Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Control and Computer Engineering (35^{th}cycle)

Wearable Sensors and Artificial Intelligence for Monitoring of Parkinson's Disease

By

Luigi Borzì

******

Supervisor(s):
Prof. Gabriella Olmo, Supervisor
Prof. Alfredo Benso, Co-Supervisor
Prof. Elisa Ficarra, Co-Supervisor

Doctoral Examination Committee:
Prof. Silvia Del Din, (Referee), University of Newcastle, United Kingdom
Prof. Veronica Cimolin, (Referee), Politenico di Milano, Italy
Prof. Florenc Demrozi, University of Stavanger, Norway
Prof. Marco Mazza, University of Fribourg, Switzerland
Prof. Fulvio Corno, Politecnico di Torino, Italy

Politecnico di Torino
2023

Declaration

I hereby declare that the contents and organization of this dissertation constitute my own original work and do not compromise in any way the rights of third parties, including those relating to the security of personal data. The content of the methods and results of the thesis has been extracted and rearranged from the present author's published papers.

Luigi Borzì
2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Abstract

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting more than 10 million people worldwide. Being a movement disorder, many motor symptoms manifest at different stages of the disease, including bradykinesia (slowness of movement), postural stability, gait impairment, and freezing of gait (FOG - a sudden motor block). PD represents a very complex disorder for several reasons, including the difficulty of early diagnosis, the heterogeneity in the manifestation of symptoms, their evolution, and their fluctuations throughout the day. For these reasons, the longitudinal monitoring of PD, possibly performed in daily life conditions, is fundamental for achieving a global picture of the disease and its evolution, allowing the planning of proper therapy adjustments. A large spectrum of technological solutions is available for the accurate and precise evaluation of motor impairment in PD. However, most of them are costly and can be used only in laboratory settings. In this context, wearable motion sensors represent a valuable and ecological solution for collecting and processing movement data, allowing to continuously monitor PD in free-living settings. Moreover, the combination of wearable sensors and machine learning (ML) methods provides great opportunities for an accurate and objective evaluation of motor impairment in PD.

This present thesis aims to provide new opportunities for monitoring PD using wearable sensors and ML. A wide range of sensor settings and signal processing, ML, and deep learning (DL) methods for PD assessment are described and discussed in this study. Indeed, their use for the evaluation and prediction of specific motor symptoms is described.

Overall, motion data were recorded from more than 200 subjects with PD, using different experimental protocols for different objectives. Specifically, this thesis describes and discusses methods for the automatic evaluation of bradykinesia, estimation of postural stability and gait impairment, and detection and prediction of

FOG. Both dedicated hardware and sensors embedded in commercial smartphones were used for data collection. Different signal processing, ML, and DL algorithms were designed to maximize performance while maintaining a low computation burden. When evaluating lower-limb bradykinesia using a single smartphone, results demonstrated that the correlation between the computer scoring system and the average clinical score was larger than the best agreement among four independent raters. This suggests that, for specific tasks, computer methods may overcome inter-rater variability. When assessing postural stability using a smartphone during simple activities, such as stance and turning, results suggested that it is possible to obtain a gross evaluation of the postural response in PD. This information may be important for predicting the risk of falls and taking proper countermeasures. The combination of a single inertial sensor and DL methods provided promising performance in FOG detection, even predicting FOG before its actual occurrence. The large number of participants involved in the analysis and the heterogeneity of activities executed by subjects strongly enforce the validity of the obtained results. Moreover, the high-speed and low-memory characteristics of the developed algorithm suggest a possible real-time implementation of the detection model in a stand-alone wearable device. The designed solution can be used for triggering some kind of auditory or tactile cue for reducing FOG or even preventing its occurrence.

Overall, the results suggest that it is possible to use simple and non-invasive technology for monitoring several motor aspects of PD. The designed solutions can be used during the normal follow-up clinical visits for a more objective estimation of motor symptoms severity. Moreover, they can be used to remotely assess the disease, providing precise and continuous measures describing the presence, severity, and fluctuations of PD motor signs.

Future studies will make use of even less obtrusive wearable solutions (e.g., smart clothes), in order to maximize patient compliance, thus allowing long-term monitoring of PD. Moreover, data will be collected and analyzed in semi-supervised and non-supervised environments, in order to evaluate the robustness of the designed algorithms in heterogeneous and complex real-life scenarios.

Contents

Acronyms

6MWT Six-minute walking test.

Adam adaptive moment estimation.

ADL Activities of daily living.

AI Artificial intelligence.

AUC Area under the curve.

CNN Convolutional neural network.

COM Center of mass.

CV Cross-validation.

CWT Continuous wavelet transform.

DBS Deep brain stimulation.

DL Deep learning.

DNN Deep neural network.

DT Decision tree.

EER Equal error rate.

EMG Electromyography.

FC Final contect.

FFT Fast Fourier transform.

FN False negative.

FOG Freezing of gait.

FP False positive.

FS Feature selection.

GD Gradient descent.

H&Y Hoehn and Yahr.

HAR Human activity recognition.

IC Initial contact.

ICC Intra-class correlation coefficient.

IMU Inertial measurement unit.

L-Dopa Levodopa.

LA Leg agility.

LB Lewy bodies.

LOSO Leave one subject out.

MAE Mean absolute error.

ML Machine learning.

MLP Multi-layer perceptron.

NPV Negative predictive value.

NWS Non-wearable systems.

PCA Principal component analysis.

PD Parkinson's disease.

PIGD Postural stability and gait difficulty.

PPV Positive predictive value.

PwPD Patients with Parkinson's disease.

QoL Quality of life.

QoM Quality of movement.

ReLU Rectified linear unit.

RMSE Root mean square error.

RNN Recurrent neural network.

ROC Receiver operating characteristic.

TN True negative.

TP True positive.

TUG Timed up and go.

UPDRS Unified Parkinson's disease rating scale.

UPS Ubiquiting proteasome system.

WMS Werable motion sensors.

WS Wearable sensors.

Chapter 1

Introduction

1.1 Parkinson's Disease

Parkinson's disease (PD) is a chronic, progressive, neurodegenerative disorder affecting more than 10 million people worldwide [1]. PD involves both motor and non-motor symptoms, associated with the degeneration of dopamine neurons, mainly in the *substantia nigra pars compacta* region of the brain stem (Figure 1.1 left). Due to its involvement in motor control, executive functions, gratification, and motivation, an impaired dopamine production (Figure 1.1 right) produces delayed and uncoordinated movements, deficits of attention, altered mood (e.g., anxiety, depression), and psychiatric disorders, severely affecting the quality of life (QoL) of patients with PD (PwPD) [2]. The main PD symptoms include bradykinesia, rigidity, tremor, and postural instability [3]. However, several non-motor symptoms (e.g., sleep disturbances, psychosis, depression, dementia, autonomic and gastrointestinal dysfunction) may occur as well [4, 5]. Current PD treatments aim to increase dopamine levels, with Levodopa (L-Dopa) being the most used one. Although this drug temporally reverts the symptoms, it does not prevent disease progression [6], and it can even produce motor complications such as dyskinesia (i.e., involuntary movements) [2].

Approximately 0.3% of the general population is affected by PD, and this value rises to 3% for subjects older than 65 years, being aging a major risk factor [8]. The mean age at onset is 55 [9], the mean age at diagnosis is 70 [10], and the incidence[1] ranges from 0.5 per 100,000 in the 30-40 age range to 120 per 100,000 in the oldest

[1] number of new cases per population at risk in a given time period

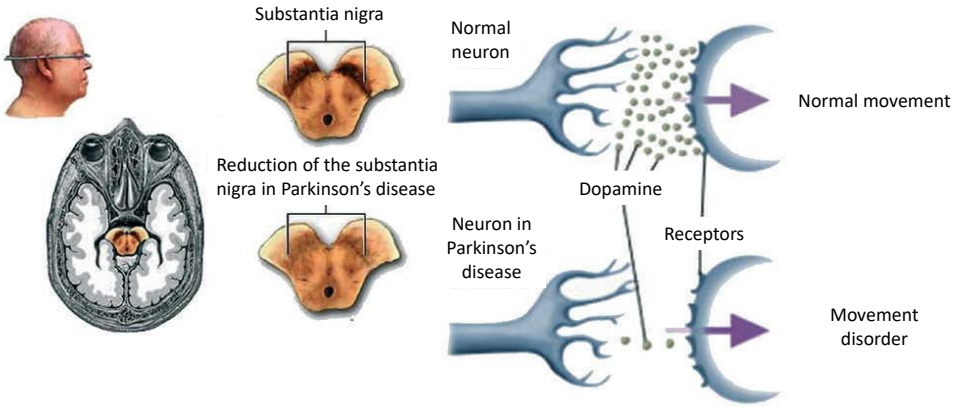


Fig. 1.1 Substantia nigra and dopamine reduction in Parkinson's disease (adapted from [7]).

population (over 70) [10]. Besides aging, the male sex is the most prominent risk factor for developing PD at all ages and nationalities [11]. Male-to-female ratios for incidence rates range from 1.37 to 3.7 and generally increase with age, suggesting that twice as many men than women suffer from PD [10]. It has been observed an increase of almost 50% in the incidence of both PD and Parkinsonism of all types over the 30-year period from 1976 to 2005, particularly in men older than 70 years. This trend needs to be interpreted with caution, as it may be due to improved access to care or increased awareness of signs and symptoms of Parkinsonism by physicians [12]. PD is typified by a degenerative process that affects dopaminergic neurons in the substantia nigra. How the degenerative processes damage both the nigrostriatal system and other brain regions are not completely clear. Literature studies suggest two major hypotheses regarding the pathogenesis of the disease. The first one indicates misfolding and aggregation of proteins to provoke the death of dopaminergic neurons [13], while the other proposes mitochondrial dysfunction and the consequent oxidative stress [14], including toxic oxidized dopamine species. The pathological hallmarks of PD include loss of nigrostriatal dopaminergic neurons and the presence of intraneuronal proteinaceous cytoplasmic inclusions, named Lewy bodies (LB) [9]. LB are α-synuclein inclusions composed of neurofilaments and proteins responsible for proteolysis. These include ubiquitin, a protein playing a primary role in targeting other proteins for clearance. Mutations in the α-synuclein

gene are responsible for some familial forms of PD in which LB are also observed [5]. It has been shown that Parkin[2] facilitates the binding of ubiquitin (ubiquitination) to other α-synuclein interacting proteins, leading to the formation of LB [15]. The ubiquitin-proteasome system (UPS) is thought to be linked to the development of cell death, as it plays an important role in intracellular proteolysis and several intracellular processes that maintain the viability of cells. Failure of the UPS leads to the abnormal aggregation of proteins including α-synuclein, which represents a major component of LB. The link between UPS and neurodegeneration has been reinforced by the discovery of mutations in genes that code for several proteins belonging to UPS in PD [5].

1.1.1 Symptoms

Figure 1.2 lists motor and non-motor signs of PD. The four cardinal symptoms of PD include tremor at rest, rigidity, bradykinesia, and postural instability.

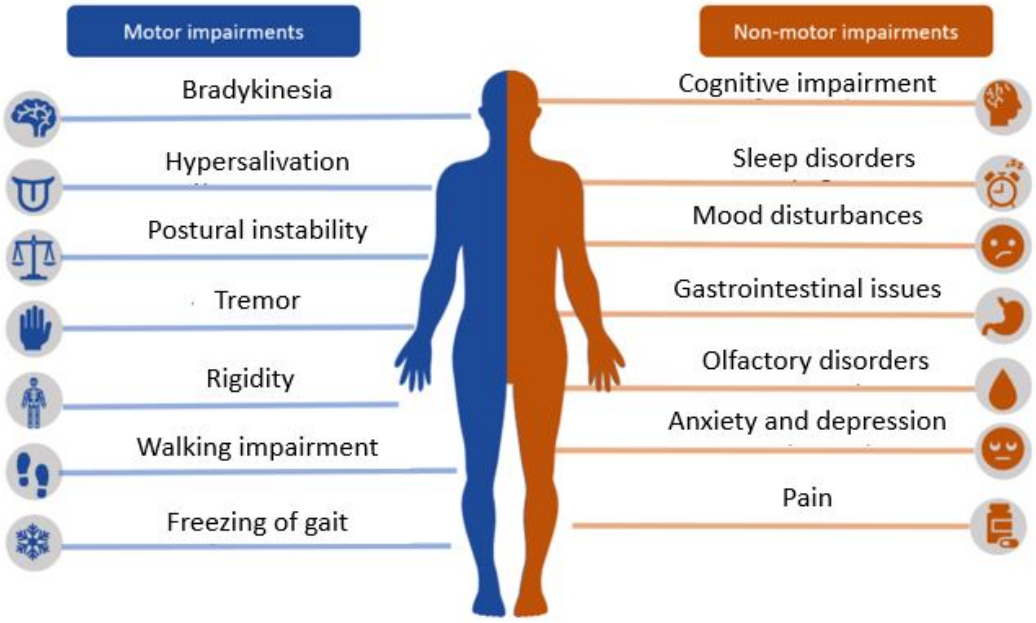


Fig. 1.2 Motor and non-motor aspects of Parkinson's disease (adapted from [16]).

Also, flexed posture and motor blocks have been included among classic characteristics of Parkinsonism, with PD being the most common form [2]. Non-motor

[2]component of a multiprotein E3 ubiquitin ligase complex which in turn is part of the ubiquitin-proteasome system

symptoms are often present before the diagnosis, impair with disease progression, and contribute to severe disability, impaired QoL, and reduced life expectancy. Non-motor symptoms are often not well recognized and consequently, inadequately treated [17]. Some of them (e.g., depression, constipation, pain, and sleep disorders) can be improved with nowadays treatments, while others need the administration of non-dopaminergic drugs.

Motor

Tremor is an involuntary, rhythmic muscle contraction leading to shaking movements in one or more parts of the body. In most cases, tremors are prominent in the distal part of an extremity. Tremor frequency ranges from low (4–5 Hz) to high (8–10 Hz)[18]. Clinically, tremor is observed in 75% of PwPD, and it can manifest in different forms. Tremor at rest is one of the PD cardinal signs. It is often asymmetric and it shows moderate amplitude and frequency ranging from 4 Hz to 6 Hz. It is caused by an agonist-antagonist alternate contraction pattern [19], and it typically disappears with action and during sleep [2]. Reemergent tremor represents an action tremor that manifests a few seconds after the transition from rest to posture and has a frequency content similar to that of a rest tremor. Postural tremor is more prominent and disabling than rest tremor and sometimes is the first manifestation of PD [20]. Its pathophysiology is linked to altered activity in the basal ganglia circuit, which is affected by dopamine neurons breakdown, and the cerebellum-thalamo-cortical circuit, which is also involved in many other tremors [18]. Tremor is not correlated with other PD motor symptoms, such as bradykinesia and rigidity, and several studies indicate tremor to be a marker of benign PD [21]. When the drug therapy is not effective, both thermocoagulation and deep brain stimulation provide good to excellent tremor control [18].

Rigidity is defined as an increased resistance during passive movement of a limb [22] and is one of the principal sign of PD. PD rigidity is marked by an increased muscle tone, increased resistance to stretching and reduced distension to passive movement [23]. The increase of rigidity during voluntary movement of other body parts and during stretching is a feature that helps differentiate PD rigidity from spasticity, with this latter worsening during fast displacement [24]. There is no proven correlation between dopamine deficiency and rigidity [23]. The mechanism underlying PD rigidity may include changes in the passive mechanical properties of

joints, muscles, and tendons, and abnormalities in peripheral sensory inputs that may influence the response to muscle stretch [25].

Bradykinesia represents the slowness of movement and is the symptom that better correlates with dopaminergic deficiency in PD [26]. Akinesia and hypokinesia refer respectively to poor spontaneous movements (e.g., in facial expression) or associated movement (e.g., arm swing during walking), and to the low amplitude of the movement itself [27]. Potentially, bradykinesia could be due to slowness in programming or executing movements [28]. However, PwPD demonstrate intact motor programming capability, yet they have difficulties in movement execution without an external trigger, which may be a loud noise or a visual cue requiring them to overcome an obstacle [2]. Although reduced muscle strength as well as other PD motor symptoms may contribute, the principal deficit seems to be the insufficient recruitment of muscle fibers during the initiation of movement [27]. In current clinical practice, the assessment of bradykinesia is carried out by observing the slowness and amplitude of movements during the execution of rapid, repetitive, alternating movements of the upper and lower limbs [29].

Postural instability is due to the loss of physiological postural reflexes and often manifests in the advanced stages of PD, after the onset of other clinical signs [30]. Together with motor blocks, it is the most common cause of falls and leads to a high risk of hip fractures [31]. It also may contribute to limitations in gait and decreased mobility [30]. Many studies have indicated rigidity, dystonia, abnormal spatial cognition, abnormal processing of proprioceptive signals [32], and side effects of pharmacological treatments as causes of postural instability [33]. The abnormal postural response is clinically evaluated during the pull test, where the clinician quickly pulls the patient backward by the shoulders and quantifies the degree of retropulsion as follows: if the patient takes more than two steps backward or does not show postural response, this indicates an abnormal postural response [34]. Although pharmacological (dopaminergic therapy) and surgical (deep brain stimulation) treatments may improve some axial signs, usually, they do not robustly improve postural instability [35].

Freezing of gait (FOG) is a form of paroxysmal akinesia (i.e., loss of movement) that affects gait in more than half of PwPD [36]. It is defined as a "brief, episodic absence or marked reduction in forward progression of the feet despite the intention to walk" [37]. FOG is quite heterogeneous in terms of clinical phenomenology

(e.g., shuffling steps, trembling legs, or complete akinesia) [38], duration of a single episode (with half of the episodes lasting less than 5 s and 90% lasting less than 20 s) [39], and triggering factors including environmental circumstances (e.g., turning, gait initiation, tight spaces) [38, 40–42], cognitive challenges (e.g., dual tasking) [43] and emotional stress (e.g., anxiety) [44]. FOG represents one of the most challenging and disabling symptoms in PD [37, 45] as it increases the risk of falls [43, 46] and is an early predictor of shorter survival [47].

Non-motor

Depression is a very common condition in PwPD, with a prevalence ranging from 20% to 70% [48, 49]. Depression in PD manifests as loss of interest, decreased energy, motivation, and appetite, sleep disorders, sadness, and suicidal thoughts [4]. Prevalence of anxiety disorders is higher in PwPD than in age-matched controls [50] and seems to be related to motor fluctuations (i.e., change of patients' motor condition based on the therapy effect) [51]. Dementia may manifest in the late stages of the disease. Although PwPD demonstrate cognitive slowing and reduced memory recall, recognition memory remains intact [52]. Psychosis and visual hallucinations are common, dose-dependent side effects of dopaminergic medications, exacerbated by the disease progression and co-morbidities [53]. Among the risk factors, advanced age, the presence of dementia, and polypharmacy are the most common. Sleep disturbances are frequent in PwPD [54]. Due to depression and/or hallucinations, rest at night is compromised, and patients have difficulty falling asleep. Higher risks for pathologic sleep include male gender, cognitive impairment, advanced stage of the disease, and drug-induced psychosis [55]. Orthostatic hypotension causes position-related dizziness, which often leads to falls in PwPD, fatigue, or even fainting. This symptom may not manifest as a major problem until late stages [56]. Dopaminergic therapies could not be effective and may even worsen the symptom [57]. Alterations in voice and speech occur in approximately 75–90% of the PD population, with voice and prosody being among the earliest indicators of PD [58, 59]. Gastrointestinal symptoms are diffused in PD. Dysphagia represents a major risk for *polmonite ab ingestis*, which is one of the principal causes of death in PwPD [60]. Constipation is the most common problem and often is one of the earliest signs of PD, manifesting even 20 years before the appearance of the motor symptoms [61]. At least 60% of PwPD suffer from constipation as compared with

21% in age-matched non-PD patients [62]. Nausea affects many PwPD suffering from the side effects of L-Dopa or other PD pharmacological treatments. Irregular peristalsis is due to the stimulation of dopaminergic receptors, especially those of the gastrointestinal tract [4]. About 30% of patients experiences incontinence, due to bladder dysfunction [63]. The severity of this latter is correlated with the progression of the disease [63]. Finally, patients, particularly those taking dopaminergic therapy, may become obsessive and compulsive in gambling, spending, shopping, or even sex [4].

1.1.2 Diagnosis and follow-up

Different non-motor dysfunctions are known to be predictive markers of PD (Figure 1.3 left). However, the diagnosis of PD is mainly clinical and based on the presence of cardinal motor symptoms (Figure 1.3 center). The clinical diagnosis is confirmed

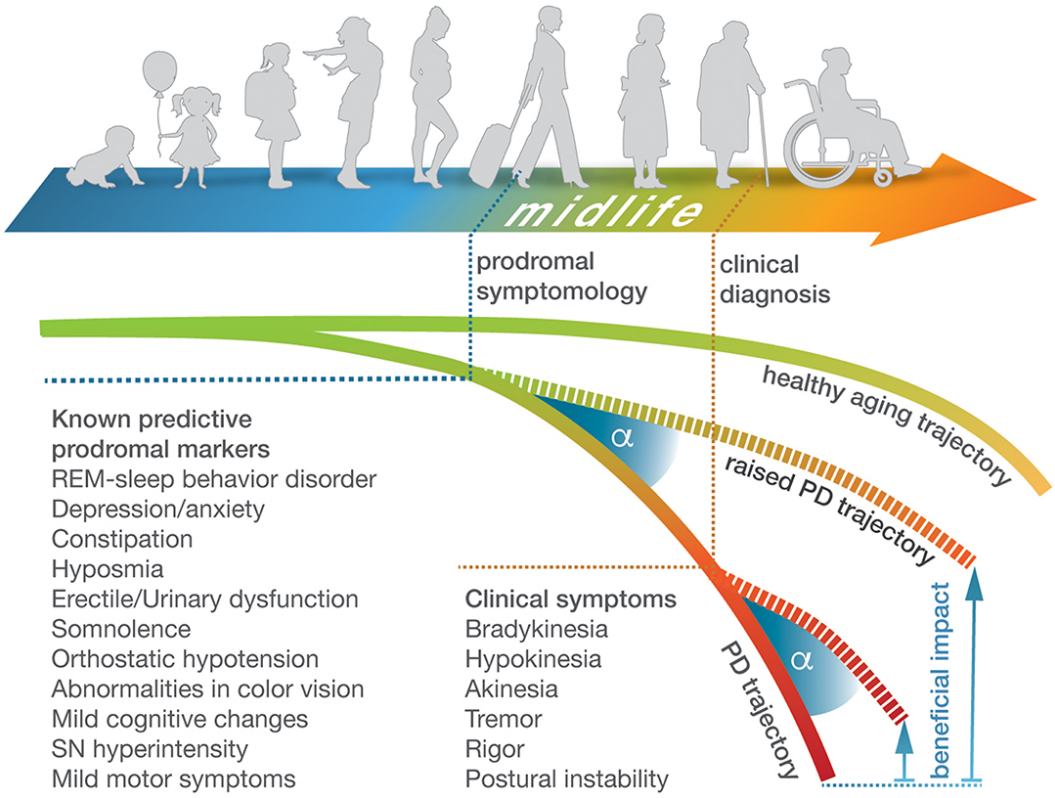


Fig. 1.3 Progression and diagnosis of Parkinson's disease (from [64]).

by an effective response to the dopaminergic therapy [65]. Some imaging tech-

niques are used to differentiate PD with motor symptoms from other disorders [66]. Dopamine transporter single-photon emission computed tomography identifies the presynaptic dopamine neuronal dysfunction present in PD and other Parkinsonisms. This is done by demonstrating reduced uptake of a radioactive tracer that binds to dopamine transporters in the basal ganglia. Dopamine transporter scans are generally useful only when the clinical diagnosis is uncertain. Functional brain imaging with positron emission tomography and the radiotracer 18-fluorodopa is capable to quantify the deficiency of dopamine synthesis and storage within pre-synaptic striatal nerve terminals. Therefore, it ascertains the diagnosis of PD in early disease stages and allows differential diagnosis between PD and other movement disorders. Additionally, this imaging technique is useful in the follow-up of the disease progression, the assessment of medical and surgical PD therapy strategies with possible neuroprotective properties, and the detection of pre-clinical disease in subjects at risk for the disorder [67].

1.1.3 Clinical scales

Hoehn and Yahr scale

The Hoehn and Yahr scale (H&Y) [21] is a widely used clinical rating scale for PD, which determines large categories based on the level of motor impairment. It is simple and of easy application, capturing typical features of progressive motor impairment, and providing a general assessment of disease progression. It ranges from stage 0 (no signs of disease) to stage 5 (wheelchair-bound or bedridden unless assisted). Figure 1.4 schematically reports the scale and briefly describes each stage. H&Y stages have been found to correlate with the progression of motor complications, reduction of QoL, and dopaminergic loss [29]. However, due to simplicity and lack of detail, it does not allow the evaluation of specific motor and non-motor aspects of the disease. The main focus of this scale is unilateral versus bilateral manifestations of motor symptoms and the impairment of postural reflexes. A modified version of H&Y is often used [68].

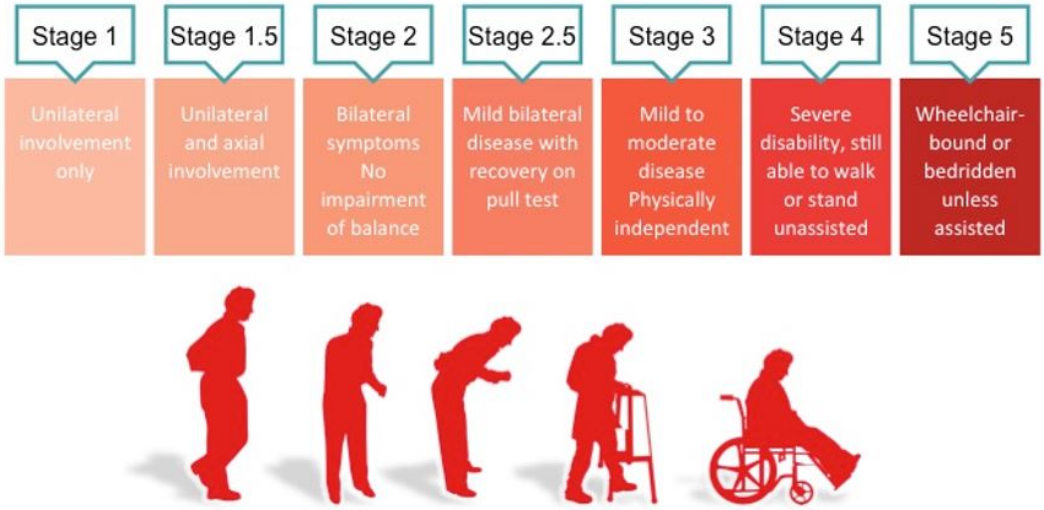


Fig. 1.4 Parkinson's disease progression as measured by the Hoehn and Yahr scale (from [69]).

