Khan, Elizabeth Graham, Peter Critchley, Anette E Schrag, Nicholas W Wood, Andrew J Lees, Kailash P Bhatia, and Niall Quinn. Parkin disease: A phenotypic study of a large case series. *Brain*, 126(6):1279–1292, 2003.
- [57] J William Langston, Hakan Widner, Christopher G Goetz, David Brooks, Stanley Fahn, Thomas Freeman, and Ray Watts. Core assessment program for intracerebral transplantations (CAPIT). *Movement Disorders*, 7(1):2–13, 1992.

- [58] 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. In *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2018*, 2018.
- [59] Kah Leong Lim and Jeanne M.M. Tan. Role of the ubiquitin proteasome system in Parkinson’s disease, 2007.
- [60] Michael Macht, Yvonne Kaussner, Jens Carsten Möller, Karin Stiasny-Kolster, Karla Maria Eggert, Hans Peter Krüger, and Heiner Ellgring. Predictors of freezing in Parkinson’s disease: A survey of 6,620 patients. *Movement Disorders*, 22(7):953–956, mar 2007.
- [61] Francesca Magrinelli, Alessandro Picelli, Pierluigi Tocco, Angela Federico, Laura Roncari, Nicola Smania, Giampietro Zanette, and Stefano Tamburin. Pathophysiology of Motor Dysfunction in Parkinson’s Disease as the Rationale for Drug Treatment and Rehabilitation, 2016.
- [62] Martina Mancini, Kelsey C. Priest, John G. Nutt, and Fay B. Horak. Quantifying freezing of gait in Parkinson’s disease during the instrumented timed up and go test. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2012.
- [63] João Massano and Kailash P. Bhatia. Clinical approach to Parkinson’s disease: Features, diagnosis, and principles of management. *Cold Spring Harbor Perspectives in Medicine*, 2012.
- [64] S. Mazilu, U. Blanke, and G. Tröster. Gait, wrist, and sensors: Detecting freezing of gait in parkinson’s disease from wrist movement. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 579–584, 2015.
- [65] S Mazilu, A Calatroni, E Gazit, D Roggen, J M Hausdorff, and G Tröster. Feature learning for detection and prediction of freezing of gait in Parkinson’s disease. *Proc. Workshop Math. Methods Biomed. Image Analysis*, 7988 LNAI:144, 2013.
- [66] Sinziana Mazilu, Ulf Blanke, Michael Hardegger, Gerhard Troster, Eran Gazit, Moran Dorfman, and Jeffrey M. Hausdorff. GaitAssist: A wearable assistant for gait training and rehabilitation in Parkinson’s disease. In *2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS 2014*, 2014.
- [67] Sinziana Mazilu, Michael Hardegger, Zack Zhu, Daniel Roggen, Gerhard Tröster, Meir Plotnik, Jeffrey M Hausdorff, and Gerhard Troester. 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*, pages 123–130, 2012.

- [68] Aravind Mittur, Suneel Gupta, and Nishit B Modi. Pharmacokinetics of Rytary®, An Extended-Release Capsule Formulation of Carbidopa–Levodopa, 2017.
- [69] Steven T Moore, Hamish G MacDougall, and William G Ondo. Ambulatory monitoring of freezing of gait in Parkinson’s disease. *Journal of Neuroscience Methods*, 167(2):340–348, 2008.
- [70] Steven T Moore, Don A Yungger, Tiffany R Morris, Valentina Dilda, Hamish G MacDougall, James M Shine, Sharon L Naismith, and Simon J G Lewis. Autonomous identification of freezing of gait in Parkinson’s disease from lower-body segmental accelerometry. *Journal of NeuroEngineering and Rehabilitation*, 10(1):19, feb 2013.
- [71] Tiffany R. Morris, Catherine Cho, Valentina Dilda, James M. Shine, Sharon L. Naismith, Simon J.G. Lewis, and Steven T. Moore. A comparison of clinical and objective measures of freezing of gait in Parkinson’s disease. *Parkinsonism and Related Disorders*, 2012.