MDS-UPDRS

The Unified Parkinson's Disease Rating Scale (UPDRS) [34] is the most employed scale in current clinical practice to evaluate several aspects of PD, including disability and motor impairment [29, 70]. It consists of a questionnaire divided into 4 parts: non-motor aspects, motor aspects, motor examination, and motor complications (Table 1.1). In each part, a score ranging from 0 to 4 is given to each item. Part I is a questionnaire including items related to cognitive impairment, hallucinations and psychosis, depressed and anxious mood, sleep disturbances, pain, fatigue, and other non-motor aspects. Part II focuses on motor aspects of PD, including speech impairment, eating, dressing, handwriting, and some mobility aspects (e.g., standing up from a chair and walking). Part III represents the clinical evaluation of the disease, where the clinicians ask the patient to perform a set of scripted tasks. These latter include some specific movements of the hands (e.g., pronation-supination, finger tapping) and legs (e.g., toe-tapping, leg agility), together with some overall mobility tasks (e.g., sit to stand, walk). Moreover, speech, facial expression, tremor, and other motor aspects and symptoms (e.g., bradykinesia, dyskinesia) are evaluated. Finally, Part IV aims to evaluate motor complications of the disease, such as dyskinesia, motor fluctuations, and dystonia.

Validity, reliability, wide utilization, application across all stages of the disease, and wide coverage of motor and non-motor symptoms make UPDRS the most established rating scale for PD [6]. Ambiguities, lack of enough detailed instructions [71] and difficulty in detecting small changes [2] led to the necessity of the development of a new version of the UPDRS. In 2001, the Movement Disorder Society (MDS) commissioned a revision of the scale, resulting in a new version, termed the MDS-sponsored UPDRS revision (MDS-UPDRS). The new scale had to keep the strengths of the previous one, modify ambiguous items, and add some new items to cover a wider spectrum of features and aspects of the disease [29].

Table 1.1 Unified Parkinson's disease rating scale sections and their description.

Part	Domain	Description
I	non-motor experiences of daily living	questionnaire
II	motor experiences of daily living	questionnaire
III	motor examination	clinical evaluation
IV	motor complications	questionnaire and clinical evaluation

1.1.4 Treatment

Current medical and surgical therapies for PD are symptomatic and lack significant disease-modifying effect [3]. At present, there is no proven neuroprotective therapy, and only symptomatic treatments are available [5]. An overview of the available pharmacological and surgical treatments is provided in the following.

Pharmacological

Until fifty years ago, ablative surgery to the controlateral thalamus was used in patients with severe tremor. Surgical treatment has then been replaced by L-Dopa, a dopamine-replacement therapy highly effective in improving the symptoms of the disease. L-Dopa (L-3,4-dihydroxyphenylalanine) is the metabolic precursor of dopamine and is combined with carbidopa[3] to limit the induced side effects (e.g., nausea) and maximize L-Dopa transport into the central nervous system [72]. L-Dopa-based treatment leads to a significant improvement in both the QoL and

[3]aromatic amino acid decarboxylase inhibitor

life expectancy of patients, although it does not revert the disease progression [73]. Although it represents the most effective drug treatment for PD [5, 74], its effectiveness decreases with disease progression [75]. Moreover, long-term therapy frequently leads to severe side effects. L-Dopa-induced dyskinesias are the most common ones, occurring in more than 50% of PwPD within 10 years of L-Dopa treatment [76], with pulsatile administration and higher doses being sources of motor fluctuations. For this reason, continuous dopaminergic delivery is preferable in order to minimize motor complications in PD [77]. The time of emergence of adverse events strongly depends on the severity of dopaminergic neuron loss at the introduction of L-Dopa [78]. Dopamine agonists[4] and monoamine oxidase (MAO)-B inhibitors[5] relief motor symptoms of PD and lead to a low risk of motor complications [79]. The MAO-B inhibitor selegiline, if administrated early in the course of PD, has been shown to improve PD motor symptoms and activity of daily living (ADL) score, and the effects persist up to 7 years or more [80]. Although dopamine agonists are not as effective as L-Dopa, they have demonstrated a reduced risk of dyskinesias [81], and this may be related to their longer half-life, compared to L-Dopa [3]. The choice of therapy has to take into account several factors, including the age of the patient, their compliance, the presence of cognitive impairment, additional medical conditions, and the degree of tolerance of the treatment. Treatment is carried out in the initial stage to ameliorate symptoms and allow the patient to be fully independent. Since the average life expectancy, from diagnosis to death, in PwPD is 17 years [73], a long-term treatment strategy is needed for most patients, and this should be discussed with the patient at early stages.

Surgical

Deep brain stimulation (DBS) is the most common surgical therapy for motor complications in advanced stages of PD and has been demonstrated to be effective in symptomatic PD therapy [82]. Benefits provided by DBS are more constant and predictable, compared with pharmacological therapy [83]. Stimulation of the subthalamic nucleus was found to reduce motor symptoms of PD, such as dyskinesias, bradykinesia, akinesia, and tremor [84, 85]. DBS requires an electrode to be inserted through the skull to stimulate the globus pallidus, subthalamic nucleus, or thalamus.

[4] compound that activates dopamine receptors

[5] class of drugs that inhibit neurotransmitter degradation, including dopamine, thus increasing their concentration in the central nervous system

A device, similar to a pacemaker, is implanted under the skin and wires connect the device to the electrode (Figure 1.5). Effective frequency of stimulation is recommended to be above 100 Hz [86]. The most evident results of this therapy are the reduction of OFF time (i.e., the time when the therapeutic effect vanishes), increased ON time (i.e., the time when the therapy is effective, thus ameliorating PD motor symptoms) without dyskinesia manifestations, reduction of required L-Dopa dose [83], and improved tremor [3].

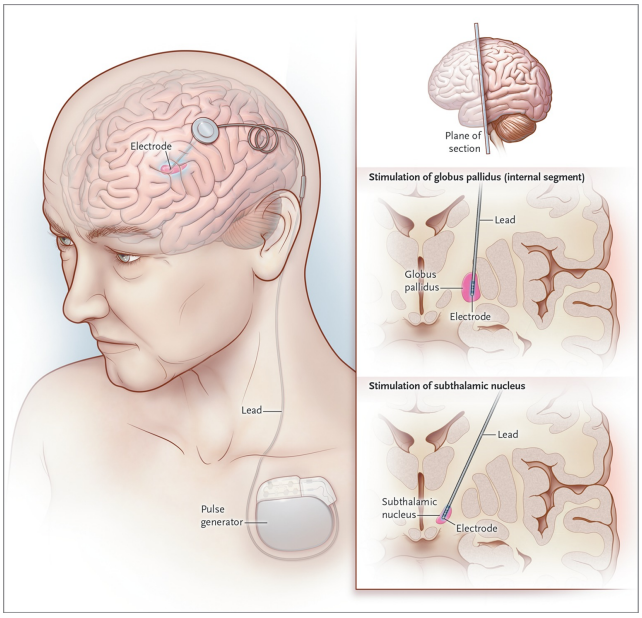


Fig. 1.5 Deep brain stimulation device and electrode location (from [87]).

Rehabilitation

Rehabilitation aims at maximizing motor and cognitive functions and minimizing secondary complications, in order to optimize the independence, safety, and QoL of patients [88]. Several rehabilitative approaches have been proposed, such as 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 standard pharmacological or surgical treatments. Nevertheless, physical

exercise is generally accepted as an adjuvant, because it has also a positive impact on non-motor symptoms, such as mood and depression [89].

1.1.5 Limitations of the current diagnostic and monitoring approaches

The diagnosis and management of PD are very complex for several reasons. First, some non-motor signs of PD can manifest up to 20 years before the clinical diagnosis (Figure 1.6), and they are rarely recognized by patients. However, the early diagnosis of the disease is fundamental for informing patients about their prognosis and initiating proper therapies where appropriate. This is particularly important nowadays, when disease-modifying drug trials provide hope for intervening in the earlier stages of the disease. The journey from the onset of motor symptoms to the diagnosis of PD depends on several aspects [90]. First, patients must be aware of the abnormal symptoms, prompting them to seek advice from their primary care physician. Second, the medical staff must recognize the reported symptoms as possibly related to PD and recommend a medical visit to the movement disorders specialist. Finally, the specialist must make a timely and accurate diagnosis, with or without the help of further investigations (e.g., medical imaging). Each of these steps may be influenced by a multitude of factors [91].

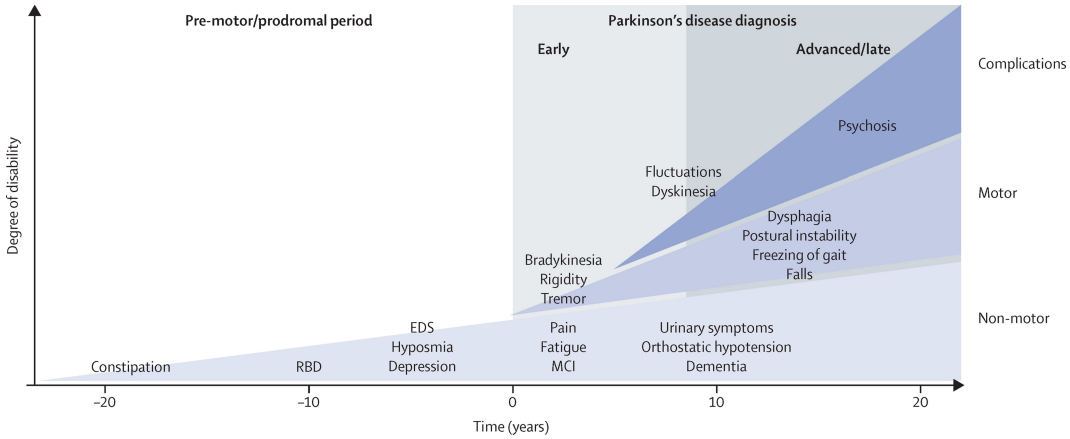


Fig. 1.6 Progression, symptoms, and complications of Parkinson's disease as the disease progresses (from [64]).

Moreover, once the disease is correctly diagnosed, some aspects of PD (e.g., motor fluctuations, FOG) are still difficult to appreciate in a medical office. Outpa-

tient visits are scheduled once or twice a year and have a limited duration, hence only gross variations are appreciated. Moreover, the visit itself may affect the actual patient status, which is conditioned by the time interval elapsed since the last drug administration, the general health conditions, emotional stress, attention [92–94], and many other subtle factors [95]. This makes it hardly possible for the neurologist to appreciate short-term variations in order to plan fine adjustments of the pharmacological treatment. In order to get further insights into the symptoms and their severity in daily living, clinicians can ask patients to compile diaries, providing detailed information regarding several aspects, including therapy effect, motor fluctuations, adverse events, presence and severity of FOG, and possible falls. However, the drawbacks of a diary include the subjectivity of the reports, their poor reliability [96, 97], and the fact that only patients with intact cognitive functionality can keep diaries, thus excluding most patients in later stages of the disease. Thus, diaries should be combined with other assessment methods. The inter-rater variability in the evaluation of a patient's motor performance should also be considered [98, 99]. The possible disagreement in assigning MDS-UPDRS scores can be justified by the complexity of discriminating between adjacent classes, due to the different aspects the clinicians are required to evaluate. From these considerations, it turns clear that long-term observations, possibly carried on during daily life, could yield a significantly improved assessment of the disease. More reliable, objective and continuous measures are advisable to detect symptoms as soon as possible, help clinicians to make the correct diagnosis, and provide continuous monitoring during daily life.

1.2 Technologies for mobility assessment in Parkinson's disease

A large number of electronic and digital solutions are available for the assessment of motor performance in healthy individuals and subjects with different types of mobility impairment. The technologies can be divided into two subgroups: non-wearable systems (NWS) and wearable sensors (WS). Optoelectronic motion-capture systems and instrumented walkways are examples of NWS, while smart insoles and wireless inertial sensors are WS. While the former are mainly used in laboratory settings, the latter can be used in unsupervised environments and in free-living conditions [100, 101]. The advantages and disadvantages of these technologies are

summarized in Table 1.2 and described in the next sections. Section 1.2.1 provides an overview of the systems currently used for assessing motor impairment in PD, while Section 1.2.2 describes and discusses the type, characteristics, and location on the body of wearable inertial sensors, together with their use for monitoring of PD.

System	Pros	Cons
NWS	Repeatability Widely accepted as 'gold standard' Good at measuring position Accurate results over short distances	Long set-up time Requires controlled environment Expensive and cumbersome equipment Limited workspace Non suitable for real life monitoring Poor at measuring acceleration
WS	Low cost Useful to monitor longer and natural movement Environment independence Wireless Promotes autonomy and active role of patient Unlimited workspace Non invasive	Power consumption restrictions Complex processing algorithms Sensibility to noise and interference Possible measurement errors Imprecise position estimation

Table 1.2 Advantages and disadvantages of non-wearable technology (NWS) for laboratory-based evaluations and wearable sensors (WS) for free living monitoring.

1.2.1 Overview

Motion-capture systems digitally track the patient's movement and capture the three-dimensional movement of the body. The system, depicted in Figure 1.7, is made of a number of synchronized cameras, several reflective markers placed over the skin of different anatomical segments, and computer software that records and processes the markers' position over time. Specifically, the three-dimensional trajectory of each marker is reconstructed from the two-dimensional images acquired by the cameras. This system allows the accurate estimation of spatial-temporal gait parameters, as well as 3D kinematics measurements [102, 103]. The video system records images of the patient during motion and provides a degree of quality control of the motion-capture data. It allows the assessment of patients' movement from multiple angles simultaneously, leading to a complete understanding of its pattern. Vicon, Qualisys, Motion Analysis, and OptoTrack represent some of the commercially available optoelectronic systems. In particular, the Vicon system provides a clinically validated solution in any gait analysis or rehabilitation environment, across different applications, like stroke rehabilitation, posture analysis, and balance studies. For

these reasons, it is the most common motion capture system [104]. To date, analysis with these optoelectronic systems is widely accepted as the *gold standard*, due to the accuracy in measuring position, as well as in providing well-quantified, repeatable, and accurate results over short distances. However, they require the markers to be within the field of view of the cameras, thus reducing the space where the motion can be captured [105, 106]. Moreover, they present a very long set-up time for marker positioning, leading to expensive and cumbersome equipment attached to the body. Finally, they are very expensive and they can be used only in laboratory settings.

Fig. 1.7 Schematic of motion capture systems including markers positioned on the subject's body and cameras (from [107]).

Cameras can be low-cost optical body-tracking sensors, with the potential to assess both healthy and pathological gait, posture, postural instability, and balance in a non-invasive way [108]. In particular, RGB cameras embedding depth sensors have become a ground-breaking vision-based motion capture system, finding application in different contexts, including medical–clinical and rehabilitation settings [109]. Acknowledged as non-intrusive tracking devices, cameras do not require subject preparation, attachment of markers to the patient's body or a dedicated handheld controller [109]. In fact, their inherent technology is able to detect and capture body movements in real time. This is done by estimating the position of the principal joints through the anatomical landmarks of a skeletal model in the 3D space [110] (Figure 1.8). As they do not require any additional equipment, subjects are free to move with their natural patterns while they perform various tasks inside the device's field of view, and their movements can be reproduced in real-time on the computer screen,

for example, to obtain visual feedback. Recently, several studies have investigated the accuracy and effectiveness of Microsoft Kinect for the assessment of posture, gesture, and gait performance in several pathological conditions, including PD [111]. The measures provided by such a device were found to be reliable for the assessment of spatial-temporal gait parameters [112, 113] and kinematic variables (e.g., trunk angle) in healthy individuals, with results comparable to those of laboratory-grade systems [109].

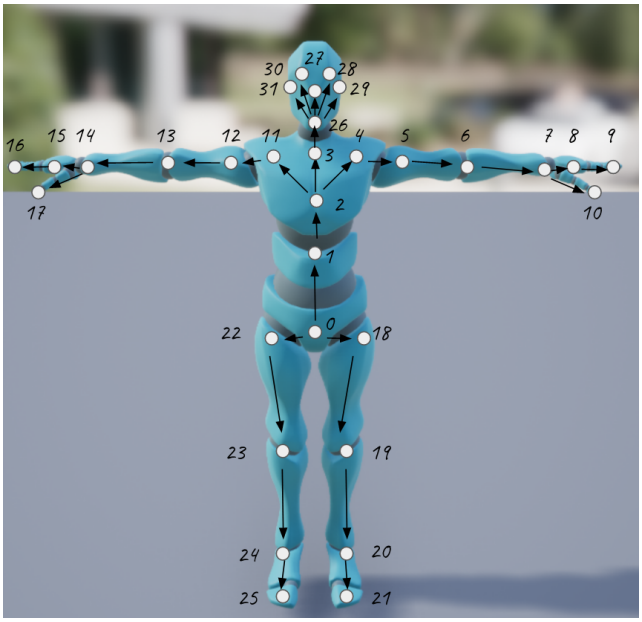


Fig. 1.8 Human body 3D reconstruction obtained from the processing of data from RGB-D camera (from [114]).

Force sensors are used to measure gait-related spatial-temporal parameters, as well as force and accelerations. They are commonly placed along the floor on some force platforms or on a single instrumented walkway, where gait is measured by pressure or force sensors when the subject walks on them [115]. Force plates measure the downward force, braking, acceleration, and the force directed medial-laterally [116]. While force sensors measure the force transmitted to the floor when walking (ground reaction force), pressure plates compute the evolution of foot pressure on the floor in real-time, which may reach up to 120%-150% of the patient's body weight in its maximum expression, when the heel touches the floor. If used individually, these devices are basic and can be used to obtain a gross evaluation of gait problems. However, when the information produced by these systems is integrated with that

obtained from motion capture systems, a robust, fine, accurate, and global evaluation of the movement and its mechanism can be performed [105].

Wireless electrogoniometers are electrical devices used to assess the flexibility and mobility of human joints. They measure angular positions through different planes of motion via attachments positioned at various joints of the body. They monitor the transduction (i.e., the conversion of voltage signals) in response to dynamic movements. Electrogoniometers can be obtained from potentiometers and strain gauges, which are devices that measure the axial rotation of a joint providing biofeedback on the range of motion. The utility of electrogoniometers contributes to diagnostic procedures in which patients possibly affected by musculoskeletal disorders become susceptible to debilitating joint impairments with the potential for permanent disability. In rehabilitation settings, the application of electrogoniometers can provide accurate results for physical therapists when making a comparative analysis of the extent of damage to an affected joint against baseline standards for improvements. Such devices have been used in PD for the assessment of rigidity [117, 118] and for the estimation of gait parameters [119].

Electromyography sensors (EMG) are devices that measure the electrical activity generated by muscle contraction. Needle EMG is an invasive electrodiagnostic technique used to accurately characterize muscle activity, and it is commonly used by neurologists. On the other hand, surface EMG (sEMG) is a non-medical procedure used to assess muscle activation by several professionals, including physiotherapists and biomedical engineers. The generated EMG signals can be analyzed to detect abnormalities, activation level, or to analyze the biomechanics of human movement. Wireless EMG sensors have been widely used in PD, for differential diagnosis [120], mobility assessment of the limbs [121], quantification of the effect of therapy [122], and the detection and analysis of FOG [123].

Smart pressure insoles are in-shoe systems embedding several pressure sensors, which measure the interaction between the foot and the shoe. They provide information critical to the understanding of gait mechanics and have a wide range of applications. In the PD context, they have been employed for FOG detection [124, 125] and gait analysis, with the objective of quantifying disease progression and daily monitoring for rehabilitation purposes [126]. Figure 1.9 depicts the pressure system (left), the positioning (center), and the plantar pressure data generated by the system (right).

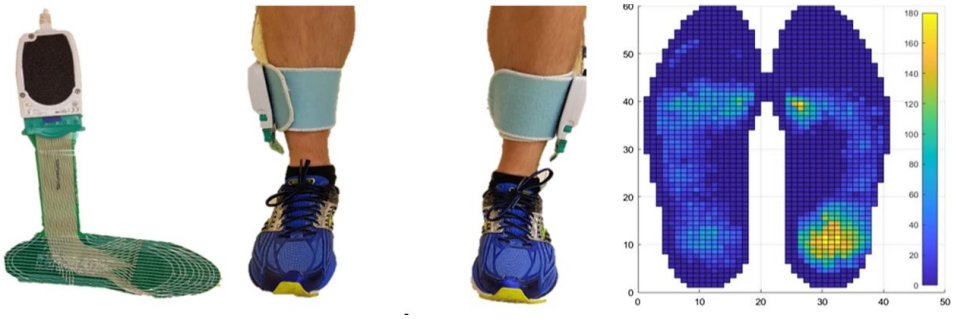


Fig. 1.9 Smart pressure insoles system, positioning, and generated foot pressure map (from [124]).

1.2.2 Wearable inertial sensors

Wearable motion sensors (WMS) have been widely used in the last two decades for assessing PD motor symptoms [127–130], following disease progression [131], and monitoring motor fluctuations [132]. This section describes and discusses the use of inertial sensor technology for monitoring PD. Specifically, the type of sensors and devices, technical characteristics, and sensors' location on the body are reported in the following.

Accelerometers are electro-mechanical devices that are able of measuring static and/or dynamic forces produced by acceleration. Static forces include gravity, while dynamic forces can include vibrations and movement. They can measure acceleration on one to three axes, with 3-axis accelerometers being the most employed. Gyroscopes can be schematically represented with a spinning wheel in which the axis of rotation is free to assume any possible orientation. During rotation, the orientation of such an axis is not affected by tilting or rotation of the mounting, following the angular momentum conservation law. Due to this principle, a gyroscope can provide the measurement of orientation and its rate of change.

Inertial measurement units (IMU) are stand-alone devices embedding inertial sensors (Figure 1.10). They commonly integrate a microcontroller, battery, and inertial units, including a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer. Moreover, the Bluetooth module and SD card allow to transmit data to other devices and save data locally, respectively. Finally, environmental sensors, such as temperature, humidity, and pressure sensors may be embedded in IMUs. Prototype or commercial IMU has been widely employed for monitoring all motor aspects of PD [133], both in laboratory settings [134] and in home environment [135].

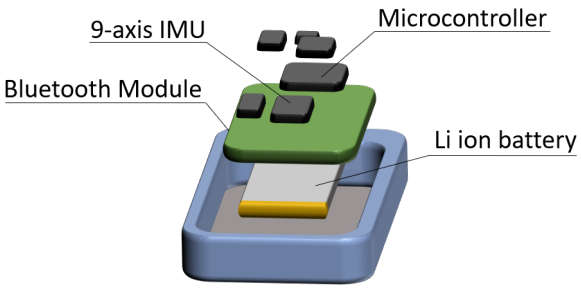


Fig. 1.10 Schematic representation of inertial measurement units.

Smartphones represent a widespread technology embedding several sensors, including motion sensors. Specifically, 3-axis accelerometer is used by apps to detect the movements of the device, as well as to allow features like shaking the phone to change music. The gyroscope is used in combination with the accelerometer to detect the rotation of the phone, for features like tilting the phone to play racing games or to watch a movie. The advantage of smartphones over other wearable devices is associated with their ability to capture and process data, transmit and receive data, and connect with other devices or sensors available in the physical environment. They have been largely used for motion analysis in several application domains, including human activity recognition (HAR) [136, 137], fall detection [138], and monitoring of PD motor symptoms [139–141].

The technical characteristics of motion sensing devices should be evaluated in terms of sample frequency, range, and resolution, in order to assess their suitability for specific data acquisition tasks. Human activity acceleration signals lay in the 0-20 Hz band [142]. According to the Nyquist sampling theorem, a sampling frequency of at least 40 Hz should be used for avoiding undesired aliasing effects. The amplitude of human motion ranges between $\pm$1g while walking and $\pm$2g during running [142, 143]. Thus, a range of at least $\pm$1g should be used for the recording task in order to avoid signal saturation. As far as concerns the gyroscope, it produces signals with a reduced frequency band, compared to accelerometers. Moreover, a range of $\pm$ 2000dps is more than adequate for motion analysis, in a large variety of tasks [144–146]. Most commercial motion devices have settable parameters, with a range from $\pm$2g to $\pm$16g for accelerometers and from $\pm$250dps to $\pm$2000dps for gyroscopes. Resolution of 8 to 16 bits is more than adequate for the analysis of common activities [140], and the sampling frequency is usually settable.

Different sensor locations are used for different purposes. The most frequent locations on the body are schematically reported in Figure 1.11, and discussed in the following. Tremor is commonly detected using sensors on the wrist [127], as PD tremor is usually marked in the upper extremities. When performing gait analysis, sensors on the ankles or shanks are commonly used, as they capture lower leg movements during walking [147]. Additional sensors on the thighs provide further insights into the human leg motion, allowing to accurately reproduce the movement pattern during locomotion [148]. A single sensor on the lower back can be used for gait parameters estimation, if combined with more complex processing algorithms [149, 150]. However, the accuracy on PD patients may vary, depending on the degradation level of the gait pattern and the presence of dyskinesia and FOG. Bradykinesia can be estimated using sensors placed on different parts of the body [151].

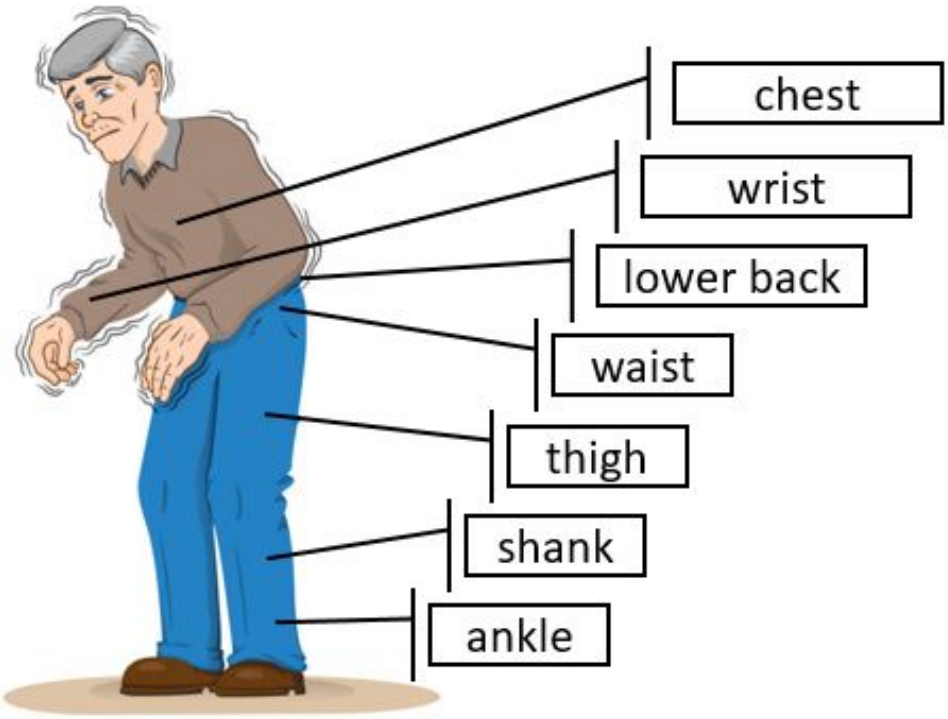


Fig. 1.11 Common locations of wearable inertial sensors on the body (from [152]).

Sensors on the upper or lower extremities may be used for assessing upper limb [153] or lower limb bradykinesia [98]. The same principle is applied to dyskinesia, with sensors placed on the upper or lower body for assessing upper-limbs or lower-limbs dyskinesia [154]. However, a single sensor on the lower back may be

suitable for detecting trunk dyskinesia [128] or providing a gross evaluation of limbs dyskinesia [155]. FOG is detected using a variety of sensors, sensors configuration, and locations on the body. Wrist [156], lower back [157], waist [135, 158, 159], thigh [160], shank [123, 161], or a combination of them [162] were all explored as possible body locations for FOG detection.

Applications for PD monitoring

FOG detection and prediction. FOG represents one of the most challenging and disabling symptoms of PD [45] and is difficult to appreciate in a medical office [37]. Automatic detection of FOG episodes has been extensively explored over the past 15 years, making use of WS [163]. The sensors used include commercial inertial measurement units [164] or prototypes [165], smartphones [157, 166], and individual accelerometers and/or gyroscopes [167, 168]. Experimental studies using gait analysis have shown that, in addition to episodes of FOG, patients with FOG are characterized by abnormal spatio-temporal gait parameters, such as slower and shorter stride length, greater spatial and temporal variability from stride to stride, and greater asymmetry between the mobility of both legs than patients without FOG [145, 169]. Specific spatio-temporal gait parameters (e.g., amplitude and step-to-step variability) progressively degrade until the occurrence of FOG, providing an opportunity to recognize typical pre-FOG periods (i.e., specific movement patterns that occur during actual gait just prior to FOG episodes). Recognition of pre-FOG periods by motion sensors would allow corrective strategies to be adopted to prevent or overcome FOG, such as the administration of external sensory cueing [170].

Postural stability. Postural stability (PS) is typically impaired in PwPD and worsens with disease progression [171, 172]. The difficulty in balancing the center of mass (COM) makes PwPD prone to the risk of falls [173]. However, in the early stages of the disease, PS is often difficult to assess during outpatient visits, so it is rarely used as a diagnostic criterion. On the other hand, the PS score is useful for monitoring disease progression. In fact, a marked postural impairment may denote a definite progression to severe disease conditions. PS is clinically assessed following the recommendations of the MDS-UPDRS part-III [71] (see section 1.1.1). However, it should be noted that the MDS-UPDRS retropulsion test represents a somewhat invasive measure, does not exclude the risk of falls for patients, and cannot be applied in uncontrolled environments. Postural control is highly related to

COM movements [173], and the latter is directly and quantitatively measurable by wearable (e.g., accelerometers and gyroscopes) or nonwearable (e.g., stabilometric platform) sensors [173]. Many studies in the literature have demonstrated the ability of WMS to quantify postural control in healthy and PD populations [131, 173]. A direct comparison between the results of stabilometric platforms, which are the "gold standard" for assessing PS, and WMS yielded a robust correlation between the two measures [174], paving the way for possible assessment of PD outside the laboratory setting. Further studies have demonstrated the possibility of discriminating PwPD from healthy controls [173], distinguishing patients with mild and severe disease progression [131], identifying individuals at high risk for falls [175], and monitoring disease progression [176].

Gait impairment. Gait is severely impaired in PwPD, and many gait cycle characteristics (e.g., single/double support, stance, swing duration; step and stride length, and their variability) are correlated with disease progression [177–179]. Consequently, the gait of PwPD has been addressed in many studies through WS. Inertial sensors have also been used to quantify fall risk from gait data recorded among the elderly [179, 180]. Among the various gait tasks performed during daily activities, many studies in the literature, as well as clinical experience, recognize turning as a preferred activity to detect motor impairment in PwPD. Turning requires the coordination of multiple limbs and continuous displacement of the COM. Spatially and temporally asymmetric stepping is required for each leg to travel a different distance while maintaining the same step time, and this provides information about the dynamic aspects of balance [181]. Not surprisingly, turning is closely related to FOG, PS, increased risk of falls and subsequent injuries, isolation, loss of confidence, and depression [180–183]. On the other hand, automatic assessment of turning is not trivial. Optical systems [183, 184] ensure repeatability due to the controlled measurement environment and lead to well-quantified and accurate results over short distances. Indeed, these systems are widely accepted as the "gold standard". However, their high cost, long installation time, and expensive and bulky equipment make their use impractical for frequent monitoring. In this context, WMS systems are an optimal solution for continuous monitoring of gait and turn under free-living conditions.

Bradykinesia. Bradykinesia correlates with disease progression and medication efficacy. Its clinical evaluation is performed following the MDS-UPDRS guidelines [71]. Specifically, patients are asked to perform a series of upper and lower-extremity

motor exercises and a short walking task. However, a point-in-time measurement such as that performed during outpatient visits does not provide any details about bradykinesia fluctuations in real-world scenarios. In this context, WMS sensors are useful for both providing objective measurements and long-term monitoring of bradykinesia during ADLs [132, 135]. Several studies in the literature have used inertial sensors for the quantification of bradykinesia, providing a comparison with current clinical assessment. The most widely used task for bradykinesia assessment is leg agility [185–187], followed by gait [188]. The results suggest that the sensors provide an accurate measure of bradykinesia, which correlates with clinical assessment.

1.3 Objectives and significance of the study

The present thesis describes a wide spectrum of computer methods and their application for monitoring PD motor symptoms. This study aims to overcome the limitations of both the current clinical monitoring approaches and the related works, providing new insights into several motor aspects of PD, and improving the performance of computer methods for PD monitoring. The main strengths of this study can be summarized as follows.

Device. In most of the experiments included in this thesis, the data collection process was conducted using a simple device as a commercial smartphone. This was done for several reasons. First, the use of smartphones avoids buying additional hardware, as they represent widespread technology embedding motion sensors. Moreover, participants are familiar with such technology, and this leads to increased acceptability of the hardware. Finally, smartphones can be used for data recording, processing, and transmission, and the development of a specific mobile application would allow interaction with patients.

Sample. A very large sample of PwPD was enrolled in this study. Overall, data from more than 200 PwPD and 20 elderly controls were analyzed to provide efficient solutions for the remote monitoring of PD.

Processing. The use of a single smartphone increases the complexity of the data analysis. To overcome this limitation, signal processing techniques were combined with machine learning (ML) and deep learning (DL) methods, providing new insights

into PD motor signs and incremental performance with respect to related work. Data processing involved the extraction of a large number of features, the use of different ML models, their optimization for maximizing performance, and the post-processing of the results. This latter step allows to interpret the results and propose future studies.

1.4 Thesis organization

The rest of the thesis is organized as follows. Signal processing, ML, and DL methods for the analysis of inertial data is provided in Chapter 2, together with an overview of the related work. This chapter serves both as a comprehensive description of the data analysis pipeline and as an introduction to the methods used in the present thesis. Chapter 3 describes materials and methods used in different PD monitoring applications. Each section of the chapter describes a different application of computer methods for PD monitoring, including FOG detection and prediction, assessment of postural stability, estimation of gait impairment, and evaluation of bradykinesia. Information regarding subjects, instrumentation, experimental protocol, preprocessing, and classification is carefully provided. The results are reported in Chapter 4, for each of the experiments included in this thesis. In Chapter 5, an overall discussion of the obtained results, their interpretation, potential, and limitations are carefully presented. Finally, in Chapter 6 conclusions are drawn and future studies are proposed.

1.5 Thesis related publications

Part of the methods and results of the present thesis is based on the following publications:

- **L. Borzì**, I. Mazzetta, A. Zampogna, A. Suppa, G. Olmo, et al. Prediction of freezing of gait in Parkinson's disease using wearables and machine learning. Sensors, 21(2):1–19, 2021. doi: 10.3390/s21020614.
- **L. Borzì**, L. Sigcha, D. Rodríguez-Martín, and G. Olmo. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural

networks and a single inertial sensor. Artificial Intelligence in Medicine, 2022. doi: 10.1016/j.artmed.2022.102459.

- **L. Borzì**, G. Olmo, C.A. Artusi, M. Fabbri, M.G. Rizzone, et al. A new index to assess turning quality and postural stability in patients with Parkinson's disease. Biomedical Signal Processing and Control, 62:102059, 2020. doi: 10.1016/j.bspc.2020.102059.

- **L. Borzì**, S. Fornara, F. Amato, G. Olmo, C.A. Artusi, et al. Smartphone-based evaluation of postural stability in Parkinson's disease patients during quiet stance. Electronics, 9(6):1–14, 2020. doi: 10.3390/electronics9060919.

- **L. Borzì**, I. Mazzetta, A. Zampogna, A. Suppa, F. Irrera, et al. Predicting Axial Impairment in Parkinson's Disease through a Single Inertial Sensor. Sensors, 22:412, 2022. doi: 10.3390/s22020412.

- **L. Borzì**, A. Manoni, A. Zampogna, F. Irrera, A. Suppa, et al. Correlation between wearable inertial sensor data and standardized Parkinson's disease axial impairment measures using machine learning. In IEEE 21st Mediterranean Electrotechnical Conference (MELECON), pages 732–736, 2022. doi: 10.1109/MELECON53508.2022.9843018

- **L. Borzì**, M. Varrecchia, S. Sibille, G. Olmo, C.A. Artusi, et al. Smartphone-Based Estimation of Item 3.8 of the MDS-UPDRS-III for Assessing Leg Agility in People With Parkinson's Disease. IEEE Open Journal of Engineering in Medicine and Biology, 1:140–147, 2020. doi: 10.1109/ojemb.2020.2993463.

Chapter 2

Machine learning for human motion analysis and classification

The term ML refers to computer programs that are capable of directly learning from the input data or past experience, without being explicitly programmed. This characteristic distinguishes ML from the more general term artificial intelligence (AI), consisting in intelligent systems that can simulate human intelligence. In the last two decades, the combination of WS and ML algorithms has demonstrated excellent performance in the detection [189] and analysis [190] of movement in PD. In the traditional (shallow) ML approaches, a process of feature extraction from the raw data is required, including the selection of the most representative features to solve a given problem [191]. These feature extraction and selection processes are required because ML models by themselves are not capable of learning from high-dimensional data in their raw forms (i.e., medical images or time series acquired by sensors) [192]. Among the ML algorithms, in the last years, the use of DL approaches has led to establishing the state-of-the-art in many domains and applications [193], and it has provided promising results in the automatic assessment of movement disorders [194]. Different from (shallow) ML methods, DL algorithms have the ability to extract high-level features directly from data. Thus, DL enables the development of end-to-end models, which decreases the time and effort required to design classical pipeline-based approaches, including selecting appropriate features [193, 195]. On the other hand, the large amount of data necessary for training DL algorithms, and the reduced model interpretability represent the main limitations to their wide application in the medical domain.

Figure 2.1 and 2.2 schematically report the processing pipeline of ML and DL algorithms. The former requires some processing steps to be performed prior to classification, including filtering, feature extraction, and feature selection. Using DL reduces the signal processing steps, as they automatically learn from raw data. However, some preprocessing procedures may be common to both approaches, aimed to remove noise from data or to split the input data into different data-frames. The principal preprocessing steps (e.g., signal processing, segmentation, feature extraction and selection) performed prior to ML classification are described in Section 2.1, while ML and DL classification algorithms are discussed in Sections 2.2 and 2.3, respectively.

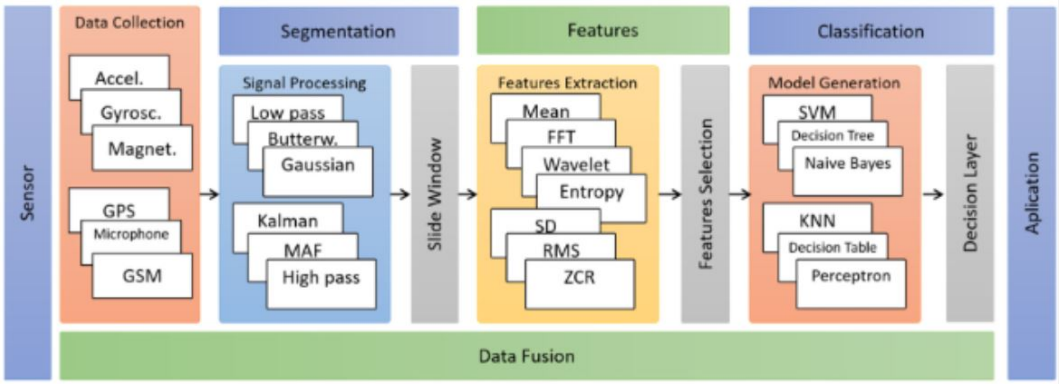


Fig. 2.1 Processing outline used by shallow machine learning algorithms (from [136]).

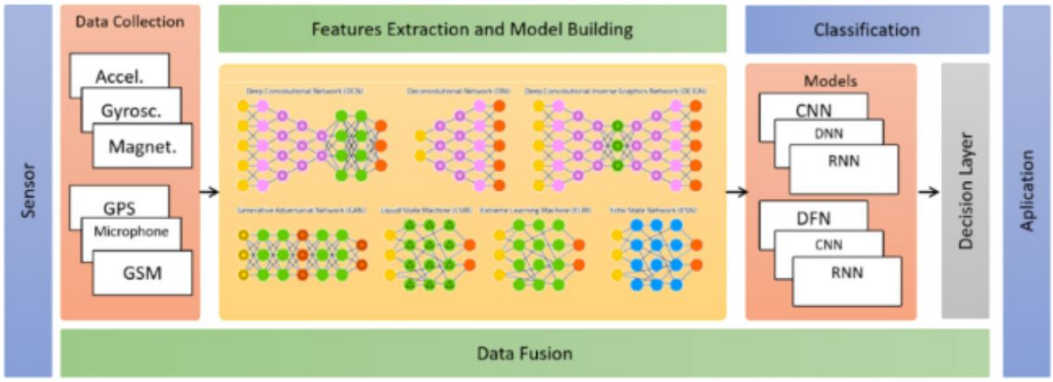


Fig. 2.2 Processing outline used by deep learning algorithms (from [136]).

2.1 Data preprocessing

Commonly, raw inertial recordings undergo some preprocessing steps, aiming to denoise the signal, segment data into pieces, possibly transform the original data, and extract and select the information that characterizes the data. A brief description of each processing step is reported in the following sections. A sample of raw accelerometer and gyroscope readings is provided in Figure 2.3 and 2.4, respectively, where different simple activities are performed, including walking, standing, turning, and standing up. Signals represent the data generated by inertial sensors embedded in a commercial smartphone, positioned on the lower back. As can be observed in Figure 2.3, the acceleration pattern produced by walking and standing can be clearly identified. However, postural transitions, such as standing up, are not clearly visible in the acceleration data, and the signals during turning can be confused with those during walking. Gyroscope recordings can provide complementary information, with postural transitions and turning producing a characteristic pattern in the angular velocity signal (Figure 2.4). On the other hand, the gyroscope's signals during walking are very noisy, and those during stance bring little information.

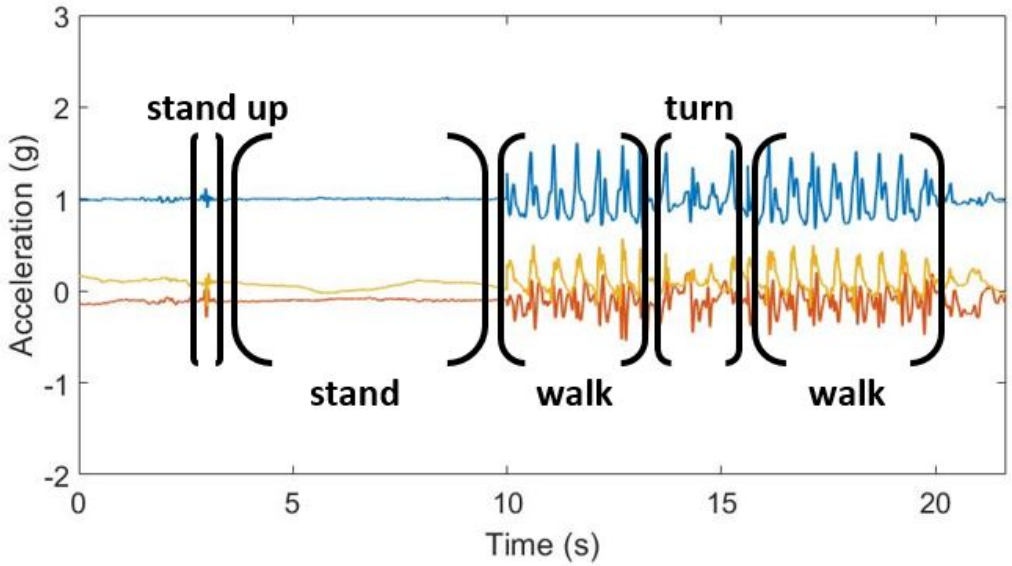


Fig. 2.3 Raw acceleration readings from a series of simple activities. The blue, orange, and yellow line represents the vertical, medio-lateral, and anterior-posterior directions, respectively.

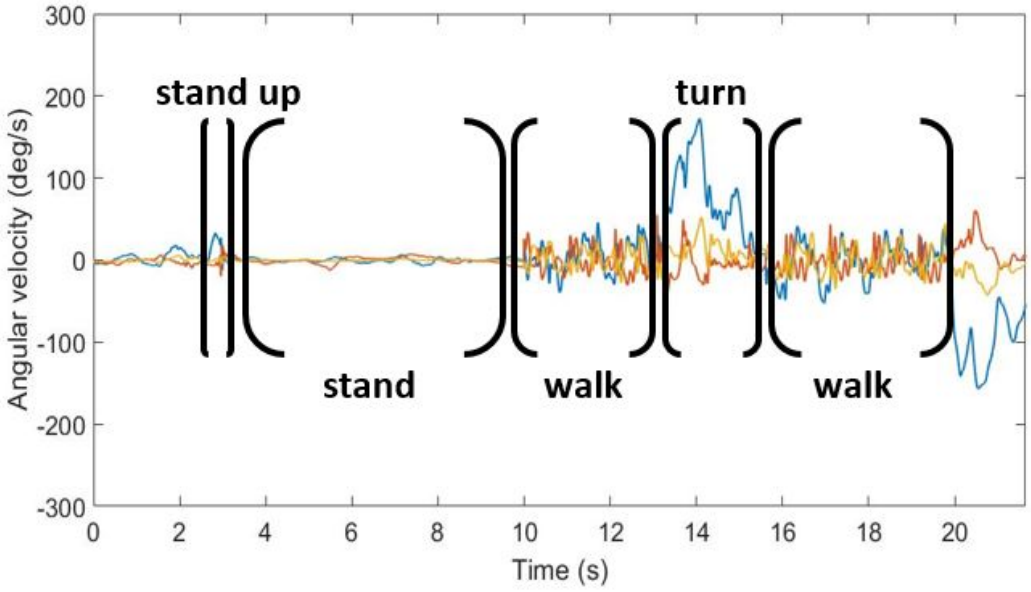


Fig. 2.4 Raw angular velocity readings from a series of simple activities. The blue, orange, and yellow line represents the vertical, medio-lateral, and anterior-posterior direction, respectively.

2.1.1 Resampling

When data are recorded with a high sampling frequency, down-sampling the data can be useful to reduce the computational time and burden of the subsequent processing steps, and useless computation. Given that human activity acceleration signals lay in the 0-20 Hz band [142], and the band of angular velocity signals is even smaller, a sampling frequency of 40-50 Hz is adequate, according to the Nyquist sampling theorem.

2.1.2 Data transformation

The observation of signals in the time domain does not provide a full picture of characteristics and patterns. Thus, data transformation methods can be used to provide different representations of the initial data. Fast Fourier transform (FFT) carries information complementary to that obtained in the original time domain signal. FFT converts a signal from its original time domain to a representation in the frequency domain. It is computed as in Equation 2.1, where X is the original signal, $W_n = e^{(-2\pi i)/n}$ is one of n roots of unity, and Y is the generated data. However, FFT

provides a static picture of the signal, not considering the evolution of patterns in the time domain. Moreover, the complex representation of the signal increase the computation burden.

$$Y(k) = \sum_{j=1}^{n} X(j) W_n^{(j-1,k-1)} \tag{2.1}$$

In this context, the continuous wavelet transform (CWT) allows a time-frequency representation of signals. The frequency content of the signals is generated for each time frame, producing a 2D (time vs frequency) data map. This approach is often used in gait analysis for walking steps detection [196, 184]. CWT uses inner products to measure the similarity between the signal $x(t)$ and an analyzing function, which is a wavelet $\psi(t)$. Equation (2.2) reports the formula for CWT computation. First, the wavelet is shifted by $b \in R$ values and stretched/compressed by $a \in R^+$ values, then the dilated and scaled versions of the wavelet $\psi^*(\frac{t-b}{a})$ is compared to the signal $x(t)$ in order to compute their similarity. This procedure is performed using a mother wavelet $\psi(t)$ and all possible values of a and b. Figure 2.5 reports a segment of acceleration signal (vertical axis) recorded during gait, together with the module of its FFT and CWT transforms.

$$X(a,b,x(t),\psi) = \int_{-\infty}^{+\infty} x(t) \frac{1}{a} \psi^* \left(\frac{t-b}{a} \right) dt \tag{2.2}$$

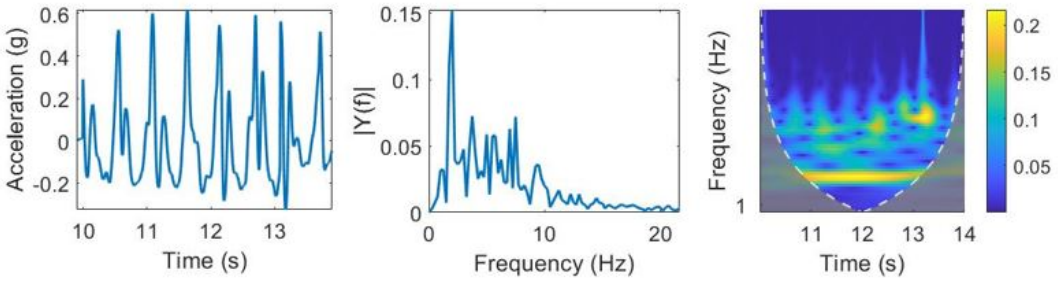


Fig. 2.5 Data transformation methods. An example of the original signal (left), its fast Fourier transform (center), and the continuous wavelet transform (right).

The selection of the analyzing wavelet depends on the type of signal features or patterns to detect. Although there is no definite relationship between scale (a) and frequency, this is generally inversely proportional. Specifically, a smaller scale

generates a compressed wavelet capable of catching rapidly changing details, thus allowing the analysis of high-frequency features.

2.1.3 Filtering

Raw inertial readings are usually filtered to keep only the frequency components of interest and remove undesired noise. First, gravity acceleration produces an offset in the acceleration recordings. This can be observed in Figure 2.3 as the mean value of the acceleration signals along each axis. Moreover, the sensor can slightly move from the original body location where it is attached, and this can generate undesired low-frequency trends. Finally, the frequency band of signals produced by human movement is between 0 and 20 Hz for acceleration [142], and much lower for angular velocity. Moreover, most of the energy of inertial signals during gait is included in the band from 0 to 3 Hz [197, 198]. High-pass filtering can be used to remove offset and low-frequency trends, while low-pass filtering can be performed to remove high-frequency noise.

2.1.4 Segmentation

Segmentation is a necessary processing step in many applications, including HAR [199], gait analysis [200], and monitoring of some PD motor symptoms, such as tremor [127] and FOG [130]. The segmentation process consists in dividing the initial signal into successive time-frames (windows); then, the information extraction process (Section 2.1.5) is performed on these windows. Figure 2.6 depicts gait acceleration signals.

In most cases, a static segmentation is used [127, 132]. In this case, the signal is divided into equal-size frames, with or without overlap between adjacent windows (Figure 2.6 left). In some cases (e.g., gait analysis), it can be more useful to perform segmentation based on events (e.g., steps). This generates non-overlapped windows of varying length, according to the time distance between adjacent events (Figure 2.6 right).

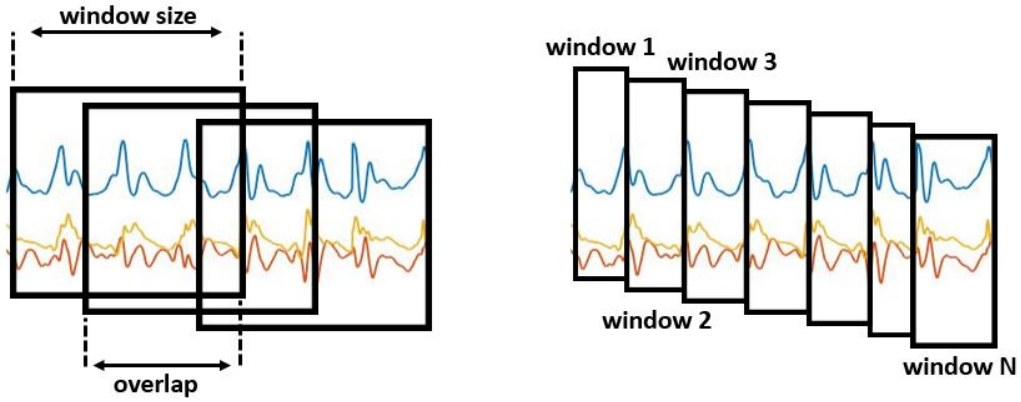


Fig. 2.6 Different segmentation procedures. Static segmentation produces fixed-length windows with a certain overlap (left), while dynamic segmentation generates non-overlapped windows of different sizes (right).

2.1.5 Feature extraction

The information can be extracted either from the entire signal or from the windows generated in the segmentation process. Feature extraction consists in identifying and computing a set of characteristics that best describe the signal. They can be extracted from the time domain or from data generated using the FFT or CWT transforms. For example, several types of features can be extracted to describe human gait. Well-known spatio-temporal gait parameters include stride and step time, stride and step length, and stance, swing, double stance, and mid-swing duration. Furthermore, it is possible to include simple features describing the amplitude of signals in the time domain (Figure 2.5 left), such as mean value, standard deviation, root mean square value, maximum, minimum, interval, and quartiles. For a more in-depth analysis of the gait pattern, it is possible to calculate the distance between peaks, the zero crossing rate, the number of peaks and the entropy, which describe the temporal evolution of the signal and its complexity. Furthermore, starting from the FFT of the signal (Figure 2.5 center), the frequency, amplitude, and width of the main harmonic allow an approximate evaluation of the frequency content. In addition, the number of harmonics, the ratio of the main harmonic energy to the total energy, and the spectral entropy can be used to describe the regularity of the signal. Most of the features described above can be used in a wide range of human motion analysis tasks involving repetitive movements, such as walking, finger tapping, and repetitive leg

movements. Depending on the specific application, other types of features can be included to better describe specific movement patterns.

2.1.6 Dimensionality reduction

Principal Component Analysis (PCA) aims to reduce the dimensionality of the generated feature set while preserving the maximum amount of information. PCA linearly transforms the data into a new coordinate system where most of the variation in the data can be described with fewer dimensions than the initial data. In order to make PCA work properly, the input variables should be independent and centered (i.e., the mean value should be equal to zero). First, the algorithm computes the covariance matrix for each feature pair (x,y), according to Equation 2.3. The covariance matrix (COV) has feature variances along the diagonal. Then, the COV is used to find the eigenvalues λ and the eigenvectors v satisfying the equations reported in Equation 2.4, where I represents the identity matrix.

$$COV_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N-1} \tag{2.3}$$

$$det(COV - \lambda I) = 0, (COV - \lambda I)v = 0 \tag{2.4}$$

Finally, the eigenvectors are sorted in descending order of the respective eigenvalues. The original feature set is multiplied by the matrix formed by the first k eigenvectors, obtaining a reduced feature set, including the projection of the original features along the principal components. The choice of k can be a-priori or based on the variance level to keep in the transformed feature set. The pseudo-code reported in Algorithm 1 describes the procedure returning k by setting the variance (*varSet*) to be kept by the first k components.