- [72] Weerasak Muangpaisan, Hiroyuki Hori, and Carol Brayne. Systematic Review of the Prevalence and Incidence of Parkinson’s Disease in Asia. *Journal of Epidemiology*, 19(6):281–293, 2009.
- [73] Alvaro Muro-de-la Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications. *Sensors*, 14(2):3362–3394, feb 2014.
- [74] Alice Nieuwboer and Nir Giladi. Characterizing freezing of gait in Parkinson’s disease: Models of an episodic phenomenon, 2013.
- [75] Alice Nieuwboer, Lynn Rochester, Talia Herman, Wim Vandenberghe, George Ehab Emil, Tom Thomaes, and Nir Giladi. Reliability of the new freezing of gait questionnaire: Agreement between patients with Parkinson’s disease and their carers. *Gait and Posture*, 30(4):459–463, nov 2009.
- [76] L. Niu, Xi Tuan Ji, J. M. Li, D. S. Zhao, G. Huang, W. P. Liu, Y. Qu, Lian Ting Ma, and X. T. Ji. Effect of bilateral deep brain stimulation of the subthalamic nucleus on freezing of gait in Parkinson’s disease. *Journal of International Medical Research*, 2012.
- [77] Jorik Nonnekes, Arno M Janssen, Senja H G Mensink, Lars B Oude Nijhuis, Bastiaan R Bloem, and Anke H Snijders. Short rapid steps to provoke freezing of gait in Parkinson’s disease. *Journal of Neurology*, 261(9):1763–1767, 2014.
- [78] Jorik Nonnekes, Anke H Snijders, John G Nutt, Günter Deuschl, Nir Giladi, and Bastiaan R Bloem. Freezing of gait: A practical approach to management, 2015.

- [79] John G Nutt, Bastiaan R Bloem, Nir Giladi, Mark Hallett, Fay B Horak, and Alice Nieuwboer. Freezing of gait: Moving forward on a mysterious clinical phenomenon, 2011.
- [80] Minh H. Pham, Morad Elshehabi, Linda Haertner, Silvia Del Din, Karin Srulijes, Tanja Heger, Matthis Synofzik, Markus A. Hobert, Gert S. Faber, Clint Hansen, Dina Salkovic, Joaquim J. Ferreira, Daniela Berg, Álvaro Sanchez-Ferro, Jaap H. van Dieën, Clemens Becker, Lynn Rochester, Gerhard Schmidt, and Walter Maetzler. Validation of a step detection algorithm during straight walking and turning in Patients with Parkinson’s disease and older adults using an inertial measurement unit at the lower back. *Frontiers in Neurology*, 2017.
- [81] Kemal Polat. Freezing of gait (FoG) detection using logistic regression in Parkinson’s disease from acceleration signals. In *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*, 2019.
- [82] E. Ray Dorsey, Alexis Elbaz, Emma Nichols, Foad Abd-Allah, Ahmed Abdelalim, Jose C. Adsuar, Mustafa Geleto Ansha, Carol Brayne, Jee Young J. Choi, Daniel Collado-Mateo, Nabila Dahodwala, Huyen Phuc Do, Dumessa Edessa, Matthias Endres, Seyed Mohammad Fereshtehnejad, Kyle J. Foreman, Fortune Gbetoho Gankpe, Rahul Gupta, Graeme J. Hankey, Simon I. Hay, Mohamed I. Hegazy, Desalegn T. Hibstu, Amir Kasaeian, Yousef Khader, Ibrahim Khalil, Young Ho Khang, Yun Jin Kim, Yoshihiro Kokubo, Giancarlo Logroscino, João Massano, Norlinah Mohamed Ibrahim, Mohammed A. Mohammed, Alireza Mohammadi, Maziar Moradi-Lakeh, Mohsen Naghavi, Binh Thanh Nguyen, Yirga Legesse Nirayo, Felix Akpojene Ogbo, Mayowa Ojo Owolabi, David M. Pereira, Maarten J. Postma, Mostafa Qorbani, Muhammad Aziz Rahman, Kedir T. Roba, Hosein Safari, Saeid Safiri, Maheswar Satpathy, Monika Sawhney, Azadeh Shafieesabet, Mekonnen Sisay Shiferaw, Mari Smith, Cassandra E.I. Szoeki, Rafael Tabarés-Seisdedos, Nu Thi Truong, Kingsley Nnanna Ukwaja, Narayanaswamy Venketasubramanian, Santos Villafaina, Kidu Gidey Weldegewergs, Ronny Westerman, Tissa Wijeratne, Andrea S. Winkler, Bach Tran Xuan, Naohiro Yonemoto, Valery L. Feigin, Theo Vos, and Christopher J.L. Murray. Global, regional, and national burden of Parkinson’s disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 2018.