Algorithm 1 Choose k

```
k set k = 1, cumVariance = 0, setVariance;
while cumVariance < setVariance do
cumVariance = cumVariance + variance(k)
k = k + 1;
```

2.1.7 Feature selection

Feature selection (FS) represents the process of reducing the number of input features when developing a predictive model. Reducing the number of features reduces both the computational complexity of the model and the over-fitting. This latter occurs when a statistical model fits exactly against its training data. When this happens, the algorithm cannot perform accurately against unseen data, defeating its purpose. FS methods can be grossly divided into three categories, described in the following.

Filter approach. Features are selected before the ML algorithm is run, thus the selection is independent of the model. Minimum-redundance maximum-relevance algorithm [201] finds features that maximize the mutual information between the features and the response variables and minimize mutual information between the features themselves. Pearson correlation coefficient has also been used as a filter approach, by sorting features based on the correlation with the target variable [98]. Moreover, strong correlations among features may be used to remove redundant information [140]. Equation 2.5 describes the computation of Pearson correlation, where x and y represent the variables, COV the covariance matrix, and σ the variance.

$$r(x,y) = \frac{COV(x,y)}{\sigma_x \cdot \sigma_y} \tag{2.5}$$

Embedded approach. In this case, FS occurs naturally as part of the data mining algorithm. Artificial neural networks (ANN), support vector machine (SVM), and the ensemble of decision trees (DT) represent ML models that intrinsically estimate feature importance during the training process, promoting the most discriminative features and reducing the weight of weak variables.

Wrapper approach. This method uses the data mining algorithm as a black box to find the best subset of attributes. In this case, different feature subsets are input to the classification model, and that providing the best performance is finally selected. Sequential FS builds up a feature set until performance stops improving. While forward FS works iteratively by adding features to the subset, backward FS starts with the complete feature set and iteratively discards features.

2.1.8 Feature scaling

The range of values of features may widely vary. As some ML models embed distance calculation, features with a larger range mostly affect the final results, while features with a very small range are not properly considered. Therefore, the range of all features should be normalized so that the contribution of each feature is independent of the initial range of values. Moreover, optimization algorithms work better and faster with normalized data. Min-max normalization (Equation 2.6) consists in subtracting the minimum value and dividing by the range. This process maps features in the range from 0 to 1. Mean normalization (Equation 2.7) consists in subtracting the average value and dividing by the range. This process maps features in the range from -1 to 1. Finally, z-score normalization (Equation 2.8) consists in subtracting the average value and dividing by the standard deviation. This process generates features with zero-mean and unitary variance.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{2.6}$$

$$x' = \frac{x - mean(x)}{max(x) - min(x)} \tag{2.7}$$

$$x' = \frac{x - mean(x)}{std(x)} \tag{2.8}$$

2.2 Shallow machine learning algorithms

This section describes the most common supervised ML algorithms. Contrary to unsupervised learning methods (e.g., clustering), they require an annotation process performed prior to training. Such annotation task consists in adding labels or tags to pieces of data (e.g., identifying some activities in the data or marking the presence of clinical signs), which provide the algorithm information that is used to properly train the model for the specific task. Regression is used to predict continuous values and find a correlation between variables (e.g., the output of ML models and a continuous clinical score). On the other hand, classification aims to group samples into categories (e.g., control subjects versus PwPD or FOG vs non-FOG). The selection of a specific ML model depends on several aspects, including the objective, size of the data set, performance target, training and testing speed, and interpretability of the results

[202]. Moreover, the linearity or non-linearity of data may suggest the use of specific ML models over others. Table 2.1 reports the characteristics of some ML models, in terms of generalization capability, execution speed, and interpretability. In the following, some suggestions are provided regarding the selection of ML models. Finally, each classification algorithm is described in the subsequent sections.

Model	Robustness	Classification speed	Interpretability
Decision tree	low	high	high
Random forest	high	high	medium
k nearest neighbor	low	low	low
Linear regression	low	high	medium
Logistic regression	low	high	medium
Support vector machine	high	medium	low
Linear discriminant analysis	high	medium	low

Table 2.1 Comparison between different machine learning models in terms of generalization capability, speed, and interpretability.

Data set size. If the number of observations is lower than the number of features, then linear algorithms (e.g., linear regression, linear-SVM) perform well. On the other hand, if the training data is sufficiently large and the number of observations is higher than the number of features, algorithms like kNN, DT, or kernel-SVM usually perform better.

Interpretability. Interpretability refers to the easy understanding of how any individual predictor is associated with the response. Usually, the ML models providing the best performance are associated with low interpretability. Thus, the classification model should be selected according to the objective, aiming to maximize accuracy over interpretability or vice-versa, or choosing a trade-off. In medical applications, the interpretability of the model is of crucial importance. Indeed, clinicians find it hard to understand and trust complex models due to the lack of intuition and explanation of their predictions. In this context, there is an emerging trend of interpretability techniques aiming to shed light and provide insights into the prediction process of the ML models [203].

Speed. In real-world applications, the choice of the algorithm is driven by training time. Algorithms like linear regression (LR) and logistic regression (LogR) are easy to implement and quick to run. On the other hand, algorithms like SVM, ANN, and RF need a lot of time to train. However, once trained, these latter algorithms are

very fast at testing time, thus they are suitable for real-time applications. In contrast, the testing time of kNN is proportional to the training set size and may be very long.

Linearity. When classes are linearly separable, all the algorithms perform well. In this case, linear models (e.g., logistic regression or linear SVM) provide a linear boundary, which is more robust than more complex separation functions. One option to test data linearity is to train such linear models and observe the residual errors. In case of large errors, data is presumably not linear, and more complex ML models (e.g., ANN, kernel-SVM, RF) are required.

Decision tree and ensemble of trees

DT is probably the simplest and more interpretable ML algorithm. A DT is a flowchart-like tree structure, where each node indicates a variable, each branch represents a decision, and each terminal leaf node denotes a class label. DT is built from the root to the leaves (i.e., from top to bottom), where the most feature power, the closer the feature is to the root. The Gini index (GI) [204] measures the quality of each split, with $GI = 0$ when all observations belong to the same class, and $GI = 1$ when the distribution of the elements within classes is random. However, it provides weak generalization capability, as the decision boundaries separating classes are very sharp (i.e., single threshold). On the other hand, DT is known for its very good interpretability, as it makes it possible to follow the flow of elements from the input down to the class allocation. RF takes advantage of a large number of individual decision trees that operate as an ensemble. Each DT in the RF provides a prediction for each item, and the final output is computed using majority voting (i.e., the class with the larger number of votes becomes the final prediction. In order to work properly, the individual DTs should be independent (i.e., different from each other), so that they protect each other from their individual errors. When implementing DT, some parameters should be tuned, including the maximum depth (i.e., number of tree levels), the maximum number of decision splits, and the minimum number of observations per leaf node. When using RF, additional parameters should be specified, such as the number of DTs.

k nearest neighbor

k nearest neighbor (kNN) is a distance-based supervised classification algorithm. The prediction of a new element is obtained based on the distance from the elements of the training set. Let X_1 and X_2 be two variables (features) and class A and B the two classes (binary classification problem). Figure 2.7 depicts elements of the two classes and the new element (red) to be assigned to either class A or B. First, the distance between the new element and all the elements in the training set is computed. Then, the k nearest neighbors are selected as the k data points which are closest to the element. Finally, the element is assigned to the most represented class among the k neighbors. As can be observed in Figure 2.7, changing the number of neighbors affects the final classification results. Specifically, the new element is assigned to class B or A if $k = 3$ or $k = 6$, respectively.

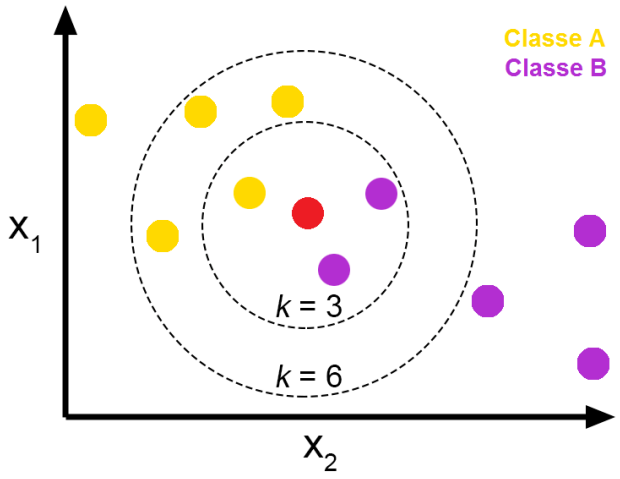


Fig. 2.7 k nearest neighbor classifier (from [205]).

Both the parameter k and the distance metric should be selected through the optimization procedure. Cityblock and euclidean distances are commonly used in the case of continuous variables. On the other hand, cosine similarity is used when dealing with categorical or binary values. The formulas for the computation of these metrics are reported in Equation 2.9,2.10 and 2.11, where x and y represent two elements, and N their dimensionality (N = 2 in Figure 2.7). kNN algorithms work better than other ML models when the input data are non-linearly separable. However, kNN does not provide any decision boundaries, thus the final prediction is severely affected by the addition of even a few elements to the data set. Moreover,

the algorithm is quite expensive, as it computes the distance of the element under test from all the other elements of the data set.

$$d_{cityblock}(x,y) = \sum_{i=1}^{N} |x_i - y_i| \tag{2.9}$$

$$d_{euclidean}(x,y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{2.10}$$

$$d_{cosin}(x,y) = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{N} x_i^2} \cdot \sqrt{\sum_{i=1}^{N} y_i^2}} \tag{2.11}$$

Linear and logistic regression

LR is used for solving regression problems, yet it can be adapted for classification problems using the discretization of the continuous output, which can be easily performed using a fixed threshold. Given the input matrix X, LR aims to find a linear relationship between the input data X and the target Y. Equation 2.12 reports the formula implemented by LR, which corresponds to the equation of a straight line, where w and b represent the weight term and the bias, respectively. The best fit line is computed by minimizing the least squares error between the prediction h_θ and the target Y. Specifically, the algorithm finds the best parameters θ by minimizing the loss function reported in Equation 2.13, where m is the size of the input data X, θ is the matrix of parameters (i.e., weights of features) and λ is the regularization parameter.

$$Y = wX + b; \tag{2.12}$$

$$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m} (h_\theta(x(i)) - y^{(i)})^2 + \lambda \sum_{i=1}^{n} \theta_j^2], h_\theta = \theta^T X \tag{2.13}$$

LogR is a widespread algorithm used for solving classification problems. The loss function and the hypothesis function for LogR are reported in Equation 2.14. It can be noticed that, differently from LR, the hypothesis function is a sigmoid rather that a straight line. The best model parameters are computed by minimizing the loss function, using the maximum likelihood method. When implementing LR and

LogR, a few training parameters should be selected, such as the regularization term (λ). Specifically, a larger value of parameters promotes a better fit of data over the generalization capability of the algorithm.

$$J(\theta) = -[\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}log(h_\theta(x(i))+(1-y^{(i)})(1-h_\theta(x^{(i)}))]+\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2, h_\theta = \frac{1}{1+e^{-\theta^T X}} \tag{2.14}$$

Support vector machine

SVM is a supervised ML algorithm used for both classification and regression. SVM aims to find a hyperplane in an N-dimensional space that distinctly separates the classes. The dimension of the hyperplane is proportional to the number of features. The hyperplane is selected as the decision boundary ensuring the largest separation (margin) between classes (Figure 2.8 left), and this represents the main strength of SVM. Given the input data X, model parameters θ, and prediction y, the loss function J is computed as in Equation 2.15, where $h_\theta = \theta^T X$. Different of the loss function of other models LR and LogR, in SVM the penalization occurs both on incorrect predictions and on those data points which are close to the decision boundary. This latter aspect ensures margin maximization. Moreover, SVM can exploit different types of kernel functions for mapping the input data into a higher-dimensional feature space (Figure 2.8 right) allowing the distinct separation of the input classes with a linear hyperplane. The following parameters can be optimized during the training-validation procedure. The regularization term (λ in Equation 2.15) controls the tolerance level for misclassification errors. Specifically, smaller values of λ promote margin maximization and allow a larger number of misclassifications. The kernel function could be linear, polynomial (e.g., quadratic, cubic), or gaussian. Finally, the kernel scale controls the exact shape of the kernel function (e.g., narrow or wide gaussian function).

$$J(\theta) = \frac{1}{\lambda}[\sum_{i=1}^{m} y^{(i)} max(0, 1-h_\theta(x(i)) + (1-y^{(i)}) max(0, 1+h_\theta(x(i))] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 \tag{2.15}$$

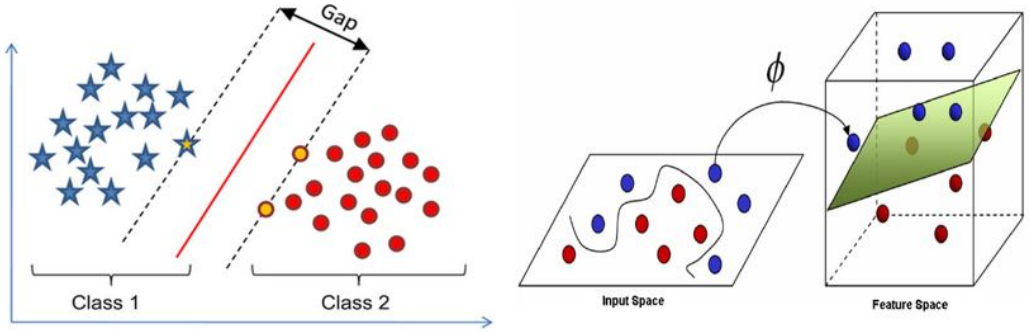


Fig. 2.8 Support vector machine. The hyperplane provides the maximum margin (left) and the application of a kernel function for mapping data into a higher-dimensional space (from [206]).

Linear discriminant analysis

The working principle of Linear discriminant analysis (LDA) is similar to that of PCA, aiming to find linear combinations of input variables that best explain the data. However, differently of PCA, LDA takes into account the class information, allowing to perform supervised classification. LDA aims to maximize the inter-class variance (i.e., the class separation) while minimizing the intra-class variance (i.e., the dispersion of data points belonging to the same class). Due to the intrinsic working principle of the algorithm, LDA provides a very good generalization capability. However, while SVM can project data into a higher-dimensional feature space, LDA projects the data into a lower-dimensional space. Thus, it is very important that the data set size (i.e., the number of elements) is larger than the number of variables/features.

2.3 Deep learning methods

DL algorithms provide the advantage of automatic feature extraction directly from raw input data, without requiring any feature engineering. This reduces the effort devoted to the definition of hand-crafted features and avoids user errors. Moreover, DL approaches are able to automatically identify hidden features, providing a more in-depth representation of the input data. On the other hand, a large amount of data is

required to work these methods properly, and their interpretability is still challenging [203].

2.3.1 Data preparation

The data preparation procedures depend on the specific DL model. The most basic ANN, named multi-layer perceptron (MLP) requires features to be extracted from the raw data and used as input to the model. In this case, some or all of the pre-processing procedures previously described are necessary for making the algorithm work properly. Specifically, at least segmentation and feature extraction should be performed before inputting the data into the model. Instead, some more complex DL architectures, such as convolutional neural networks (CNN) and recurrent neural networks (RNN) can be fed directly with raw data, as the feature extraction and selection steps are embedded into the DL model. This allows to save time and reduces the computational steps necessary for data preparation. However, some transformations (see Section 2.1.2) may be applied to the raw data before they are input to the DL model. Moreover, raw inertial data need to be segmented (see Section 2.1.4). Finally, centered data makes the algorithms speed up the computation and promote proper convergence of the solution.

2.3.2 Deep learning algorithms

Deep multi-layer perceptron

MLP represents the first ANN architecture, made of fully connected layers. This means that each neuron of a layer is connected to every neuron in the subsequent layer. MLP is composed of at least three layers, namely the input, hidden, and output layers (Figure 2.9). MLPs with a number of hidden layers equal to or larger than two are called deep neural networks (DNN).

Features are input to the MLP input layer (which has the same dimension as the feature set), processed through the hidden layer, and the final prediction finally occurs in the output layer (in which the dimension is equal to the number of different classes). The output y of each neuron is computed as in Equation 2.16, N is the number of neurons in the preceding layer, x_i the i^{th} input, w_i is the i^{th} weight, σ

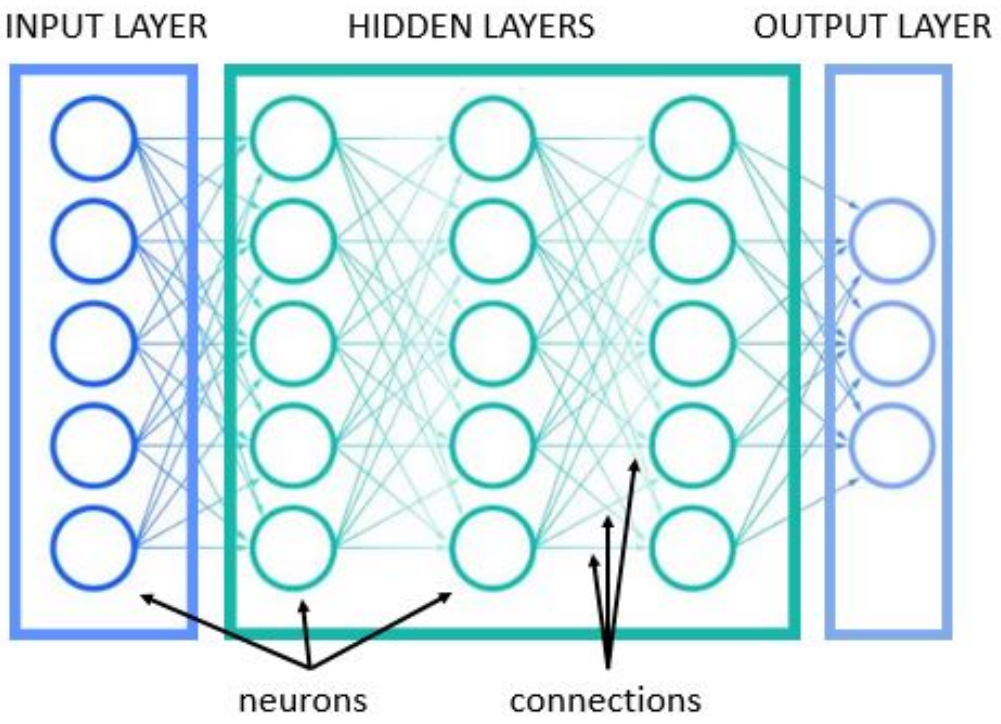


Fig. 2.9 Deep neural network.

the activation function, and b is the bias term. Different activation functions can be used. Hyperbolic tangent (Equation 2.17) and rectified linear unit (ReLU) are commonly used in the hidden layers, with the latter being much faster than the former. Sigmoid is used in the output layer in case of binary classification problems. On the other hand, the softmax activation function is used when dealing with multi-class problems.

$$y = \sum_{i=1}^{N} \sigma(\omega_i x_i) + b \tag{2.16}$$

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.17}$$

$$\sigma(x) = max(0, x) \tag{2.18}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.19}$$

$$\sigma(x)_i = \frac{e^{x_i}}{\sum_{c=1}^{C} e^{x_k}}, j = 1, 2, ..., k \tag{2.20}$$

The training process of DNNs is carried out using optimization algorithms aimed to find a local minimum of the loss function. The binary cross-entropy loss, commonly used for binary classification problems, is computed as in Equation 2.21, where $\hat{y}$ is the prediction, y the target class, and N is the total number of elements in the training set. While the number of neurons in the input and output layers are fixed, based on the number of input features and classes, the number of hidden layers and their size should be tuned to optimize the performance.

$$J(\hat{y},y) = -\frac{1}{N}\sum_{i=1}^{N}[y_i\dot{l}og(\hat{y}_i) + (1-y)_i\dot{(}log(1-\hat{y}_i))] \tag{2.21}$$

Convolutional neural network

CNN can learn a high level of abstraction and features from large data sets by applying convolution operations to the input data. In fact, CNN leverages three important ideas: sparse interactions, parameter sharing, and equivariant representations [207]. CNNs are competent to automatically extract features from images and signals and actually achieve state-of-the-art results in image classification, speech recognition, and text analysis. When applied to time series classification like HAR, CNNs have some advantages over other models, including local dependency and scale invariance [208]. Figure 2.10 schematically reports the principal layers of CNN architecture, listed and described in the following.

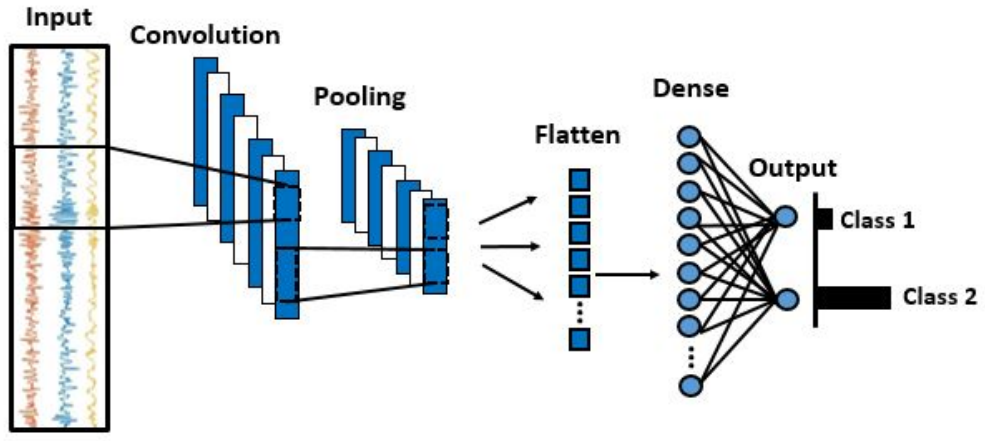


Fig. 2.10 Convolutional neural network layers and components.

Input layer. In the case of inertial signals, the input layer is usually represented by windows generated in the segmentation step. Input dimensions depend on the window size and number of channels, with the latter corresponding to the components of the recorded inertial data. When using all components from the 3-axis accelerometer, the number of channels is equal to three. This number increases up to six if using the additional components of the 3-axial gyroscope. The window size is obtained by multiplying the time duration (s) of the window by the sampling frequency. Given that the computational complexity and testing time of the algorithm are affected by the input dimension, it is advisable to resample the original signal using the lowest yet possible sampling frequency (see Section 2.1.1).

Convolutional layer (1D-CNN). Given a one-dimensional (1D) signal of length m, convolutional layers perform the convolution between the signal and a number n_f of filters of size f, sliding with a stride s. The generated output has dimensions $\left(\frac{m-f}{s}+1, n_f\right)$. Both the f weights and the bias term of each filter are learned during the training stage. The number of filters, size, and stride should be tuned and selected according to the optimization process.

ReLU activation function. ReLU represents the most common activation function in CNNs, increasing non-linearity and speeding up the computation. It is defined in the [0,+ ∞] interval, and computes the output as $a = max(0, z)$, where z is the input value.

Pooling layer. It provides a reduction in the size of the representation generated by the convolutional layer. The main advantages include speeding up the computation and summarizing the presence of features in patches of the feature map. The pooling layer applies a filter of size p with a stride s to the input data of dimensions (d, n_f), generating an output of dimensions ($\frac{d-p+1}{s}$, n_f). While max-pooling outputs the maximum value of the f values, average-pooling computes the mean of these values. The type, size, and stride of pooling should be tuned and selected according to the optimization process.

Flatten layer. It consists in unrolling the multidimensional matrix of dimensions (d, n_f) into a 1D vector of size (1, $d \cdot n_f$). It provides a mechanism to adapt the outputs of a CNN layer to dense layers. Each unit of the layer represents a neuron, which is then connected to each neuron of the subsequent dense layer.

Dense layer. It represents the fully connected layer of the network, in which each neuron is connected to each neuron of the preceding flattened layer. Given a

number of neurons d in the preceding layer and n_n in the dense layer, the number of parameters required for the computation is $d \cdot n_n + n_n$, where $d \cdot n_n$ accounts for the neuron weights and n_n for the neuron bias. Both the n_n weights and the bias term of each filter are learned during the learning stage of the algorithm. The number of dense layers and their size should be tuned and selected according to the optimization process.

Softmax layer. It represents the classification layer, in which the final continuous output of each neuron $\sigma(z_i)$ is computed from the input vector z, as shown in Equation 2.22. Given a number of classes k, the normalization term $\sum_{j=1}^{k} e^{z_k}$ ensures that all the output values of the function will sum up to 1, thus representing a valid probability distribution.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{2.22}$$

Moreover, the following regularization methods are used to avoid over-fitting and improve the generalization capability of the classification algorithm.

Dropout. Dropout regularization consists in randomly removing a given percentage of units (1 - dropout rate), thus training the CNN with a smaller number of neurons. It reduces over-fitting and increases the generalization capability of the classification model. The dropout rate should be tuned and selected according to the optimization process.

Regularization. Similarly to dropout, regularization aims to reduce the variance, hence the over-fitting. In this study, $L2$ regularization was used in the softmax layer, updating the general cost function by adding an additional term $\frac{\lambda}{2m} \cdot \sum \|\omega\|^2$, where λ is the regularization term, m the input dimension and ω represents the weights vector. The effect of $L2$ is the reduction of the connection weights. The regularization term should be tuned and selected according to the optimization process.

2.3.3 Training

In order to enable the learning process of the classification model, some preliminary settings should be adjusted, as discussed in the following.

Learning rate. The learning rate α is one of the most important parameters in the learning process, as it controls the learning step size at each iteration while moving toward the minima of the loss function. The larger is α, the faster the model weights are updated and the learning process turns out to be. However, an excessive α may lead to divergence of the solution. On the other hand, a very small α avoids divergence but slows down the computation and may lead to convergence to local minima. The learning rate should be selected by observing the training-validation loss learning curve to ensure proper model training and convergence of the solution.

Batch size. Training the model using mini-batches represents a compromise between the batch gradient descent (GD) algorithm and the stochastic GD. In the former approach, all the data are passed to the network at one time, while in the latter only one item at a time is passed to the network during the learning process. The former approach takes advantage of the vectorization of the input data, yet it is very slow, as all the input data are required to take a learning step. The latter approach speeds up the movement toward the minimum of the loss function but loses the speed-up provided by vectorization. The mini-batch GD consists in dividing the input data into $\frac{m}{b_s}$ batches of data, where m is the input data size and b_s is the mini-batch size. The training is then performed using a single batch at each learning step.

Number of epochs. The number of epochs defines the maximum number of iterations the model should undergo before the training process is stopped. A reduced number of iterations may lead to poor performance, while a too large number of iterations may result in model over-fitting.

Early stopping. In order to avoid over-fitting and reduce unnecessary computation during training, early stopping conditions may be defined in the learning process. In this case, the training is stopped when no reduction in the validation loss is observed for a certain number of iterations. The best parameters (i.e., those providing the minimum validation error) are saved and then used in the testing stage.

Optimizer. Adaptive moment estimation (Adam) [209] optimization was used in this study for weights and biases of the neural network. It combines both the Momentum and the RMSprop GD algorithms, thus it is very effective and commonly used in deep neural network architectures. First, Momentum GD computes gradients $d\omega$ at each iteration for the t_{th} mini-batch. Then, it computes $V_{dw} = \beta_1 V_{d\omega} + (1 - \beta_1) d\omega$, with $\beta_1 = 0.9$. Finally, weights are updated with $\omega = \omega - \alpha V_{d\omega}$. RMSprop

works in a similar way, computing $S_{dw} = \beta_2 S_{d\omega} + (1-\beta) d\omega^2$, and finally updating weights as in Equation 2.23, where ε is 10^{-8} and β_2 is set to 0.999.

$$\omega = \omega - \alpha \frac{d\omega}{\sqrt{S_{d\omega} + \varepsilon}} \tag{2.23}$$

Adam first computes $V'_{d\omega} = \frac{V_{d\omega}}{1-\beta_1^t}$ and $S'_{d\omega} = \frac{S_{d\omega}}{1-\beta_2^t}$, and finally updates weights ω as in Equation 2.24, where α is the learning rate.

$$\omega = \omega - \alpha \frac{V_{d\omega'}}{\sqrt{S'_{d\omega} + \varepsilon}} \tag{2.24}$$

Loss function. For the classification task, the categorical cross-entropy loss function is commonly used. It is defined as in Equation 2.25, where N is the total number of samples, y_i is the i^{th} class label, and $\bar{y}_i$ is the i^{th} prediction.

$$E = -\frac{1}{N} \sum_{i=1}^{N} y_i \cdot log(\bar{y}_i) + (1-y_i) \cdot log(1-\bar{y}_i) \tag{2.25}$$

2.4 Performance evaluation

The evaluation of the prediction performance represents one of the most important stages of classification and/or regression tasks. A comprehensive performance evaluation allows to fully describe the strengths and weaknesses of the developed algorithm and to compare the results with past or future literature studies. The performance metrics are described in Section 2.4.2 and Section 2.4.3 for classification and regression, respectively. Section 2.4.4 discusses the effect of class imbalance on classification performance, together with an overview of possible solutions.

2.4.1 Validation methods

In order to prepare the data set generated in the previous steps for the subsequent classification process, a splitting procedure should be applied. Specifically, the data set is divided into training, validation, and test set. The former is used to train the classification model, while the latter is used for performance evaluation. All classification algorithms involve a set of parameters, intrinsic to the model itself,

that should be correctly tuned for maximizing the performance. This optimization procedure is usually carried out on the training set. Specifically, the training set is further divided into training and validation sets; for each combination of parameters' values, the model is trained on the training set, and its performance is evaluated on the validation set; then, the combination of parameters providing the best results is selected. Finally, the optimized model is tested on the independent test set, which represents completely new unknown data.

There are different types of validation and test methods. Hold-out validation consists in randomly dividing the data set into training and test sets, according to selected proportions (e.g., 0.8 for training, 0.2 for testing). K-fold cross-validation (CV) divides the data set into k folds. Then, the model is iteratively trained using data from k-1 folds and tested on data from the remaining fold. Common values for k are 5 and 10, where the procedure is repeated 5 and 10 times, respectively. Leave-one-subject-out (LOSO) validation is a particular case of k-fold CV, in which k is equal to the number of subjects. This method ensures subject independence, as data from the same patient belongs either to the training or to the test set. Finally, leave-one-task-out is used when subjects perform more than one experiment. In this case, the training phase involves all tasks except for one, which is used as a test. This validation procedure is commonly performed when the objective is to develop subject-specific models, as in this case both training and testing involve data from the same subject.

2.4.2 Classification metrics

In the following, classification metrics are listed and described for a classic binary classification problem. In this context, the ground truth (e.g., the clinical score, the presence or absence of symptoms) corresponds to a binary variable, named in the following with the term "label". Similarly, the prediction is a binary value, obtained as the output of the classification model (in case of discrete output) or computed as the binary approximation of the continuous probability (in case of continuous output). Positive and negative samples indicate the presence or absence of a particular symptom (e.g., FOG).

True positives (TP) are defined as the true samples correctly identified by the model. False positives (FP) represent negative samples wrongly predicted as positive.

False negatives (FN) correspond to positive samples not detected by the model. Finally, true negatives (TN) represent correctly classified negative instances. Figure 2.11 schematically describes these metrics, where both label and prediction can assume only binary integer or categorical values (e.g., 0 or 1, yes or no).

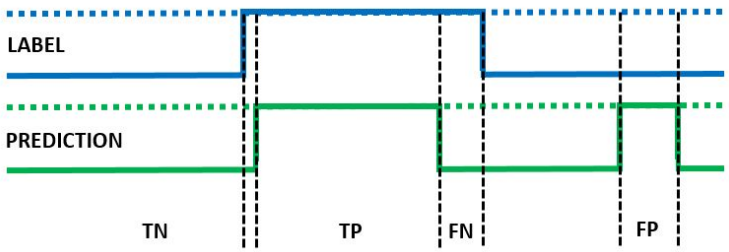


Fig. 2.11 Definition of true (TP) and false positives (FP) and true (TN) and false negatives (FN).

Sensitivity/recall (Equation 2.26) evaluates how many true samples are recognized by the model. Specificity (Equation 2.27) measures how efficiently negative samples are discarded. Accuracy (Equation 2.28) is an overall performance evaluation metric providing the percentage of correct classification. The geometric mean of sensitivity and specificity (Equation 3.1) is useful to appreciate the situations in which one of the two measures is far smaller than the other. F-score (Equation 2.30) is computed as the harmonic mean between sensitivity and precision/positive predictive value, with positive predictive value (PPV) computed as in Equation 2.31. Finally, negative predictive value (NPV) is computed as in Equation 2.32.

In the case of unbalanced data sets (see Section 2.4.4), the F-score is preferable to accuracy as a global correct classification metric.

$$sensitivity = \frac{TP}{TP+FN} \tag{2.26}$$

$$specificity = \frac{TN}{TN+FP} \tag{2.27}$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.28}$$

$$geometric-mean = \sqrt{sensitivity \cdot specificity} \tag{2.29}$$

$$F-score = \frac{2 \cdot sensitivity \cdot precision}{sensitivity + precision} \tag{2.30}$$

$$PPV = \frac{TP}{TP+FP} \tag{2.31}$$

$$NPV = \frac{TN}{TN+FN} \tag{2.32}$$

Moreover, the receiver operating characteristic (ROC) curve is a graphical plot describing the diagnostic capability of a binary classifier as its discrimination threshold varies. Specifically, the ROC curve is created by plotting the true positive rate (i.e., sensitivity) against the false positive rate (i.e., 1 - specificity) at various threshold settings. Finally, the area under the curve (AUC) measures the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve, while the equal error rate (EER) corresponds to the error observed at the point in the ROC curve where sensitivity equals specificity.

2.4.3 Regression metrics

Differently of classification, where the objective is to predict specific classes, the objective of regression algorithms is to predict a continuous score. In this case, proper performance metrics are defined, different from that used in classification tasks. The correlation coefficient (Equation 2.33) measures how well the model fits the dependent variable, i.e., how much variability in the dependent variable can be explained by the model; it ranges between 0 and 1, with larger values indicating better performance. Root mean square error (RMSE, Equation 2.34) and mean absolute error (MAE, Equation 2.35) are absolute measures of the goodness of fit, providing the entity of deviation from the target values. While MAE treats all errors the same, RMSE gives larger penalization to big prediction errors.

$$r = \sqrt{1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})}} \tag{2.33}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2.34}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \bar{y}| \tag{2.35}$$

2.4.4 The problem of class unbalance

In most real-world applications, classes are not balanced, which means that different classes can encompass a different number of elements. ML algorithms aim to minimize the misclassification error, regardless of the distribution of the training items in the different classes. In the case of very unbalanced classes, this will result in a high accuracy value and poor sensitivity or specificity in the smallest represented class. Figure 4.9 reports an example of a confusion matrix, where the positive class includes 10 samples, and the negative class contains 100 elements. The computation of performance metrics results in an accuracy of 0.87 and a specificity of 0.95, which are good results. However, sensitivity is equal to 0.1 and precision is less than 0.2, leading to poor usability of the classification model. Different methods can be used to solve this problem, balancing the original data set and providing balanced results.

	Predicted positives	Predicted negatives
True positives	1	9
True negatives	5	95

Fig. 2.12 Example of the confusion matrix.

Under-sampling consists in randomly selecting only a certain percentage of the most represented class. In the case reported in Figure 4.9, the negative class would be sampled with a ratio of 0.1, thus discarding 90% of the negative elements. This procedure makes the data set balanced, however, a huge loss of information occurs. Over-sampling works in the opposite direction. In this case, the elements of the smallest class are repeated a certain number of times, depending on the balance ratio. However, this would produce over-fitting (i.e., the model is fitting on the provided samples and it is unlike to have good generalization capabilities).

This issue can be partially solved by generating artificial data, similar but not identical to that of the smallest class but not identical. New data points are generated using linear combinations of the training samples. Synthetic minority oversampling technique (SMOTE) [210] is an example of an over-sampling algorithm.

Data augmentation techniques include also arbitrary rotations of the inertial signals. Starting from each single data window (see Section 2.1.4), additional

windows can be obtained by inverting signals along the time dimension, changing the sign of positive and negative recorded samples, or changing the order of axes (e.g., invert x and y axis). The additional windows are then merged with the original data set, resulting in a larger and more heterogeneous sample.

The cost function of the ML classification model may be modified to assign different weights to errors in different classes. The cost matrix is a 2D matrix reporting the cost of errors occurring in each field of the confusion matrix (Figure 4.9). The error weight of TP and TN is set to zero, and that of FP and FN is equal by default (e.g., equal to 1). Changing one of the two weights (e.g. the FN) would produce different results, affecting the performance metrics (e.g., sensitivity). However, despite some metrics can be maximized using this approach, it is unlikely that the overall accuracy would improve.

Finally, sequential data, such as inertial signals, commonly undergo a segmentation procedure (see Section 2.1.4), consisting in dividing the original signal into different time frames. An alternative approach for class balancing could be to use a different overlap for signals belonging to different classes. A larger overlap can be used for the less-represented class, resulting in balanced data. This would not produce any loss of information, as the most represented class is not affected by the balancing procedure. Moreover, any data are synthetically generated or repeated identically multiple times.

2.5 Related work

The combination of WS and ML techniques has been successfully employed for the monitoring of several motor aspects of PD. It has been employed for assessing bradykinesia [151, 211], dyskinesia [212, 213], postural stability [140, 176], tremor [127], and FOG [157, 158, 135, 214, 162]. Moreover, some studies implemented ML methods for predicting single MDS-UPDRS items related to motor signs [98, 215] or a set of MDS-UPDRS motor scores [216]. A detailed overview of the literature works focusing on different aspects of PD is provided in the following.

FOG

Automated FOG detection methods based on WMS and ML have received increasing attention over the last two decades. Different sensors, sensor configurations, experimental procedures, and classification algorithms have been proposed, providing incremental improvement in detection performance. The number of sensors ranges from 1 [157, 158, 217] to 6 [218], and several sensor locations have been proposed [156–158, 160–162]. Experimental procedures include a large variety of walking tasks. Some studies addressed the timed-up-and-go test (TUG) [219, 220], as it includes gait initiation and turns. In other studies, self-designed protocols included walking [161], together with turns and stops for eliciting FOG [160, 221, 214]. Finally, few studies included activities similar to ADLs in addition to free-like walking tasks [135, 217, 159]. As far as concerns the sample size, most published works enrolled a population ranging from 7 to 12 PwPD [219, 160, 161, 220, 221, 162, 214], with few studies addressing larger samples of 21 [135, 159, 217] and 38 [157] patients. A large variety of FOG detection algorithms has been proposed. The first approaches were based on the computation of a few indices characterizing FOG and the use of a simple threshold to distinguish FOG from other activities [222, 223]. The freeze index [222] represents the first FOG index described in the literature and is evaluated as the ratio of power in the freezing band (i.e., 3-8 Hz) and that in the locomotion band (i.e., 0-3 Hz). Subsequent works aimed to improve the classification performance by extracting a larger number of features and using ML algorithms for the classification task, including SVM [135, 157], kNN [162, 214], and random forest [224]. Finally, DL was proposed to detect FOG, outperforming classical ML models [158, 217, 156]. Recent studies have taken advantage of the high potential of DL algorithms to perform FOG detection and prediction. More specifically, CNN [158, 156, 161], LSTM [217, 221], and deep autoencoders [160] achieved good to excellent performance in FOG detection, with sensitivity ranging from 0.63 [161] to 0.92 [217] and specificity ranging from 0.75 [135] to 0.98 [161]. The major limitations of the mentioned studies include the small PD patients cohort, the lack of external validation data sets, the use of laboratory settings and supervised experiments, the high computational complexity of the designed classification algorithms, and/or the non-suitability for real-time implementations.

WMS also enables monitoring of spatial-temporal gait degradation in patients with FOG, possibly useful for the recognition of pre-FOG periods (i.e., prediction of

FOG before its actual occurrence). Indeed, their usage, in combination with ML, has recently paved the way for FOG prediction [162, 225–229]. By examining several time- and frequency-domain gait features, these studies have achieved the real-time detection of pre-FOG periods [225–227, 161]. However, the reported performance of ML analysis in predicting FOG is suboptimal in terms of accuracy, probably reflecting the clinical heterogeneity of the cohorts under study. For example, accuracy in predicting FOG in PD would benefit from assessing the effect of L-Dopa, which is known to improve spatio-temporal gait parameters (e.g., stride length and speed). Accordingly, the patient's condition with respect to dopaminergic therapy would influence the effectiveness of the algorithm and thus the accuracy of FOG prediction. None of the previous studies using wearable inertial sensors and ML analysis to predict FOG evaluated and compared PwPD with and without dopaminergic therapy.

Postural stability and gait impairment

Over the past two decades, wearable sensing systems based on accelerometers and gyroscopes have been increasingly used for objective monitoring of gait and balance in PwPD [230, 231, 145]. These technologies have provided highly accurate data through analysis of inertial data recorded during various activities, including stance [140, 129], postural transitions [232, 233], gait [234, 179, 123], and turning [235, 176]. In addition, several ML algorithms have been used previously to objectively assess the severity of PD [236, 237]. However, so far only a few studies have used sensor-based recordings in PD to predict specific MDS-UPDRS items regarding balance and gait or even the postural stability and gait difficulty (PIGD) score, achieving suboptimal performance. Currently, clinical assessment of axial impairment in PD involves measurement of the PIGD score, which is an accurate indicator of disease severity and prognosis [238]. To obtain quantitative markers of disease progression, it would be important to measure the PIGD score objectively in PwPD. In [215], gait data of 75 PwPD were recorded using a single IMU on the waist. An SVM classifier was used for gait detection, and power spectra in the range of 0–10 Hz were calculated and used as output. The results showed that the output of the algorithm was correlated with MDS-UPDRS gait scores ($r = -0.73$, $p < 0.001$) and PS scores ($r = -0.42$, $p < 0.001$). In [239], 31 participants were equipped with 3 WMS on the lower back and on each foot. Inertial data were recorded in the home environment and gait, turning, and stance activities were registered and analyzed.

Several measures related to the quantity and quality of movements were extracted, selected, and input to a multivariable LR model. The outcome of the algorithm was found to be correlated with UPDRS-part III total score ($r = 0.48$, $p = 0.007$) and PIGD score ($r = 0.61$, $p < 0.001$). To overcome the limitations of the aforementioned studies, a DL model fed with raw inertial data was validated and tested on a large number of PwPD, obtaining good performance [240]. Specifically, 119 PwPD were fitted with a single triaxial accelerometer on the lower back and asked to walk back and forth for 2 minutes. A CNN model was used to predict the MDS-UPDRS part III total score, obtaining an $r = 0.82$ and a $RMSE = 6.3$. These studies did not consider several clinical biases that have a major impact on sensor-based measures, such as the effect of L-Dopa and the presence of FOG. In addition, the use of multiple WMS reduced the noninvasiveness and comfort of the adopted sensing systems [239].

Among the various walking tasks performed during daily activities, many studies in the literature, as well as clinical experience, recognize turning as a privileged activity to detect motor impairment in PwPD. Despite numerous studies on the gait characteristics of PwPD using WS, only a few have focused on turn detection and analysis. Among them, the authors of [184] used a single sensor on the lower back and recorded data from a cohort of 20 PwPD and 13 elderly people with the goal of detecting turns in both ON and OFF states. The experimental protocol included simulated ADLs including free walking, sit-to-stand, brushing teeth, ironing, and turning. The obtained detection performance, measured in terms of sensitivity, specificity, and accuracy, was 92%, 89%, and 92% in the ON state and 92%, 78%, and 83% in the OFF state. In [180], 45 elderly and 10 young adults were equipped with four WMS, placed on the sternum, lower back, and thighs. Participants were asked to perform multiple 360^{circ} turns in both directions; the aim was to identify people characterized by a high anamnestic risk of falling. Dynamic time warping was performed on the angular velocity signals of elderly participants and controls, and significant differences emerged between the two classes. However, the method was not applied to patients with neurodegenerative diseases. All of the cited studies have some limitations, including a small cohort of PD patients (maximum 30 PD patients), supervised experimental protocols, the use of dedicated hardware, and, in general, the use of multiple IMUs. In addition, no postural control index or objective measure with assessed correlation with clinical parameters has been proposed.

Bradykinesia

The combination of WMS and ML techniques was also exploited to recognize the severity of bradykinesia. In [185, 241], 34 and 24 subjects were enrolled, respectively. Three IMUs were mounted on the patient's chest and each thigh to assess the leg agility (LA) task in the clinical setting. Time and frequency domain features were extracted and selected to feed the classification algorithms, namely SVM and kNN. The accuracy was found to be 43% in both studies. A correlation coefficient $r = 0.49$ between the automatic scoring system and the MDS-UPDRS clinical assessment was reported in [241]. Bradykinesia was assessed in [242] through a specific smartphone application tested on 14 PwPD. Several items of the MDS-UPDRS Part III were analyzed, including finger touch, hand pronation-supination movements, and LA. Focusing on the latter, the best correlation with the clinical score was found with leg movement power ($r = -0.5,\ p = 0.015$). However, no classification tasks designed to assess the severity of bradykinesia were performed. In [188], a six-month clinical study of 44 PwPD was conducted to assess many MDS-UPDRS motor tasks using a smartphone in the home environment. Although LA was not included in the assessment, the walking task can be partially considered an indicator of bradykinesia; the authors found an intra-class correlation coefficient (*ICC*) of 0.88 with the MDS-UPDRS walking score. In addition, in [186] 19 PwPD were monitored with ankle-mounted WMS for assessment of LA and response to treatment. Time and frequency domain characteristics were calculated to feed different classifiers, i.e., SVM, DT, and LR. Performance was expressed in terms of *ICC*, Pearson correlation coefficient, and RMSE with respect to MDS-UPDRS bradykinesia score (item 3.14); $ICC = 0.89$, $r = 0.83$ and $RMSE = 0.53$ were reported. Finally, in [187] 50 PwPDs were tested by wearing IMU on each ankle for LA quantification. A fuzzy logic inference model was constructed by exploiting both the most significant features and rules based on the recommendations of MDS-UPDRS item 3.8. Unfortunately, classification results in terms of accuracy, r, *ICC*, and *RMSE* were not reported. It is worth noting that all of the cited studies have some limitations, including the small sample size (maximum 44 PwPD), the use of dedicated hardware (only one study employs smartphones), and the number of sensors (most studies use three sensors). In addition, some relevant details about the dataset (e.g., the cardinality of MDS-UPDRS classes) were not reported.

Chapter 3

Materials and methods

This chapter describes the data and the processing methods used in this study. Different sections refer to any specific objective. FOG detection and prediction are addressed in Section 3.1. The automatic assessment of postural stability while performing different tasks is described in Section 3.2. The evaluation of gait impairment is described in Section 3.3, while Section 3.4 reports the automatic estimation of bradykinesia. Each of the following sections is structured as follows. First, an overview is provided regarding the objective of the study, along with a brief summary of the employed materials and methods. Then, the patients' cohort enrolled in the study, data acquisition procedures (i.e., sensors, experimental protocol), preprocessing steps (i.e., denoising, feature extraction, feature selection), and implementation details of classification and regression models are described.

3.1 Freezing of gait

The following two studies focus on the detection and prediction of FOG. In the first work (Section 3.1.1), data were recorded using sensors on the shanks, and the analyses were performed using an optimized feature set and ML model. Performance was evaluated in both ON and OFF conditions, and the effect of therapy was evaluated. In the second work (Section 3.1.2), data were recorded using a single sensor on the waist, and DL was used to develop a real-time FOG detection algorithm.

3.1.1 Prediction of freezing of gait using lower limbs inertial sensors

Despite a large number of studies focused on the detection of FOG, few of them aimed to recognize the typical degradation of gait patterns preceding FOG. Moreover, none of them considered the effect of therapy on the detection and prediction performance. In this work, the combination of ML and WMS data was exploited for the detection and prediction of FOG [219]. Participants were equipped with two IMUs on their shins and asked to perform a TUG test. Data from a single axis of the gyroscope were processed and fed to different ML models. Classification models were optimized and combined for improving the algorithm performance. Finally, the results of both FOG detection and prediction algorithms were assessed, further considering the effect of therapy.

Subjects

Eleven PwPD affected by FOG were enrolled in the Movement Disorder outpatient clinic of the Department of Human Neurosciences, Sapienza University of Rome, Italy. The inclusion criteria were as follows. A clinical diagnosis of idiopathic PD based on current consensus criteria [243], lack of dementia (mini-mental state examination - MMSE > 24), presence of FOG directly verified by physical examination of two neurologists, ability to walk independently, and lack of comorbidities possibly affecting gait (e.g. neuropathies, rheumatic and orthopedic disorders). To assess patients' motor, cognitive, and emotional functions, the clinical examination included the following standardized scales and scores: H&Y, MDS-UPDRS part III, FOG questionnaire (FOG-Q), MMSE, frontal assessment battery (FAB), Hamilton depression rating scale (HAM-D) and Beck anxiety inventory (BAI). During the experimental sessions, patients were studied both under (1 hour after L-Dopa intake) and not under (after L-Dopa withdrawal for at least 12 hours) dopaminergic therapy (i.e., ON and OFF state of therapy, respectively). Finally, the L-Dopa equivalent daily doses (LEDDs) were calculated for each patient [244]. Table 3.1 summarizes the demographic and clinical features of the population enrolled in this study. In agreement with the Declaration of Helsinki, the experimental procedures were approved by the institutional review board of Sapienza University of Rome, Italy. Also, all the patients gave written informed consent to experimental procedures.

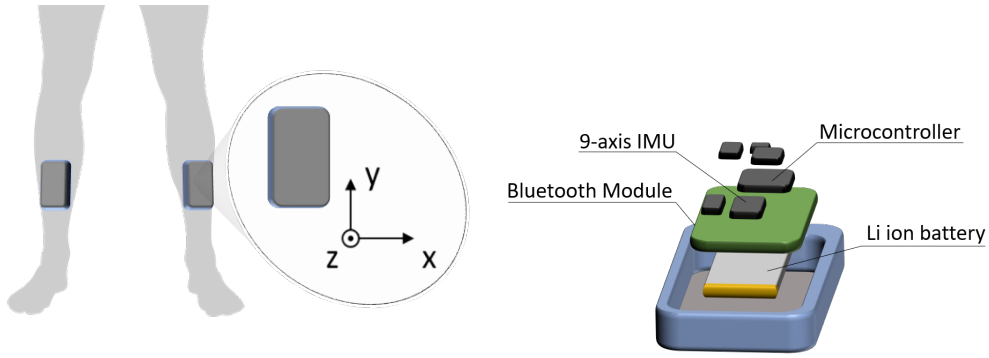
Table 3.1 Demographic and clinical features of patients enrolled in this study (mean ± standard deviation). MDS : movement disorder society; UPDRS : unified Parkinson's disease rating scale; ON : under dopaminergic therapy; OFF : not under dopaminergic therapy.

Number of subjects (male)	11 (7)
Age (years)	73 ± 7
Disease duration (years)	10.5 ± 7
FOG duration (years)	6.7 ± 1.6
Hoehn & Yahr score	2.7 ± 1
MDS-UPDRS - part III ON	37.9 ± 15.1
MDS-UPDRS - part III OFF	44.5 ± 16.9
FOG questionnaire	18.6 ± 2.9
Mini-mental state examination	28.3 ± 2.1
Frontal assessment battery	14.4 ± 2.8
Hamilton depression rating scale	17 ± 7.8
Beck anxiety inventory	16.5 ± 13
Levodopa equivalent daily dose (mg)	741 ± 272

Data acquisition

The motor task consisted of 7-meter TUG test requiring patients to get up from a chair, walk in a straight line for 7 m, turn, walk back, and sit down. To increase the probability of FOG occurrence, the TUG test was performed in a free-living-like environment equipped with factors that commonly elicit FOG in a domestic setting. More in detail, the TUG test implied the passage from a spacious room to a narrow and furnished corridor (about 1.5 meters wide) with the interposition of an open door [145]. During TUG tests, PwPD were video-recorded through a camera and monitored by two IMUs placed and fixed on the shins (Figure 3.1a) through elastic bands, which allowed a good and permanent adhesion during the tests. Video recordings were used for the offline clinical assessment by two independent neurologists, experts in movement disorders, serving as gold standard evaluation for FOG detection. More in detail, two independent neurologists separately identified the start and end of FOG episodes and, in case of discrepancy, performed a common assessment to resolve the ambiguity. The IMUs positioning on the patient was implemented so that when the patient was standing the y-axis represented the inverse gravity vector and the x-axis lay in the frontal plane. Hence, the angular velocity around the x-axis enables a good representation of human motion during linear gait. The STMicroelectronics system-on-board prototypes neMEMSi [245] were equipped

with a 9-axis IMU (LSM9DS0), integrating a ± 16 g 3D accelerometer, a ± 12 Gauss 3D magnetometer, and a ± 2000 dps 3D gyroscope; a Bluetooth V3.0 module (BT33); a lithium-ion battery; an ultralow-power 32-bit microcontroller (STM32L1) (Figure 3.1b). Also, the neMEMSi device included temperature, hygrometer, and pressure sensors that were not used for this study. A preliminary conventional calibration process of the inertial sensors was performed. It consisted of a software correction of the displacement of the IMUs framework with respect to the earth framework, before their positioning on the patient. Real-time IMU data were acquired with a sampling frequency of 60 Hz, acceleration full scale of $\pm$ 2 g, and angular velocity full scale of $\pm$ 245 dps. No additional analog/digital filter was added besides the ones specified in the datasheets. The resulting data were sent in real-time to a personal computer through the neMEMSi Bluetooth module and progressively saved in CSV format.



(a) Sketch of wearable device positioning and inertial reference frame

(b) Exploded view of the neMEMSi device.