- [83] Saba Rezvanian and Thurmon Lockhart. Towards Real-Time Detection of Freezing of Gait Using Wavelet Transform on Wireless Accelerometer Data. *Sensors*, 16(4):475, apr 2016.
- [84] Priscila A. Rocha, Gustavo M. Porfírio, Henrique B. Ferraz, and Virginia F.M. Trevisani. Effects of external cues on gait parameters of Parkinson’s disease patients: A systematic review, 2014.

- [85] Daniel Rodríguez-Martín, Albert Samà, Carlos Pérez-López, Andreu Català, Joan M. Moreno Arostegui, Joan Cabestany, Àngels Bayés, Sheila Alcaine, Berta Mestre, Anna Prats, M Cruz Crespo, Timothy J Coughlin, Patrick Browne, Leo R Quinlan, Gearóid Laighin, Dean Sweeney, Hadas Lewy, Joseph Azuri, Gabriel Vainstein, Roberta Annicchiarico, Alberto Costa, and Alejandro Rodríguez-Molinero. Home detection of freezing of gait using Support Vector Machines through a single waist-worn triaxial accelerometer. *PLoS ONE*, 12(2):e0171764, 2017.
- [86] Maria C Rodriguez-Oroz, Marjan Jahanshahi, Paul Krack, Irene Litvan, Raúl Macías, Erwan Bezard, and José A Obeso. Initial clinical manifestations of Parkinson’s disease: features and pathophysiological mechanisms, 2009.
- [87] J C Rothwell, J A Obeso, M M Traub, and C D Marsden. The behaviour of the long - latency stretch reflex in patients with Parkinson’s Disease. *Journal of Neurology, Neurosurgery, Psychiatry*, 46:35–44, 1983.
- [88] Sapir S., Ramig L., and Fox C. Speech and swallowing disorders in Parkinson disease, 2008.
- [89] A Samà, C Pérez-López, D Rodríguez-Martín, A Català, J M Moreno-Arostegui, J Cabestany, E de Mingo, and A Rodríguez-Molinero. Estimating bradykinesia severity in Parkinson’s disease by analysing gait through a waist-worn sensor. *Computers in Biology and Medicine*, 84(March):114–123, 2017.
- [90] Rubén San-Segundo, Honorio Navarro-Hellín, Roque Torres-Sánchez, Jessica Hodgins, and Fernando de la Torre. Increasing robustness in the detection of freezing of gait in Parkinson’s disease. *Electronics (Switzerland)*, 2019.
- [91] Joseph M Savitt, Valina L Dawson, and Ted M Dawson. Diagnosis and treatment of Parkinson disease: molecules to medicine. *The Journal of clinical investigation*, 116(7):1744–1754, 2006.
- [92] J D Schaafsma, Y Balash, T Gurevich, A L Bartels, J M Hausdorff, and Nir Giladi. Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson’s disease. *European Journal of Neurology*, 10(4):391–398, 2003.
- [93] Scikit-developers. Feature selection in scikit-learn. [https://scikit-learn.org/stable/modules/feature\\_selection.html](https://scikit-learn.org/stable/modules/feature_selection.html).
- [94] Ana Lígia Silva de Lima, Luc J W Evers, Tim Hahn, Lauren Bataille, Jamie L Hamilton, Max A Little, Yasuyuki Okuma, Bastiaan R Bloem, and Marjan J Faber. Freezing of gait and fall detection in Parkinson’s disease using wearable sensors: a systematic review. *Journal of Neurology*, 264(8):1642–1654, 2017.