Fig. 3.1 Sensor positioning and composition

Each CSV file was related to a single test. Data in CSV files were processed offline as described in the next Section. For the synchronization of the two devices, data collection started when the patient was sitting down. When the patient stood up, an evident peak in the collected data from the 3-axis gyroscopes took place, as can be observed in Figure 3.2 for the x-axis of the gyroscope. In that plot, the normalized angular velocity around the x-axis is drawn versus time. At time t=2.6 s the patient stood up and a peak from each device can be detected. In the following few seconds (till t=6 s) data are not meaningful because the patient was arranging their position. After t=6.5 s the patient started walking. By superimposing the standing-up peaks related to the two legs, the relative delay from each other can be calculated. The

mentioned method allows a perfect synchronization between signals from the two shins, which conserves their phase shift along the whole test duration.

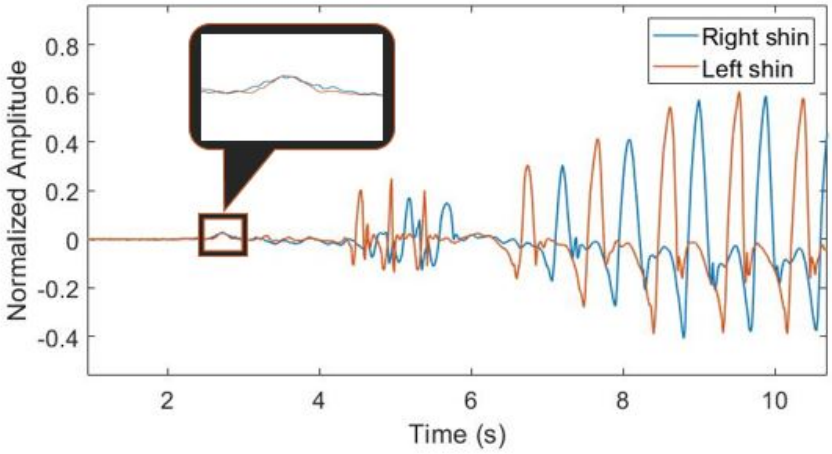


Fig. 3.2 Normalized angular velocity around the x axis versus time for the two shins during the transition from sitting to standing up.

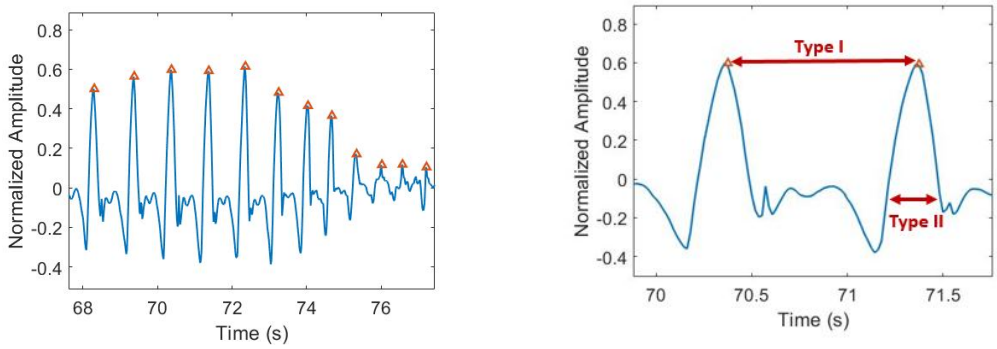
Preprocessing

In this study, a single component of the angular velocity signal was employed, which describes the principal angular movement of the leg. Specifically, the angular velocity signal around the x-axis was addressed, in the following reported as ω_x.

Normalization. The raw input data ω_x were normalized using the mean-normalization formula (see Section 2.1.8). Signal normalization allows to process a homogeneous range of motion data for the entire population, and to perform data segmentation in a subject-independent way.

Segmentation. In order to get information about the traits and qualities of each step accomplished by the patient, a step-to-step data segmentation process was performed. To this end, signal peaks were identified as anchor points for segmentation; indeed, points in which the angular velocity reaches the maximum are known to represent the mid-swing phase in gait analysis [246, 247]. This procedure was recognized to ease step detection and limit detection errors. In more detail, signals from the right and left leg were kept separated, and only signal peaks with an amplitude $\geq$ 20% of the maximum value and at least 350 ms apart were considered. The amplitude threshold was heuristically selected in order to catch both normal and anomalous steps such as those preceding FOG. The temporal threshold was set to

avoid duplicated peak detection during normal gait. In Figure 3.3a, an example of the outcome of our peak detection algorithm is reported. Once having identified signal peaks, two data segmentation tasks were performed in order to arrange data frames for subsequent feature extraction. Type I segmentation catches data between two subsequent peaks (i.e., the current and the previous one), while Type II segmentation measures the positive portion of data encompassed by the current peak (Figure 3.3b). Type I segments yield frequency information, while range of motion and movement intensity can be computed using the Type II segmentation.



(a) Signal peaks identified by the peak detection algorithm (red triangles)

(b) Type I and Type II segments.

Fig. 3.3 Peak detection and signal segmentation.

As the objective is to catch walking pattern degradation preceding FOG events, the features to be extracted from inertial data should be able to represent subtle details of each step. The selected features in both time and frequency domains are reported in Table 3.2. Some of them, e.g., standard deviation, range, and root mean square, are self-explaining. In the following, a description of those features requiring some comments is provided.

Angular Jerk. It represents the rate of variation of the angular acceleration, defined as $\frac{1}{2} \cdot \int \ddot{\omega}_x^2 \cdot dt$, where $\ddot{\omega}_x$ is the second derivative of the angular velocity around the x-axis.

Normalized Jerk. It represents the Angular Jerk normalized by the time in which it is computed.

Stride similarity. It is computed using the Dynamic Time Warping (DTW) algorithm. It provides a scalar output that is inversely proportional to the similarity between the two input signals. Thus the output represents the similarity between the actual stride

Table 3.2 List of extracted features, along with the employed segmentation type.

Domain	Feature	Segmentation
Temporal	Standard Deviation	Type I
	Range	Type I
	Root Mean Square	Type I
	Angular Jerk	Type II
	Normalized Jerk	Type II
	Stride Similarity	Type I
	Step Time	Type II
	Stride Time	Type I
	Peak height	Type II
	Peak width	Type II
Spectral	Power Spectral Entropy	Type I
	Principal Harmonic Frequency	Type I
	Principal Harmonic Amplitude	Type I
	Principal Harmonic Width	Type I
	Weighted Power Spectral Frequency	Type I
	Low Power Frequency	Type I

and the previous one.

Step Time. It is computed as the temporal distance between each peak and the previous contralateral peak.

Stride Time. It is computed as the temporal distance between two subsequent peaks in the signal measured on either the right or the left leg.

Peak height and *width*. They are computed respectively as the height (with respect to zero) and half-power width of the positive portion of the signal peak. The former represents the maximum angular velocity reached in each step while the latter is proportional to the swing time.

Power Spectral Entropy. It is the spectral Shannon Entropy, computed as $-P \cdot \log(P+\varepsilon)$, where P is the normalized squared amplitude of the signal Fast Fourier Transform (FFT) and ε an arbitrarily small value (0.001) ensuring real output values. It represents a measure of the quantity of information carried by the signal spectrum.

Principal Harmonic Amplitude and *Frequency.* They are computed from the signal FFT, as the peak value and its corresponding abscissa (frequency).

Principal Harmonic Width. It is obtained as the half-power width of the principal harmonic component.

Weighted Power Spectral Peak. It is the product of amplitude and frequency of the principal harmonic.
Low Power Frequency. It represents the ratio between the power in the bandwidth 0-2 Hz and the total signal power.

Classification

A binary supervised classification problem was set up for FOG detection, namely gait vs FOG. This allowed us to get an insight into the capability of the extracted feature set to discriminate between normal and abnormal gait patterns. In view of the subsequent pre-FOG detection task, the implemented algorithm should be robust and easily interpretable. In this context, DT represents a simple and fast algorithm, providing straightforward interpretability of its outcome. Nevertheless, DT is known to implement a very sharp margin separating the two classes, thus increasing the risk of over-fitting. On the other hand, SVM seeks the hyperplane providing the largest margin for separating the two classes and it has been widely employed in similar problems [248] [249]. In this work, both models were combined by exploiting DT for feature selection and SVM for classification. As for DT, features close to the tree root achieve the best classification of the training set. Hence, this algorithm can be used to rank features in decreasing order of relevance. The implemented DT has the following parameters: split criterion based on Gini-Simpson diversity index [204], minimum leaf size equal to 1, and a maximum number of splits of 15. A tuning procedure was performed to identify the best feature subset and model parameters, i.e., those minimizing the misclassification error in a 10-fold CV. The number of selected features varied from 1 to N (i.e., the entire feature set), while the SVM regularization term ranged from 1 to 20.

The generated feature set was split into 70% training and 30% test sets to avoid over-fitting. Then, the model optimization process was performed using the training set and the resulting optimized model was tested on the test set. Moreover, in order to ensure subject independence and to achieve results representative of more realistic working conditions, LOSO validation was performed. The classification performance was evaluated for each performed validation/test, in terms of sensitivity, specificity, accuracy, PPV and NPV, F-score, and AUC. Training, validation, and test were carried out on data related to patients ON and OFF therapy separately. Then, the results obtained in the two conditions were compared. Finally, the final model

configuration was trained with data related to patients ON (OFF) therapy and tested on OFF (ON) data. Thus, the performance obtained in the two testing conditions was compared.

As for pre-FOG detection, i.e., capturing typical degradation of gait pattern preceding FOG episodes, a binary classification problem was set up to differentiate between gait and pre-FOG. The steps employed to select the final model configuration, which is used for pre-FOG identification, are described in Figure 3.4.

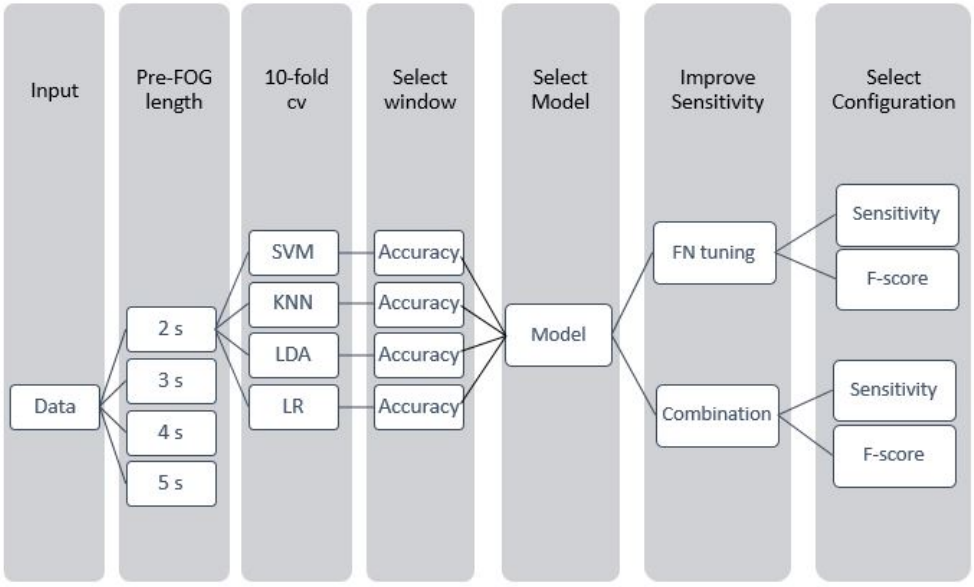


Fig. 3.4 Workflow of the procedure used for the identification of the final model configuration. SVM : support vector machine; kNN : k-nearest neighbor; LDA : linear discriminant analysis; LogR : logistic regression; FN : false negative.

First of all, as the pre-FOG window length cannot be determined *a-priori*, different window lengths in the range of 2-5 s were considered. For each value, the corresponding gait data were labeled as belonging to the pre-FOG class. As the identification of the most suitable classification algorithm for this task is not straightforward, several tests were performed using the model implemented for FOG detection as well as different models, namely kNN, LDA, and LogR. In order to jointly identify the most suitable window length and classification model, a 10-fold CV was performed for each model-window length pair and computed accuracy. For each classification algorithm, the corresponding hyperparameters were optimized employing a Bayesian optimization algorithm. The hyperparameters and the corre-

sponding range used for model optimization in the 10-fold CV procedure are reported in Table 3.3. Once having identified the model providing the best performance, two approaches were exploited for improving the algorithm sensitivity. The first one consists in tuning the FN cost of the algorithm in the range of 0-10, in order to reduce the number of pre-FOG samples that are not recognized by the algorithm. The second approach employs a combination of detection models; data samples are classified as pre-FOG if at least one model yields a pre-FOG decision.

Table 3.3 Parameters employed for model optimization, along with the corresponding range.

Model	Parameter	Range
SVM	kernel function	linear, quadratic, gaussian
	kernel scale	0.001 - 100
	box-constraint	0.01 - 100
kNN	number of neighbors	1 - 50
	distance metric	euclidean, manhattan
	distance weight	equal, inverse, squared inverse
LDA	Gamma	0.01 - 1
	Delta	0.01 -100
LR	Lambda	0.01 - 100

In both cases, sensitivity, accuracy, and F-score were computed and the configuration providing the best performance was selected. Then, LOSO validation was performed, and several classification evaluation metrics were computed. Finally, in order to provide interpretability of the developed classification model, the relevance of features in discriminating normal from abnormal (i.e., pre-FOG) gait steps was assessed. Specifically, the most frequently selected features were addressed (i.e., in at least 80% of cases in the LOSO validation) and the Spearman correlation coefficient and the associated p-value were computed between those features and the binary class label (i.e., 0: gait, 1: pre-FOG). This method allows quantifying the statistical dependence between features and target, thus providing a measure of increase or reduction in variables' value while approaching a FOG event. Specifically, a positive/negative correlation coefficient indicates an increase/decrease of features value as approaching FOG. Finally, the latency between the pre-FOG detection

and the actual FOG occurrence was computed. The above-mentioned analysis was performed on data related to patients' ON and OFF therapy separately and the results were compared. Then, the final model configuration was trained with patients ON (OFF) therapy and tested on patients OFF (ON). Finally, the performance obtained in the two testing conditions was compared.

3.1.2 Real-time detection of freezing of gait using a single accelerometer

This study aims to develop a robust, fast, and lightweight FOG detection algorithm for real-time applications in the home environment. Data from three different data sets were used for the analysis. A novel CNN was trained and optimized to maximize performance while reducing the computational complexity. A detailed performance evaluation was provided, including information regarding the detection rate, FPs number and duration, and the prediction time or detection latency. Finally, testing time, memory requirement, and battery consumption were estimated.

Subjects

In this study, three data sets were employed for the analysis. As far as concerns the main data set (REMPARK), only binary (i.e., FOG, non-FOG) class labels were available. Given a large number of FOG episodes recorded, it was used for training, validation, and testing of the classification model. The additional data sets (6MWT,ADL) yield the advantage of including class labels for different activities. The first one includes a large amount of gait data and some FOG episodes. In the second one, despite no FOG episodes being recorded during the data collection procedures, several ADLs were performed during the recordings. Thus, this latter data set was included to test the robustness of the FOG detection algorithm to FPs, which is of fundamental importance for a real-life-oriented detection tool. Both data sets were used in this study as independent test sets. A summary of the characteristics of the three data sets is reported in Table 3.4, together with the list of labeled activities.

The REMPARK data set [135] includes data from 21 PD PwPD. The inclusion criteria were a clinical diagnosis of PD with motor symptoms, H&Y stage larger than

Table 3.4 Characteristics of the data sets used in this study.

Data set	REMPARK	6MWT	ADL
Subjects (% male)	21 (86%)	38 (75%)	59 (63%)
Total recording time	9.1 h	2.4 h	5.9 h
FOG duration	93 min	5.3 min	0
# FOG episodes	1058	52	0
Labeled activities	FOG non-FOG	FOG gait stance	gait stance sit sit-to-stand stand-to-sit toe tapping leg agility retropulsion test

2 in OFF state of therapy, absence of dementia or vision impairments preventing them from accomplishing the required tasks, and FOG-Q score larger than 6. Subjects needing gait assistance (e.g., walking stick, crutch) were included in the study. The experiments were conducted at patients' homes. Data were recorded both ON and OFF dopaminergic therapy. In detail, the sample included 18 males and 3 females, with age 69.3 $\pm$ 9.7, disease duration 9 $\pm$ 4.8, H&Y score 3.1 $\pm$ 0.4, FOG-Q 15.8 $\pm$ 4.1, MMSE 27.8 $\pm$ 1.9, and total MDS-UPDRS part-III 16.2 $\pm$ 9.7 in ON and 36.3 $\pm$ 14.4 in OFF.

The 6MWT data set [132, 157] includes data from 38 PwPD and 21 control subjects. The inclusion criteria for the PD sample were a clinical diagnosis of PD with motor symptoms (either with a medical history of FOG events or not), and no major comorbidities or vision/cognitive impairments preventing them from accomplishing the required tasks. Subjects needing gait assistance were included in the study. The experiments were conducted during pre-scheduled outpatient visits, and all PwPD were in a daily ON state, meaning that they had taken their usual drug dose, and a variable time had elapsed since then. The sample included 28 males and 10 females, with a mean age of 70.7 $\pm$ 8.2, disease duration of 9 $\pm$ 4.8, and H&Y score of 2.5 $\pm$ 0.8. The inclusion criteria for the controls were no clinically evident sign of Parkinsonism, no severe vision impairment, dementia, and other significant neurological disorders. Subjects needing gait assistance were included in the study.

The control sample included 7 males and 14 females, with a mean age of 85.6 $\pm$ 7.2. Being enrolled in a retirement home, the age of control subjects was significantly superior to that of PwPD. However, these controls represent a challenging situation for the detection algorithm in correctly classifying the gait features of PD patients. In fact, elderly people may exhibit challenging gait patterns, in terms of gait velocity and turn amplitude.

The ADL data set [176] includes data from 59 PwPD. The inclusion criteria were a clinical diagnosis of PD with motor symptoms, no major comorbidities, or vision/cognitive impairments preventing them from accomplishing the required tasks. Subjects needing gait assistance were included in the study. All PD participants were in a daily ON state. The sample included 37 males and 22 females, with an age of 69.2 $\pm$ 10.2, disease duration of 6.7 $\pm$ 5.3, and H&Y score of 2.14 $\pm$ 0.8.

Data acquisition

REMPARK. The performed activities included walking tasks (e.g., showing home, stand up and go test, walking outdoors) and some tasks designed for FP analysis (e.g., brushing teeth, painting/drawing/erasing on a sheet of paper, cleaning windows). Acceleration data were recorded using an IMU mounted on the left side of the waist (Figure 3.5) by means of an elastic band and locally stored on the device. Sensor range was set to $\pm$ 6g and sample rate to 200Hz, with data subsequently down-sampled to 40 Hz. A total of 9.1 h of inertial data were recorded during the experiments, including 93 minutes of FOG.

6MWT. Participants were asked to perform the 6-minute walking test (6MWT), consisting of walking back and forth along a 10-m hallway for 6 min at their preferred pace. Data from a 3-axial accelerometer and 3-axial gyroscope were recorded using a smartphone mounted on the lower back by an elastic band (Figure 3.6). A range of $\pm$ 2g and 2000 dps was used for the accelerometer and the gyroscope respectively, and a sampling rate of 200 Hz was selected. Inertial data were locally stored in the smartphone. A total of 2.4 h of inertial data were recorded from PwPD during the experiments, including 97.6 min of gait, 17.4 min of stance, and 5.3 min of FOG. An additional 1.4 h of data were recorded from control subjects, including 72 min of gait and 4 min of stance.

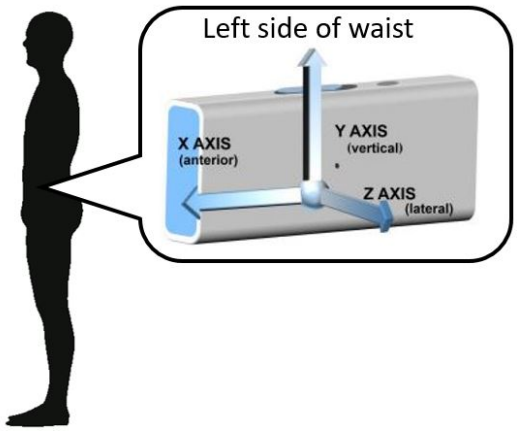


Fig. 3.5 Sensor position and axes orientation in the REMPARK data set. Anterior, vertical, and lateral (left) direction corresponds to the x-axis, y-axis, and z-axis of the sensor reference system, respectively.

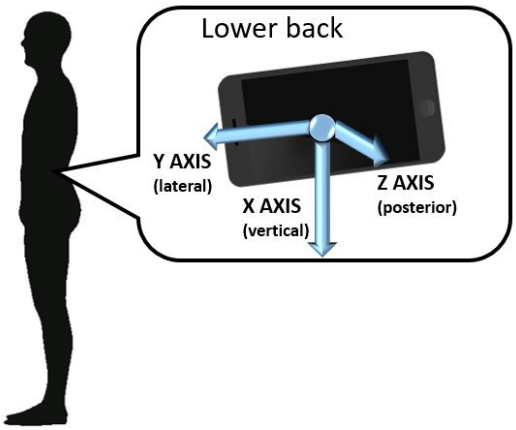


Fig. 3.6 Sensor position and axes orientation in the independent data sets (6MWT, ADL). Vertical, lateral (left), and posterior direction corresponds to the x-axis, y-axis, and z-axis of the sensor reference system, respectively.

ADL. The same sensor configuration as in the 6MWT data set was used. Experiments were conducted during outpatient visits, and participants were asked by the clinicians to perform several activities, including free walking, turning with different angular amplitudes, standing up, sitting down, standing for several seconds, and other tasks required for the MDS-UPDRS evaluation. These tasks, performed in semi-supervised conditions, are rather representative of activities carried out in a domestic environment. A total of 5.9 h of inertial data were recorded during the experiments, including 32.8 min of walking, 40.2 min of stance (i.e., sit, stand), and

13.5 min of postural transitions (i.e., standing up, sitting down), while the remaining activities included tasks related to the MDS-UPDRS evaluation and unlabeled activities.

Preprocessing

Raw data of the main data set (REMPARK) were segmented using fixed-length sliding windows of 2s. As the data set is intrinsically unbalanced, due to the different proportions of FOG and non-FOG events, a differential segmentation process was performed for training and validation set generation. The segmentation procedure, reported in Figure 3.7, consists in using different overlaps for FOG and non-FOG data. More specifically, 50% overlap (1s advance) and 87.5% overlap (0.25s advance) were used for non-FOG and FOG data, respectively. Windows including only non-FOG data were labeled as non-FOG, windows including at least 50% of FOG were labeled as FOG, and the remaining windows were discarded.

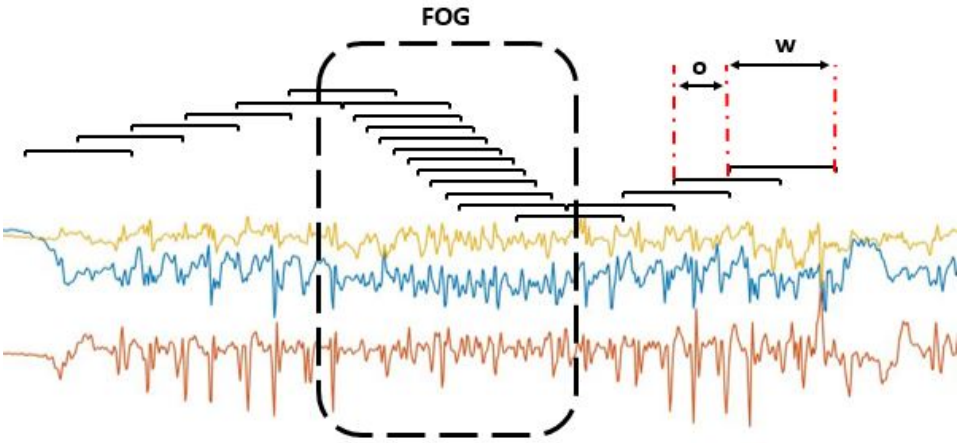


Fig. 3.7 Differential segmentation process used for training and validation set generation. Window size (w) and overlap (o) are different during FOG and other activities.

The result of the segmentation process is reported in Figure 3.8 (right), in terms of the proportion of FOG and non-FOG instances. As evident from Figure 3.8 (left), using a fixed-length overlap generates an unbalanced training set, with 75% and 25% of non-FOG and FOG instances, respectively. On the contrary, the implemented differential segmentation procedure provided a balanced distribution of non-FOG and FOG instances, i.e., 52% and 48%, respectively. As for the test set generation,

segmentation was performed using a fixed overlap of 75% (0.5s advance), resembling the actual working condition of the FOG detection system (every 0.5s the algorithm processes data from the preceding 2s). After segmentation, mean-removal was performed on each window separately, in order to allow the classification model to work properly with centered data.

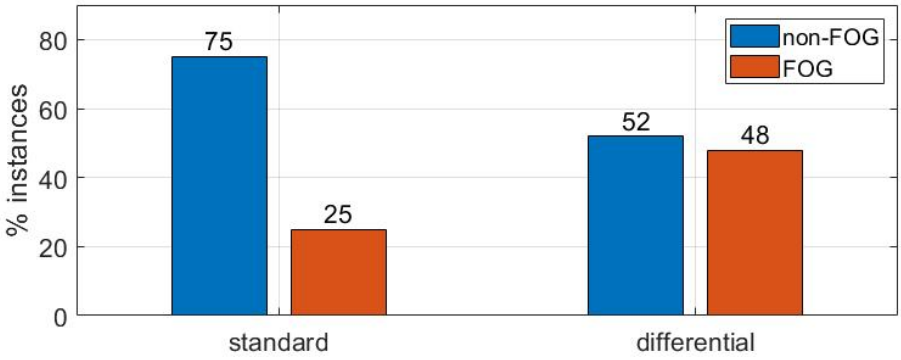


Fig. 3.8 Proportion of FOG using standard segmentation (left) and class balancing obtained from the differential segmentation procedure (right).

As for the independent data sets (6MWT, ADL), the axes' orientation and sampling rate were different from those of the main data set. Hence the following procedures were performed to achieve uniform data. First, data were resampled at 40 Hz, by under-sampling the original data collected at 200 Hz. Then, the order of the axes was adjusted to match that of the main database. After data resampling and reshaping, segmentation was performed using 2s sliding windows with 75% overlap (0.5s advance), as done for the test set of the main data set. Finally, the mean value was removed from each window separately.

In order to prepare the data for the subsequent optimization and testing procedures, the entire REMPARK data set was initially split into training, validation, and test set, including respectively 12, 4, and 5 patients (approximately 60% for training, 20% for validation, and the remaining 20% for testing). The subsets were generated so that patients in the training and test sets had similar characteristics in terms of age, disease duration, H&Y, MMSE, and MDS-UPDRS-III, while PD patients with more severe FOG were assigned to the test set. This represents a conservative situation, useful to test the generalization capability of the algorithm when data from patients with severe gait impairment are input to the classification model. The optimization of the classification model architecture, parameters, and learning process settings

was performed using the training and validation sets. Then, the optimized model was tested on the test set and further tested on the additional independent data sets.