- [95] Anke H Snijders, Maarten J Nijkrake, Maaike Bakker, Marten Munneke, Carina Wind, and Bastiaan R Bloem. Clinimetrics of freezing of gait. *Movement Disorders*, 23(S2):S468—S474, 2008.
- [96] Movement Disorder Society. MDS-UPDRS-Italian-Official-Working-Document. 2008.
- [97] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 2009.
- [98] Joke Spildooren, Sarah Vercruysse, Kaat Desloovere, Wim Vandenberghe, Eric Kerckhofs, and Alice Nieuwboer. Freezing of gait in Parkinson’s disease: The impact of dual-tasking and turning. *Movement Disorders*, 2010.
- [99] Sigurlaug Sveinbjornsdottir. The clinical symptoms of Parkinson’s disease, 2016.
- [100] Dean Sweeney, Leo R. Quinlan, Patrick Browne, Margaret Richardson, Pauline Meskell, and Gearóid Ólaighin. A technological review of wearable cueing devices addressing freezing of gait in Parkinson’s disease, 2019.
- [101] Weijun Tao, Tao Liu, Rencheng Zheng, and Hutian Feng. Gait analysis using wearable sensors, 2012.
- [102] Hoang Minh Thang, Vo Quang Viet, Nguyen Dinh Thuc, and Deokjai Choi. Gait identification using accelerometer on mobile phone. In *2012 International Conference on Control, Automation and Information Sciences, IC-CAIS 2012*, 2012.
- [103] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 2018.
- [104] Kaiyu Tong and Malcolm H. Granat. A practical gait analysis system using gyroscopes. *Medical Engineering and Physics*, 1999.
- [105] Evanthia E Tripoliti, Alexandros T Tzallas, Markos G Tsipouras, George Rigas, Panagiota Bougia, Michael Leontiou, Spiros Konitsiotis, Maria Chondrogiorgi, Sofia Tsouli, and Dimitrios I Fotiadis. Automatic detection of freezing of gait events in patients with Parkinson’s disease. *Computer Methods and Programs in Biomedicine*, 110(1):12–26, apr 2013.
- [106] Alexandros T Tzallas, Markos G Tsipouras, Georgios Rigas, Dimitrios G Tsalikakis, Evaggelos C Karvounis, Maria Chondrogiorgi, Fotis Psomadellis, Jorge Cancela, Matteo Pastorino, María Teresa Arredondo Waldmeyer, Spiros Konitsiotis, and Dimitrios I Fotiadis. PERFORM: a system for monitoring, assessment and management of patients with Parkinson’s disease. *Sensors (Basel, Switzerland)*, 14(11):21329–21357, nov 2014.

- [107] S K Van Den Eeden. Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity. *American Journal of Epidemiology*, 157(>11):1015–1022, 2003.
- [108] Richard Williamson and Brian J. Andrews. Gait event detection for FES using accelerometers and supervised machine learning. *IEEE Transactions on Rehabilitation Engineering*, 2000.
- [109] Kristian Winge, L Friberg, L Werdelin, K K Nielsen, and H Stimpel. Relationship between nigrostriatal dopaminergic degeneration, urinary symptoms, and bladder control in Parkinson's disease. *European Journal of Neurology*, 12(11):842–850, nov 2005.
- [110] Konstanze F Winklhofer and Christian Haass. Mitochondrial dysfunction in Parkinson's disease, 2010.
- [111] 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*, 2018.
- [112] Heidemarie Zach, Arno M Janssen, Anke H Snijders, Arnaud Delval, Murielle U Ferraye, Eduard Auff, Vivian Weerdesteyn, Bastiaan R Bloem, and Jorik Nonnekes. Identifying freezing of gait in Parkinson's disease during freezing provoking tasks using waist-mounted accelerometry. *Parkinsonism and Related Disorders*, 21(11):1362–1366, nov 2015.