Classification

Classification model. The simplest CNN architecture includes a convolutional layer followed by a pooling, subsequent flattened and dense layers, and finally a softmax classification layer. Starting from this configuration, additional convolutional, pooling, and dense layers were iteratively added to the architecture. Moreover, different heads - CNN were used to obtain different spatial resolutions in the analysis of the input signals to capture useful features, from local to global level [250].

Optimization. For each configuration, a grid search procedure was used to optimize the hyper-parameters (i.e., number of filters and kernel size for convolutional layers; type, size, and stride for pooling layers; number of neurons for dense layers). Finally, the dropout rate and regularization parameters were adjusted to optimize the training-validation performance. AUC and memory requirements were monitored during the validation stage, in order to identify the architecture providing the best performance while not significantly increasing the computational burden. Specifically, an increase in AUC by at least 1% was deemed necessary to justify a significant increase in computational complexity, whereas no threshold was set in case of a reduction in computational complexity. The range of values and the step size used for parameter optimization are reported in Table 3.5. While some parameters were optimized using the grid-search tuning process (i.e., number of filters, filter size, pool size, pool stride, number of neurons in the dense layers), others were manually adjusted to ensure a proper training process and to increase performance. The parameters range and the respective step size were selected considering both studies focusing on human activity recognition tasks [251, 252] and literature works proposing FOG detection algorithms [158, 217]. However, the range was limited in some cases in order to control the model complexity. Specifically, the upper limit for the number of filters, the number of dense layers, the number of neurons in the dense layers, and the number of convolutional heads were limited to 32, 3, 128, and 3, respectively.

Performance evaluation. A comprehensive performance evaluation procedure was implemented to assess the classification results of the proposed model. Sensitivity, specificity, accuracy, F-score, AUC, and EER were computed. Moreover, the

Table 3.5 Range of values and steps used for the optimization of the model architecture, model training, and regularization parameters. Some parameters were tuned using the automatic grid-search optimization procedure (top), while others were manually adjusted (bottom).

Parameter	Range	Step
# filters	4-32	4
filter size	3-39	3
pool size	2-3	1
pool stride	2-3	1
# neurons	[16, 32, 64, 128]	-
# dense layers	1-3	1
# convolutional heads	1-3	1
pool type	[average, max]	-
dropout rate	0.2-0.8	0.1
regularization term	[0.001, 0.01, 0.1]	-
learning rate	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	-
batch size	[64, 128, 256, 512, 1024]	-
# training epochs	20-250	10

geometric mean of sensitivity and specificity (Equation 3.1) is useful to appreciate the situations in which one of the two measures is far smaller than the other.

$$geometric-mean = \sqrt{sensitivity \cdot specificity} \tag{3.1}$$

In order to provide further details on the prediction performance of the classification algorithm, some post-processing procedures were performed using predictions and class labels. As far as concerns the REMPARK data set, only a binary class label was available, being either 1 or 0 in the case of FOG or activities other than FOG, respectively. As for the additional data sets, several class labels were provided, including gait and stance for both the 6MWT and ADL data sets, with the latter including also labels for sitting down, standing up, and some MDS-UPDRS-related tasks.

Postprocessing. The following measures were computed for all the data sets. First, the percentage of FOG episodes detected was computed as the number of real FOG episodes in which at least one data window was classified as FOG. The proportion of FOG windows detected within each episode represents a complementary

measure with respect to the percentage of FOG episodes detected. It was computed as the percentage of true FOG windows classified as FOG by the algorithm in each episode (Figure 3.9 B). As far as concerns the FPs, their number, duration (Figure 3.9 D), and the distance of false FOG episodes from the real FOG (Figure 3.9 C) were computed for two out of three data sets, where FOG data was available. Finally, FOG detection latency represents the temporal resolution in detecting FOG episodes. It was computed as the difference between the onset of the real and detected FOG episodes (Figure 3.9 A). Such measure is expressed in seconds (s) and can be either negative or positive, depending on the detection being predicted or delayed with respect to the true episode. As far as concerns the data sets for which the activity label was provided, additional analyses were performed to assess FPs. More specifically, the occurrence of false FOG episodes was counted for each activity. This is important to assess which activities are most frequently misclassified by the algorithm.

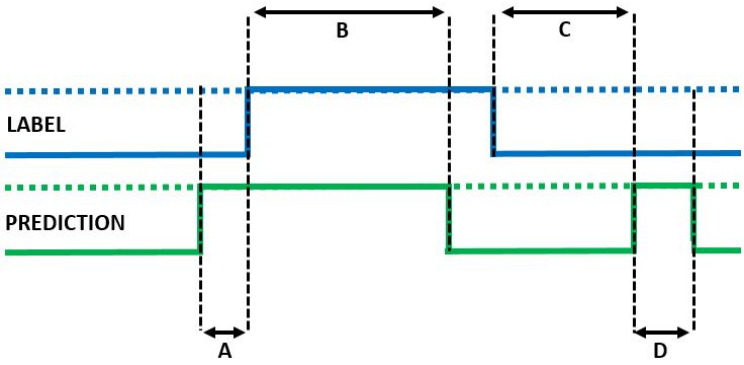


Fig. 3.9 Schematic of the measures computed for post-processing analysis. A: prediction time; B: amount of FOG detected in the episode; C: distance between false FOG episode and the nearest real FOG episode; D: duration of false FOG episode.

Computational complexity. As the FOG detection algorithm is required to work in real-time, testing time was computed for different input data dimensions. It was computed separately for the preprocessing steps (i.e., reshape, mean removal) and for the classification step. Moreover, in order to check the feasibility of implementing the algorithm in a stand-alone device, memory requirements were calculated for input data and classification model parameter storage. For further analysis of the computational complexity, the floating-point operations (FLOPs) were evaluated on the multi-head CNN model. Finally, in order to provide an estimation of the battery consumption, the computational complexity and the processing time of the proposed

algorithm were compared to those of the detection model presently embedded in the commercial STAT-ON monitoring device [135].

The effect of activity threshold. The FOG detection algorithm was designed to process and classify every single time window. However, this does not represent the most energy-efficient solution for data analysis in real-life. Instead, a simple thresholding method could be used for distinguishing activity from inactivity periods. Then, the designed algorithm can be run only during activity periods. In this way, a percentage of FPs (registered in this study during inactivity periods) may be avoided. Moreover, the energy consumption would be significantly reduced, as inactivity data are processed using a simple threshold approach instead of a DL algorithm. To this end, the magnitude M of the 3D acceleration signal for each window j was computed according to Equation 3.2, where α_x, α_y, and α_z represent the acceleration signal along each axis, and the sum is performed for every sample i of every window of length w.

$$M_j = \sqrt{\sum_{i=1}^{w} (\alpha_{x_i}^2 + \alpha_{y_i}^2 + \alpha_{z_i}^2)} \tag{3.2}$$

Then, the performance and the percentage of discarded windows were computed on the validation set of the REMPARK data set. The threshold was selected in order to discard data windows without degrading the performance of the algorithm. Finally, the effect of the magnitude thresholding was evaluated separately on the test set of the REMPARK data set, and on the 6MWT and ADL data sets.

The experiments were performed on a computer with a 2.3 GHz processor, 8 GB RAM, and 4 GB GPU. Preprocessing and post-processing were performed using MATLAB (version R2020a), while training, optimization, and testing of the classification model were performed in Python (version 3.6), using Keras (version 2.4), Keras-flops (version 0.1.2), and TensorFlow (version 2.3) libraries.

3.2 Postural stability

The following two studies aim to automatically evaluate postural stability in PwPD. In both studies, a single smartphone placed on the lower back was used for data acquisition. In the first work (Section 3.2.1), data related to turning were analyzed, aiming both to distinguish patients with different disease progression and to provide

an index correlated with the level of postural stability. In the second work (Section 3.2.2), participants were asked to keep a static upright position. ML models were optimized to classify patients according to the clinical postural stability score.

3.2.1 Turn quality and postural stability assessment using smartphones

Despite a large number of studies focused on gait analysis during straight walking, only a few of them analyzed the gait pattern during turning. Moreover, most of these latter performed statistical tests to highlight differences in populations (e.g., PwPD vs healthy controls) or in the level of motor impairment. In the present study, data recorded from a single smartphone was used to develop a binary classification model for distinguishing patients with different levels of motor impairment. Moreover, an index for assessing the quality of movement (QoM) and the level of postural stability was proposed.

Subjects

The study was performed at the Regional Reference Center for Parkinson's Disease and Movement Disorders, University Hospital *Città della Salute e della Scienza*, Turin, Italy. A total number of 72 PwPD were enrolled in the experiment. About 60% were males, and this reflects the prevalence of this pathology, which is unbalanced between the two genders. However, the demographic and clinical characteristics are not significantly different between the two groups; hence, for the sake of brevity, these were not separately reported. The inclusion criteria were a clinical diagnosis of PD with motor symptoms, no major comorbidities or vision/cognitive impairments preventing them from accomplishing the required tasks. All PD participants were in *daily ON* state, meaning that they had taken their usual drug dose, and a variable time had elapsed since then. The first group of patients has been enrolled during the pre-scheduled follow-up clinical visit. Hence, detailed clinical scores related to the entire MDS-UPDRS part III were available, measured simultaneously with the experiments by clinicians with expertise in movement disorders. The second set of 13 PwPD was explicitly enrolled for the experiments, independently of their periodical follow-up neurological visit. The main characteristics of the two patient subgroups (labeled G-PD1 and G-PD2 respectively) are summarized in Tables 3.6

and 3.7. Given that patients from G-PD2 did not undergo a simultaneous complete neurological examination, the whole MDS-UPDRS part III was not evaluated. However, a recent H&Y score was available from anamnestic data and confirmed by the clinicians supervising the experiments. The patients were binary classified into two groups: *mild (M)* vs *moderate/severe (S)* motor and postural conditions. The first class corresponds to H&Y less than or equal to 2, whereas the second class encompasses patients with H&Y larger than 2. It is worth noticing that, as the objective was to devise an index capable of correlating with the motor status and the disease progression of PD patients, all the experiments were carried out on patients characterized by different stages of the disease. The inclusion of a control group was not deemed pertinent, whereas the generalization to other conditions (e.g. frailty in the elderly population), although extremely interesting and challenging, is beyond the scope of this experiment.

Table 3.6 Demographic and clinical characteristics of PD patients enrolled during prescheduled outpatients visit (G-PD1).

# subjects (male)	Age (years)	Disease duration	H&Y
59 (37)	69.2 ± 10.2	6.7 ± 5.3	2.14 ± 0.8

Table 3.7 Demographic and clinical characteristics of PD patients performing the 6MWT (G-PD2).

Condition	# subjects (male)	Age (years)	Disease duration	H&Y
Mild (M)	6 (4)	68.2 ± 3.9	6.2 ± 1.7	≤ 2
Moderate/Severe (S)	7 (5)	75.2 ± 5.3	15.2 ± 4.5	> 2

The study was conducted in accordance with the Declaration of Helsinki and approved by the local Ethics Committee. Participants received detailed information on the study purposes and execution, and written informed consent for the observational study was obtained. Demographic and clinical data were noted anonymously. Tests were performed under the supervision of clinical personnel to ensure patients' safety.

Data acquisition

Participants were asked to perform walking tests encompassing several 180° turns in both directions while wearing a smartphone, in order to measure inertial data and

work out the QoM index.

For G-PD1, inertial data were registered during the entire visiting time. Patients were equipped with the smartphone secured around the third lumbar vertebra with an elastic band, ensuring its adherence to the body (see Fig. 3.10). Then, they were asked by the clinicians to perform several activities relevant to the MDS-UPDRS scoring, including free walking, turning with different angular amplitudes, standing up, sitting down, standing for several seconds, and so on. These tasks, performed in semi-supervised conditions, are rather representative of activities carried out in a domestic environment.

Fig. 3.10 Smartphone position adopted for the experiments.

Patients from G-PD2 were asked to perform a 6MWT, selected for its easy setup, patient's tolerance and reproducibility [253], and, most importantly, its suitability for measuring multiple 180° turns in both directions. Subjects were equipped with the smartphone as previously described, and asked to walk back and forth along a 10-meter hallway for 6 minutes at their preferred pace. They were free of using their usual walking aids, quitting the test at any moment and possibly resuming it.

All patients agreed to the videotaping of the procedure after receiving suitable explanations and being guaranteed that they could be identified and the videotapes were not made available to persons different from the authorized ones. However, video recordings were not used in this work. In fact, given that the objective was to assess a QoM index using inertial data from turns, the only relevant information in

this phase was the fact that the patient was performing a turn. The annotation of this specific piece of information was deemed not sufficient to justify the time possibly spent by a clinician to review the whole video sequence, given also the fact that a clinician was in any case present during the test. In any case, the video was recorded to enable possible future research. A chronometer was run simultaneously with the test session, and turns and gait abnormalities (e.g., FOG episodes) were manually annotated by the clinical and technical personnel. Inertial data from the embedded sensors were collected and locally stored by means of SensorLog, a commercial app for Android 6.0. Once collected, data were exported in CSV format and processed offline using `MATLAB` version 2018a for Windows 10.

Preprocessing

Turn detection. Based on the annotations of the registered activities, turning events were identified using a graphical user interface (GUI) developed in `MATLAB`. In order to avoid confounding factors and increase the detection performance in view of an automated turn detection, only 180° turns were considered. It is worth noticing that 180° turns allow insights into the patient's motor condition. Figure 3.11 reports sample amplitudes of the acceleration signal. The correlation between axes decreases during turning, compared to straight walking. Moreover, a neat increasing and subsequently decreasing pattern in the x-axis angular velocity (black line) can be observed. The area under this curve yields information on the amplitude of the turn, whereas the green line denotes the identified signal segments related to turns.

Filtering. Signals were re-calibrated to compensate for possible deviations of sensors from the initial position and to make gait signal patterns uniform. Following the method proposed in [254], the 3-axis accelerometer orientation correction was performed by applying a quaternion rotation transformation to the device raw data. After mean removal, inertial signals were de-trended and filtered with a second-order zero-lag Butterworth low-pass filter with a cut-off frequency of 20 Hz.

Feature extraction. The set of features reported in Table 3.8 was extracted from each turn of every participant. The reported features were identified from literature studies for their recognized capability to describe gait characteristics, postural control, and motor impairment in healthy, frail, and/or PD subjects, also affected by complications such as FOG. All acceleration (α_x,α_y,α_z) and angular velocity (ω_x,ω_y,ω_z) components were kept separated. Step and stride variability, step

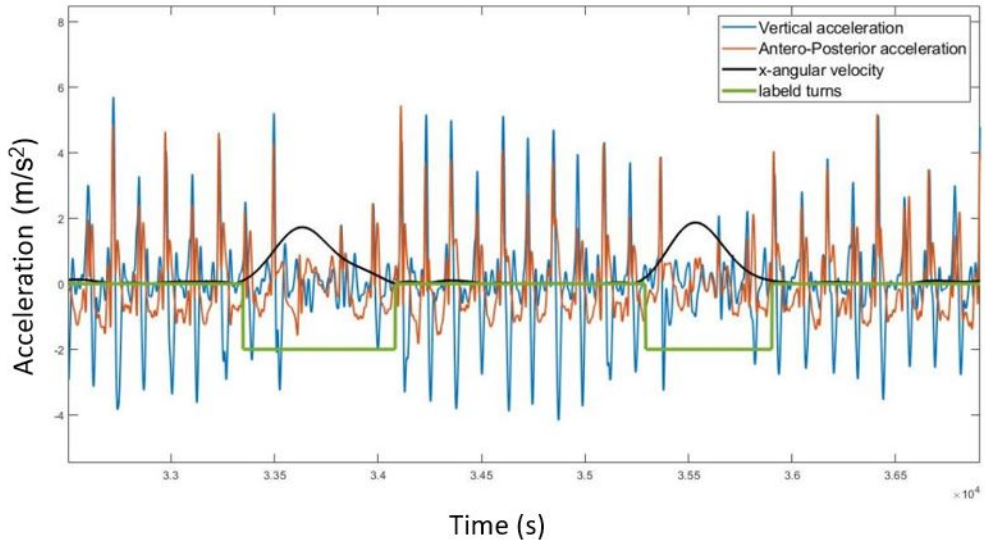


Fig. 3.11 Acceleration and angular velocity signals related to walking and turning activities. For better visualization, the x-component of angular velocity was low-pass filtered (fc = 0.5 Hz), removing high-frequency noise.

and stride time, and gait symmetry were computed from the vertical component of the acceleration signal. All the other features were computed for all components of the accelerometer and gyroscope, generating a final set of 77 features. The spectral entropy is the Shannon entropy computed on the signal FFT, whereas the weighted power spectrum peaks represent the product of amplitude and frequency values corresponding to the dominant harmonic in the FFT domain. The other features are self-explaining.

Feature selection. The Pearson correlation coefficient r was computed between each feature reported in Table 3.8 and the class of patients from G-PD2, i.e., 0 and 1 for mild and severe motor impairment, respectively. Then, only features with a mild-to-moderate correlation (i.e., $r > 0.4$) were selected for the subsequent binary classification task. Moreover, features exhibiting strong mutual correlation were discarded, in order to keep only relevant and non-redundant features.

Classification

A binary classification algorithm was addressed, capable of sorting patients into two classes: *mild (M)* and *moderate/severe (S)* motor conditions, defined as a H&Y

Table 3.8 List of all features extracted for each turn.

Feature ID	Feature	Source
1	Root Mean Square	[255–257]
2	Range	[258]
3	Standard Deviation	[255, 256]
4	Jerk	[256]
5	Normalized jerk	[256]
6	Spectral Entropy	[259]
7	Spectrum Peaks	[260, 259]
8	Normalyzed Spectrum Peaks	[259]
9	Weighted Power Spectrum Peaks	[260, 259]
10	Harmonic Index	[261]
11	Low Power Frequency	[262]
12	Step variability	[263, 36]
13	Stride variability	[260, 262, 263]
14	Symmetry	[262, 263]
15	Step time	[263]
16	Stride time	[262, 263]
17	Dominant frequency	[260, 259, 36]

stage ≤ 2 and > 2 respectively. Different ML models were considered, namely kNN, SVM, and LDA. For each model, the main parameters were exhaustively tuned, namely: kernel type, kernel scale, and cost for SVM; number of neighbors, distance metric, and distance weight for kNN; gamma and delta parameters for LDA. The classifiers were trained using G-PD2 patients, i.e. those who performed the 6MWT and whose binary classification based on the H&Y scale was available. The optimization was performed by minimizing the 10-fold CV error (i.e. maximizing the accuracy), and the performance was evaluated in terms of accuracy and AUC. Moreover, the confusion matrix was computed to assess the number of FPs and FNs. The ML model providing the best performance was selected and tested on patients from the G-PD1 group. Then, the continuous output provided by the model was used as QoM and correlated with the clinical scores. Specifically, the QoM is defined as the *a posteriori* probability of belonging to either class, averaged on all turns performed by each patient involved in the experiment. The idea is to create a scale to measure the degree of postural instability. The lower and upper limits of the scale are obtained using the binary (i.e., 0 or 1) classification between mild and severe motor impairment (G-PD2). Then, the additional patients (G-PD1) are represented

on such a scale, using the continuous output provided by the classification model. Specifically, the closer the output is to 1, the more severe the motor condition of the patient is.

3.2.2 Postural stability assessment during quiet stance using smartphones

In this work, both healthy young subjects and PwPD were enrolled and asked to keep a static upright position with a smartphone secured on their lower back. Several time- and frequency-domain features were extracted and selected, and different ML models were optimized to classify subjects according to their level of postural stability.

Subjects

Data acquisition was carried out at the Regional Reference Center for Parkinson's Disease and Movement Disorders, Azienda Ospedaliera Universitaria em Città della Salute e della Scienza, Turin (Italy). The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee. Participants received detailed information about the purpose and execution of the study and written informed consent was obtained for the observational study. Demographic and clinical data were noted anonymously. Patients consented to the videotaping of the procedure after receiving appropriate explanations and assurances of privacy. The experiments were conducted in the hospital during periodically scheduled outpatient visits; therefore, patient safety was ensured by the presence of medical staff. A total of 42 PwPD were recruited. Inclusion criteria were a clinical diagnosis of idiopathic PD with motor signs and symptoms, absence of major cognitive impairment or other conditions that prevented the patient from performing the task correctly, ability to maintain an upright position without assistance for at least one minute, and absence of dyskinesias and other comorbidities or conditions affecting balance. Because the experiments were conducted during an outpatient visit, most patients were in the daily on condition. Data acquisition was also performed on 7 healthy young subjects. The choice of a control population that did not correspond to the age-matched PD sample was dictated by the need to select some subjects with recognized optimal postural control. This made it possible to define a scale on which to place the different levels of postural stability, with the controls representing the best possible

value. The number of controls was chosen to match that of PwPDs with the worst possible level of postural control. The population characteristics are summarized in Table 3.9 for all PwPD and control subjects, while in Table 3.10 they are divided according to the postural stability score.

Table 3.9 Demographic and clinical characteristics of subjects involved in this study. Measures are reported in terms of mean $\pm$ standard deviation. n.a.: not applicable

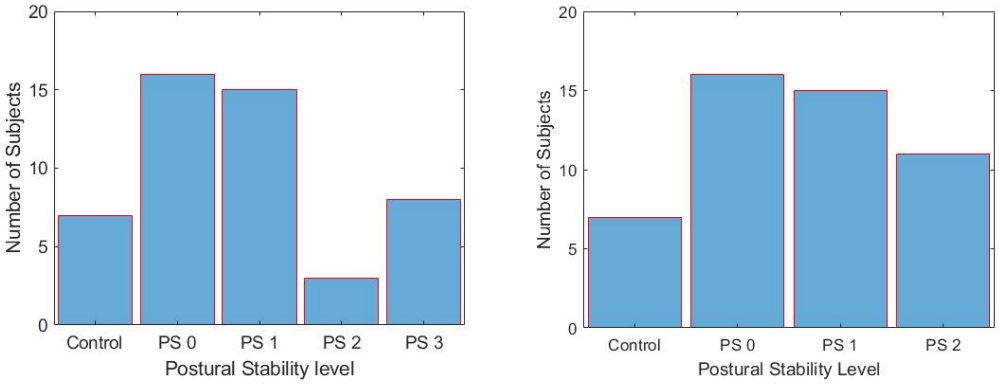
Sample	# Subjects (male)	Age (years)	Disease duration (years)	H&Y	P.S. score
PD	42 (31)	68.6 ± 10.7	10.3 ± 6.6	2.3 ± 0.6	1.1 ± 1
Control	7 (5)	27.2 ± 2	n.a	n.a	n.a

Table 3.10 Demographic and clinical characteristics of PD subjects divided for PS score. Measures are reported in terms of mean (range).

P.S. score	# Subjects (male)	Age (years)	Disease duration (years)	H&Y
PS 0	16 (10)	64.2 (52-83)	8.2 (2-21)	1.4 (1-2)
PS 1	15 (10)	69.3 (53-87)	9.3 (1-22)	2 (2)
PS 2	3 (3)	78 (74-81)	15.5 (11-20)	3 (3)
PS 3	8 (6)	67.7 (55-81)	18.4 (12-24)	3 (3)

Clinical assessment of postural stability (PS) was performed by experienced neurologists using the retropulsion test (MDS-UPDRS Part III, item 3.12). Neurologists assigned a score between 0 and 4, following the recommendations of the MDS-UPDRS. Based on this clinical score, PwPD were divided into classes. The distribution of patients in each class, along with the control population, is shown in Figure 3.12. As can be seen from Figure 3.12-a, no patient is included in class 4, despite the fact that the score of item 3.12 ranges from zero to four. This is in line with the MDS-UPDRS recommendations. In fact, a score of 4 is assigned in case the subject is largely unstable and unable to regain stability after losing it. Therefore, such patients usually do not perform the retropulsion test. As for patients in classes 0 and 1, they take a maximum of 2 and 5 steps to recover balance, respectively. As for class 2 and 3 patients, they have a drastic lack of postural reflexes. Class 2 patients should be able to recover their balance by taking a maximum of 5 steps backward, while Class 3 patients should not. However, it is not common among neurologists to wait for patients to take 5 steps backward before grasping them safely. As a result, classes 2 and 3 are largely overlapping and the distinction between the two is somewhat arbitrary. Moreover, the variance v of the data distribution calculated

both by keeping classes 2 and 3 separate and by joining them together revealed a negligible difference, i.e., $deltav < 1\%$. Therefore, in agreement with the expert neurologists who participated in this study, classes 2 and 3 were merged into a single class, referred to as 2 in the remainder of this paper. The resulting distribution is shown in Figure 3.12-b.



(a) *Original* (b) *Class 2 and 3 merged*

Fig. 3.12 Distribution of subjects based on clinical PS score.

Data acquisition

Data acquisition was performed using inertial sensors, i.e., triaxial accelerometer and triaxial gyroscope, embedded in a commercial smartphone. The smartphone was placed inside an elastic band and attached to the patient's lower back at the L3-L5 level. The smartphone recorded and stored inertial data locally via the SensorLog app for Android 6.0. Once collected, the data were exported to CSV format and processed offline using *MATLAB*, version 2019b for Windows 10. The recorded data were analyzed visually and computationally to verify that they were not limited by the operating system or application used. Subjects were asked to maintain a stable upright position, with feet about 10 cm apart, arms relaxed along the body, and eyes open. Data recording was done once for at least 30 minutes. First, the technical characteristics of the smartphone used were analyzed in terms of resolution and noise. These specifications are very important in this context, since the experimental protocol involves only a stance phase, with subjects maintaining a static position. Consequently, it is necessary to appreciate the small variations of the inertial sensors, which require adequate sensor resolution and noise levels. Table 3.11 shows the

technical specifications of the sensors embedded in the Samsung Galaxy S5 mini smartphone used in the experiments, in terms of range, resolution/sensitivity, noise, and sampling rate. The minimum range of the accelerometer ($\pm$ 2g) and gyroscope ($\pm$ 250dps) is more than adequate for analyzing static activities, and sensor noise was found to be negligible if compared to the expected signal amplitude, obtained by performing some preliminary data acquisition tasks.

Table 3.11 Smartphone embedded inertial sensors specifications. Fs: sampling frequency.

Sensor	Range (min,max)	Resolution	Noise	Fs
Accelerometer	$\pm$2 g, $\pm$16 g	16 bits	70 ($\frac{\mu g}{rt-Hz}$)	200 Hz
Gyroscope	$\pm$250 dps, $\pm$2000 dps	16 bits	3.8 ($\frac{mdps}{rt-Hz}$)	200 Hz

Preprocessing

Feature extraction. A wide range of features in the time and frequency domains was extracted. An extensive literature search, together with a visual inspection of the signals, resulted in the collection of 414 features (i.e., 69 features for each acceleration and angular velocity component) from the recorded data. The complete list of features is given in Table 3.12, along with a brief description. Note that, for the sake of brevity, when a given feature is calculated on different frequency bands (e.g. RAPP, N_p band), it is reported only once in the table. Some features (e.g., N_p, F_0 amplitude, F_0 width) were selected for their ability to represent the spectral characteristics of the signal. Other parameters (e.g., P_b, RAPP, N_p band) were selected after visual inspection of the power spectral density (PSD) of signals grouped by class, as they were considered significant for capturing intra-class similarities and inter-class differences. For example, the PSD of class 3 has a higher peak and wider distribution along the signal bandwidth than the other classes. Specifically, in classes 0 and 1, most of the signal power is in the band below 1 Hz, while a shift toward higher frequencies (i.e., up to 5 Hz) was observed for class 2 data. Based on these observations, features were extracted that can represent the signal power in some specific frequency bands (e.g., P_b [0-1] Hz, P_b [1-2] Hz), along with the power ratio between different bands (e.g., RAPP [0-1]/[1-5] Hz), and the number of spectral peaks in some specific bands (e.g., N_p band [0-1] Hz, N_p band [1-5] Hz).

Table 3.12 Data set of features extracted from each component of acceleration and angular velocity signals

ID	Study	Feature (Description)
1,2,3	Present Study	Mean, SD, Variance
4,5,6	[264],[265],[266]	Minimum, Maximum, Range
7	[267],[175]	Root Mean Square
8	[267],[173]	Average COM velocity
9	[173],[266]	Jerk (measure of postural adjustments)
10	Present Study	Mean Jerk
11	[268]	MAD (mean absolute deviation)
12	[268]	MALA (mean absolute linear acceleration)
12	[268]	SMA (simple moving average)
13	[268]	ZCR (zero crossing rate)
14	Present Study	Displacement
15,16	Present Study	Kurtosis, Skewness
17,18	[267],[266]	Total Power, Maximum Power
19,20	[267]	Mean Frequency, Peak Frequency
21	Present Study	Shannon spectral entropy
22,23	[267] [173]	F_{50}, F_{95} (F_n is the frequency value below which is present n% of total power)
24	Present Study	P_b (power in different bands)
25	Present Study	RAPP (ratio of power in different bands)
26	Present Study	N_p (number of spectral peaks)
27	Present Study	$N_p band$ (number of spectral peaks in different bands)
28	Present Study	F_0 amplitude (amplitude of the fundamental harmonic)
29	Present Study	F_0 width (width of the fundamental harmonic)
30	Present Study	F_0 (frequency-value of the fundamental harmonic)
31	Present Study	$N_p th$ (number of peaks over 50% of the maximum value)

Feature selection. The purpose of this task is to identify the most significant features, defined as those that obtain the highest correlation with the clinical score. Starting from the entire set of features, features with high correlation with the target (most significant) and low correlation with each other (non-redundant) were selected. For

this purpose, the vector containing all N features is defined as $\underline{f} = \{f_1, f_2, \cdots, f_N\}$. The vector $\underline{r}_{ft} = \{r_{f_1 t}, r_{f_2 t}, \cdots, r_{f_N t}\}$ contains the Pearson correlation coefficient r between the features and the target t, while the correlation between each pair of features is defined as $\underline{r}_{ff} = \{r_{f_i f_j}\}, i, j = 1, \cdots, N$. Given the large dimensionality of the initial data set, features with a $r < 0.4$ were discarded, as this represents a weak correlation. Therefore, redundant features were removed, keeping only those features that achieve a r_{ft} much higher than the maximum correlation with the other features ($\underline{r}_{ff}$). For clarity, the feature selection procedure is described in Algorithm 2.

Algorithm 2 Feature selection Algorithm

Require: $f_1, f_2, \ldots f_n, t$ ▷ N features, target vector

Ensure: $f_1, f_2, \ldots f_m$ ▷ M<N features

procedure FS($f_1, f_2, \ldots f_n, t$) ▷ Feature Selection function

while $i < N$ **do** ▷ Iterate for each feature

if $r_{f_i t} > 0.4$ & $r_{f_i t} > 1.33 \cdot max_{j \neq i}\{r_{f_i f_j}\}$ ▷ Look for significant and non-redundant features

select f_i ▷ Add i^{th} to the feature subset

end while

$f_1, f_2, \ldots f_m$ ▷ End up with a subset of M features

end procedure
=0

To test whether further dimensionality reduction was possible, PCA was performed on the resulting subset of features after checking the normal distribution of each feature. The first three principal components explained 82% of the total variance; this was not considered sufficient to justify further dimensionality reduction of the feature subset. The feature selection process concluded with 8 features, reported in Table 4.27, along with the original ID, source, component specification, and correlation coefficient.

Classification

The selection of one of the many available ML algorithms is often inappropriate and/or difficult to justify. For this reason, several ML models were implemented, including kNN, DT, SVM, and RF. The models were fed with the entire initial data set, i.e., all features extracted from all components of subjects belonging to all classes. For each, the optimization procedure was based on a Bayesian approach aimed at minimizing the misclassification rate; the number of iterations was set at 30.

First, a multi-class classification problem was set up. However, considering the simple data acquisition protocol, i.e., subjects at rest in the standing position with a single smartphone mounted at the waist, multi-class classification (i.e., able to distinguish both controls from PD subjects and PD patients with different PS scores) is presumably very challenging. Therefore, the multi-class classification problem was divided into multiple binary classification tasks [269], in order to take advantage of the well-known generalization ability of SVM, which has been used in several literature studies on PD [248, 270, 135]. In this study, a new approach was proposed to address multi-class classification problems. It consists of using a first classification layer, employed to obtain a coarse assessment of postural stability. Subsequently, a finer classification is performed in the second layer. em First layer. This classification step aims to create a scale whose lower and upper limits represent subjects with extreme (optimal or severely impaired) postural control. Given the availability of inertial data from the control population, a binary classification problem was set up, employing control subjects (i.e., people with the best possible postural control) and class 2 PwPD (i.e., people with the worst level of postural control among the considered population) as elements of the two classes. The performance of the ML models was evaluated and eventually the one that provided the best performance was selected. Once the model was trained and subjected to a LOSO validation, a subsequent test was performed on PwPD belonging to classes 0 and 1. Then, the soft output, i.e., the *a posteriori* probability that a data point belonged to one of the two classes, was obtained from all subjects tested. The correlation between this soft parameter and the labeled clinical classes was calculated in order to evaluate the accuracy of this measure in classification tasks.

em Second Layer. To go beyond the simple correlation value and perform finer classification, the algorithm was refined by implementing three classifiers, thus reducing the initial single multi-class problem to three binary classification tasks.

The input to each SVM was the entire set of features, with reference only to the classes that were to be distinguished by the specific SVM model (e.g., classes 0 and 2 were input to the classifier that was to distinguish subjects in class 2 from those in class 0).

3.3 Gait impairment

The following two studies aim to automatically assess gait impairment in PwPD. The studies share subjects, data acquisition procedures, and preprocessing steps. Specifically, participants were equipped with a sensor on the thigh and asked to perform the TUG, both in ON and OFF conditions. While the objective of the first work (Section 3.3.1) is to predict the score provided by the clinical assessment, the second work (Section 3.3.2) aimed to assess the self-perceived gait impairment, i.e., that reported by patients.

3.3.1 Prediction of postural instability and gait difficulty using a single inertial measurement unit

In this study, patients were equipped with a single IMU on the thigh and asked to perform a TUG test. The walking pattern of PwPD was analyzed using different feature sets and dimensionality reduction methods. A regression model was optimized for predicting the level of gait impairment and postural stability in PD, further evaluating the effect of therapy and FOG.

Subjects

Thirty-one PwPD were enrolled by the Movement Disorders Outpatient Clinic of the Department of Human Neuroscience at the Sapienza University of Rome, based on the following inclusion criteria: diagnosis of idiopathic PD; absence of dementia (i.e., MMSE > 24); ability to walk independently; absence of neurological, orthopedic, or rheumatic comorbidities that may affect walking. Two neurologists with expertise in movement disorders verified the presence of FOG in all patients. Overall, 17 PwPD showed definite FOG (*FOG+*), while the other 14 never experienced FOG (*FOG-*).

The following standardized scales were used for clinical assessment: H&Y, MDS-UPDRS part III, FOG-Q, MMSE, FAB, HAM-D, and BAI. To further assess axial impairment, the postural stability and gait difficulty (PIGD) score was calculated, measured as the sum of items 2.12, 2.13, 3.10, 3.11, and 3.12 of the MDS-UPDRS, in both the OFF and ON states of therapy. These items are related to ambulation, balance, and FOG, both reported by patients and assessed during a direct clinical examination. Patients were clinically evaluated both in the OFF state (i.e., after discontinuation of L-Dopa for at least 12 hours) and in the ON state (1 hour after taking L-Dopa). In addition, LEDDs were calculated for each patient according to standardized procedures. The demographic and clinical characteristics of the PwPD enrolled in this study are summarized in Table 3.13. The experimental procedures were approved by the institutional review board and performed according to the Declaration of Helsinki.

Table 3.13 Demographic and clinical features of patients enrolled in the present study (mean $\pm$ standard deviation). Measures are reported in terms of mean $\pm$ standard deviation. H&Y: Hoehn and Yahr; MDS-UPDRS-III: Movement Disorder Society—unified Parkinson's disease rating scale part III; OFF: not under dopaminergic therapy; ON: under dopaminergic therapy; PIGD: postural instability/gait difficulty score.

# Subjects (Male)	31 (23)
Age (Years)	71.9 ± 6.9
Disease Duration (Years)	10.9 ± 5.9
Hoehn & Yahr score	2.4 ± 0.8
Mini-mental state examination	28.1 ± 1.9
Frontal assessment battery	14.7 ± 2.8
Hamilton depression scale	12.9 ± 6.8
Levodopa equivalent daily dose (mg)	819 ± 406
MDS-UPDRS-III OFF	35.9 ± 13.9
MDS-UPDRS-III ON	27.9 ± 13.7
PIGD OFF	7.3 ± 5.7
PIGD ON	6.3 ± 4.6

Data acquisition

Patients were asked to perform a 7 m TUG test that consisted of the following procedures: get up from a chair, walk in a straight line for 7 m, turn around, walk back, and sit down. To maximize the ecological value of the recordings and trigger

the possible occurrence of FOG, the 7-m TUG test was performed in a free house-like environment with a number of factors simulating a home environment (e.g., a transition from a spacious room to a narrow, furnished hallway with the interposition of an open door) [145]. The patient's gait was videotaped through a camera and monitored by a single IMU positioned and attached to the thigh through an elastic band (Figure 3.13). The positioning of the IMU on the patient's thigh was made so that when the patient was standing, the y axis represented the inverse gravity vector and the x axis was in the frontal plane. Therefore, angular velocity around the x axis provided a good representation of thigh motion during linear gait. The prototype system-on-board neMEMSi [245] from STMicroelectronics was used for the experiments. The sensor range was set up to 16 g and 2000 dps for the accelerometer and gyroscope, respectively. A sampling rate of up to 200 Hz can be used. The dimensions of the device (including the battery) are 25 mm $\times$ 30 mm $\times$ 4 mm (Figure 3.13). Table 3.14 shows the technical characteristics of the inertial sensors incorporated in the IMU (specifications refer to those established in this study). Prior to placement, a preliminary conventional calibration of the inertial sensors was performed, including software correction of the displacement of the IMU with respect to the Earth framework. Specifically, static acquisitions of the accelerometer and gyroscope data were performed as indicated in [271, 272]. The IMU was arranged in specific positions on a table. Operations to correct or align the sensor with the reference frame were performed in real-time, with the NeMEMSi transmitting data via Bluetooth to the PC. Orientation was obtained from the measurements and compared with the earth observation frame. The rotation between the sensor and earth quaternions was calculated at each position tested by the IMU and used for orientation correction. Once the calibration procedure was completed, the IMU was placed on the patient. The data obtained were sent in real-time to a personal computer through the neMEMSi Bluetooth module and saved progressively in CSV format. Each CSV file was related to a single test. The data in the CSV files were processed offline as described in the next section.

Table 3.14 Inertial sensors technical characteristics.

Sensor	Range	Sensitivity	Sampling rate
Accelerometer	$\pm$2 g	61 μg/LSB	60 Hz
Gyroscope	$\pm$245 dps	8.75 mdps/digit	60 Hz

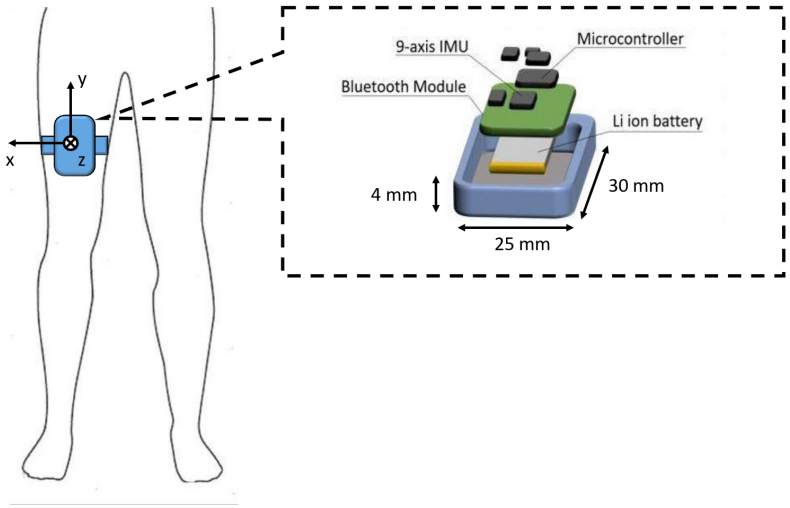


Fig. 3.13 Sketch of sensor position, together with an exploded view of the neMEMSI device.

Preprocessing

In this Section, the signal processing steps performed prior to the statistical analysis and the regression task are described. First, a sensor fusion process was performed to compute the orientation signal from the raw accelerometer and raw gyroscope readings. Then, the orientation signal was used to detect walking bouts from the entire TUG recording. Finally, inertial data were segmented, and temporal and spectral features were extracted from each stride.

Orientation estimation. A Kalman filter [273] was used to estimate the sensor orientation from the fusion of raw acceleration and angular velocity records. The sensor fusion algorithm iteratively alternates between two processes, including a prediction step and a correction step. The first consists of an approximation of the orientation estimate, performed by integrating the gyroscope readings; the second exploits the accelerometer readings to correct for the drift due to the integration of the slow-changing bias that affects the gyroscope measurements [274]. Figure 3.14 shows the raw gyroscope (a) and accelerometer (b) readings and the orientation estimate (d) obtained with the Kalman filter (c).

Filtering. After orientation estimation, the acceleration, angular velocity, and orientation signals were filtered to retain only the frequency components of inter-

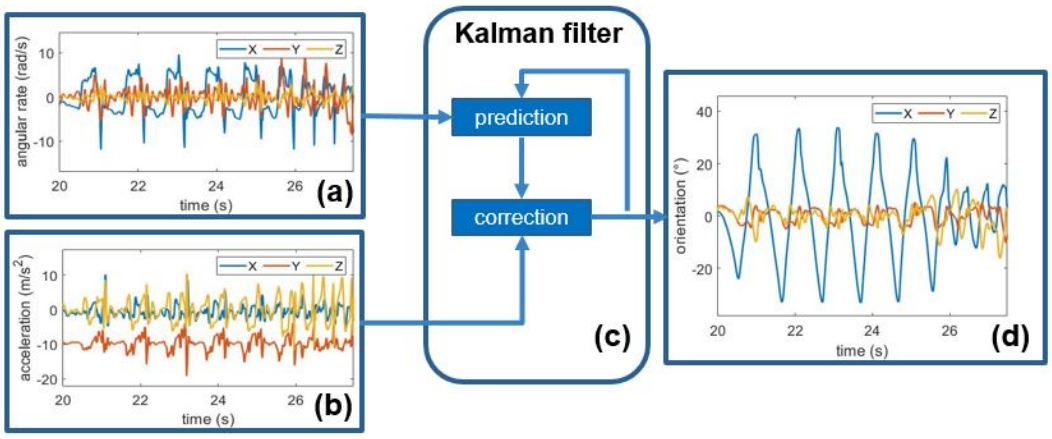


Fig. 3.14 Schematic of the Kalman filter, together with input and output. Raw gyroscope (**a**) and accelerometer (**b**) readings are input to the Kalman filter (**c**) to provide an estimate of orientation (**d**).

est, eliminating average values, low-frequency trends, and high-frequency noise. A second-order zero-lag Butterworth bandpass filter was used to retain only the components in the 0.5-20 Hz band, avoiding phase distortion.

Detection of walking bouts. In order to select only the walking data segments, an approach based on CWT, which is often used for walking step detection algorithms [196, 184], was implemented. A *Morse* mother wavelet was used in this study because of its similarity to the pattern of the orientation signal during walking. In addition, the scaling parameter a was set to perform the frequency analysis in the range of 0.5-2 Hz. This was done considering that stride time is quite heterogeneous in PwPD, due to the variability of motor characteristics among patients [235], pharmacological condition [145], and gait speed [257]. In [275], stride time in PD was found to be 1.13 *pm* 0.21 s, taking into account eleven studies on parkinsonian gait. The scalogram obtained from CWT is shown in Figure 3.15 (restricted in the frequency range 0–1 Hz), where the yellow areas correspond to the signal walking segments. To identify the walking intervals, the intensity profile was calculated for each value of the frequency scale; then, the obtained profiles were averaged; finally, the regions where the average intensity profile exceeded the standard deviation value were selected. The result of this procedure is shown in Figure 3.16, where the walking intervals in the orientation signal are identified.

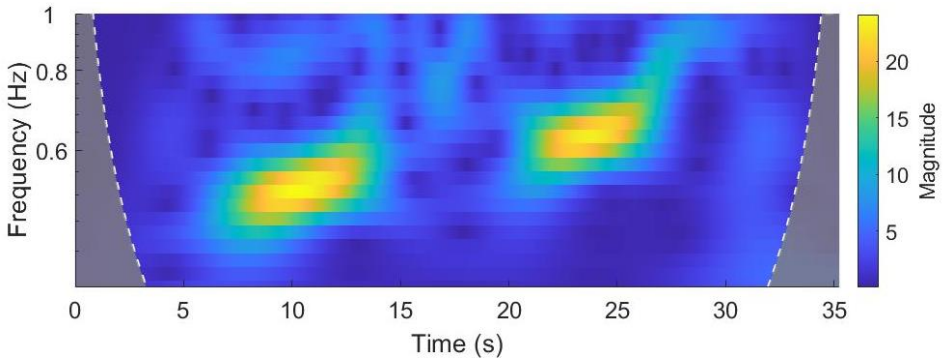


Fig. 3.15 Absolute value of the CWT plotted as a function of time and frequency. Yellow zones represent the walking segments of the signal.

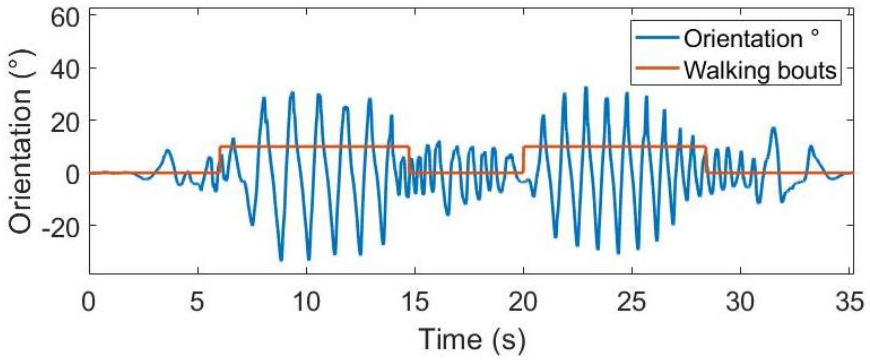


Fig. 3.16 Orientation signal (blue) and walking bouts detected by the algorithm (orange).

Detection of gait events. In each walking segment of the orientation signal, the initial contacts (ICs) were identified as the positive peaks of the orientation signal [276]. To avoid the detection of any double peaks, the orientation signal was filtered with a second-order zero-lag Butterworth low-pass filter, with a cutoff frequency of 2 Hz. In addition, only peaks greater than the standard deviation of the signal and at least 0.5 s apart were selected. As suggested in [276], the final contacts (FCs) correspond to the negative peaks following the ICs. The recordings of acceleration, angular velocity, and orientation were segmented into windows corresponding to the strides (i.e., from one IC to the next IC) in order to prepare the data for the subsequent feature extraction step.

Feature extraction. From each stride, a total number of 102 features were extracted from the acceleration, angular velocity, and orientation signal. Features include spatial-temporal gait parameters, and both time- and frequency- domain features. For each stride i, stride time, stance time, and swing time were computed

as follows:

$$T_{stride} = IC_{i+1} - IC_i \qquad T_{stance} = FC_i - IC_i \qquad T_{swing} = IC_{i+1} - FC_i$$

Tables 3.15 and 3.16 list features extracted from the time and frequency domains, respectively. The listed features describe different aspects of gait motion. For example, $Range, Std$, and RMS relate to the amplitude and intensity of the motion; E_{tot} and $binEnergy$ measure the energy content of the signal; $Entropy$ and $sEntropy$ describe the signal complexity; $DHwidth$ and $DHratio$ relate to the movement regularity. Spectral characteristics were calculated from the FFT of the signal. In order to have homogeneous spectral representations of all stride durations of all patients, the number of points at which to represent the FFT was set to $n = \bar{T}_{stride} \cdot F_s$, where $\bar{T}_{stride}$ is the mean stride time found in the PwPD [275] and F_s is the sampling rate. For stride durations longer than $barT_{stride}$, there is a small loss of spectral resolution, while for stride durations shorter than $barT_{stride}$, a few points, obtained as a linear interpolation of the actual data points, are added to the FFT. In any case, a spectral resolution of at least 1 Hz is expected, which is adequate for calculating the features listed in table 3.16.

Feature selection. The Pearson correlation coefficient r between the extracted features and the PIGD score was calculated in patients both OFF and ON treatment. To reduce the dimensionality of the entire feature set (i.e., 102 features), the least significant features (i.e., those with $r < 0.4$) were discarded. To further reduce the dimensionality of the set, features were ranked according to their predictive ability. This was done by exploiting two different approaches and evaluating their effect on the final prediction ability. The first approach consisted of sorting the features in descending order of r and selecting the first N features. The second approach used PCA to reduce the dimensionality of the feature set, keeping only the first N principal components. The N parameter was adjusted in a range from 5 to 25 in order to estimate the effect of different dimensionality of the feature set on model performance.

Table 3.15 List of time-domain features extracted in the present study, together with equations and some brief explanations. α: acceleration; ω: angular velocity; θ: orientation.

Feature	Component	Equation	Explanation
Min	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	minimum value
Max	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	maximum value
Mean	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$	average value
Std	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$	standard deviation
RMS	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$x_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	root mean square value
Range	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$r_x = x_{max} - x_{min}$	range of values
Entropy	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$E_x = x\log(x+\varepsilon), \varepsilon = 10^{-5}$	Shannon signal entropy
nPeaks	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	number of peaks higher than Std
hPeaks	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	average height of nPeaks
vPeaks	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	standard deviation of hPeaks
Zc	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	zero-crossing rate
Corr	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$r(i,j) = \frac{cov(i,j)}{\sigma(i)\sigma(j)}$	correlation between pair of axes

Regression

This section describes the statistical processing following the extraction of the entire feature set for each patient stride, in order to investigate the clinical significance of the extracted features. First, a correlation analysis was performed between the engineered features extracted from the strides and the clinical scores; this was done by calculating the Pearson correlation coefficient and the corresponding p-value for each feature-clinical score pair. Then, a regression model was implemented to predict the PIGD score of PwPD. The analysis was performed on patients in both the OFF and ON states of therapy to evaluate the effect of drug treatment on the performance of the prediction model. Finally, to also evaluate the effect of FOG on model performance, patients were divided according to the clinical presence of FOG. Figure 3.17 schematically reports the entire process, including the test populations and dimensionality reduction methods.

Feature scaling was applied to each feature using z-score normalization, which consists in removing the mean value and dividing by the standard deviation. This

Table 3.16 List of spectral-domain features extracted in the present study, together with equations and some brief explanations. α: acceleration; ω: angular velocity; θ: orientation.

Feature	Component	Equation	Explanation
DH frequency	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	frequency of the principal harmonic
DH height	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	amplitude of the principal harmonic
DH width	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	width of the principal harmonic
E_{tot}	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$\sum_{f=1}^{Fs/2} X_f$	total signal energy
DH ratio	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	ratio between the energy of the principal harmonic and E_{tot}
sEntropy	$\alpha_y, \alpha_z, \omega_x, \theta_x$	$X_f \log(X_f + \varepsilon)$	Shannon entropy of the signal FFT
binEnergy	$\alpha_y, \alpha_z, \omega_x, \theta_x$	-	ratio between energy in specific frequency bands and E_{tot}

was done to uniform the feature range while reducing the effect of possible outliers. Then, range normalization was performed both on the feature set and on the target vector (i.e., PIGD score) to rescale data in the range [0, 1].

Regression. Regarding the regression model, a support vector regression (SVR) model was implemented [277, 278]. In order to provide a robust performance evaluation, the model was tested using the LOSO CV, which resembles the realistic working condition of the model. To optimize the model parameters, a LOSO-based training-validation procedure was performed, selecting the parameters that provided the best performance on the validation set. The kernel function, kernel scale, and misclassification cost (box-constraint) parameters were optimized for each SVR model, while the tolerance margin (ε parameter) was set to the default value corresponding to one-tenth of the standard deviation of the PIGD score. The goodness of fit was evaluated using r, $RMSE$, and MAE.

Effect of therapy. The inertial data of PwPD were divided according to the pharmacological condition. Two independent data sets were obtained from patients on OFF and ON therapy. The motor conditions of the OFF and ON patients were compared by performing the Wilcoxon test on the MDS-UPDRS part III and the PIGD score in the two pharmacological conditions. The analysis shown in Figure 3.17 was then performed to optimize the model. The performance obtained on

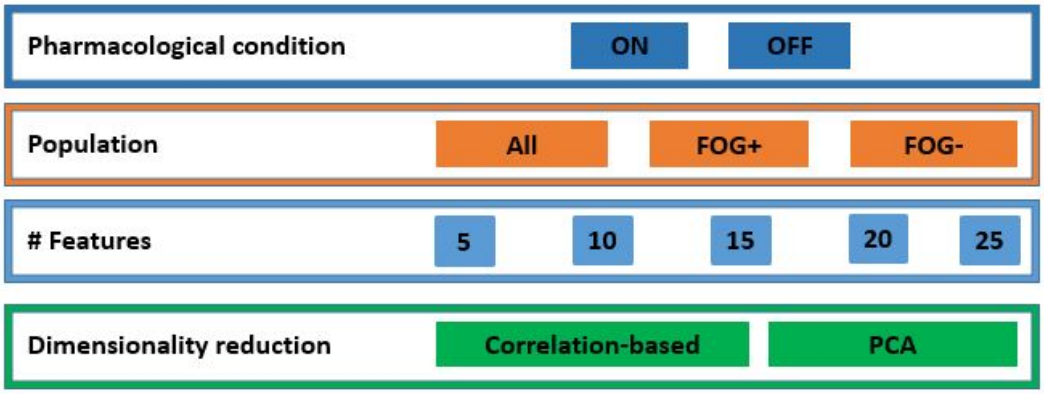


Fig. 3.17 Schematic representation of the analysis performed. The processing was performed for both pharmacological conditions, different populations, different sizes of the feature set, and different dimensionality reduction methods. ON: under dopaminergic therapy; OFF: not under dopaminergic therapy; FOG: patients with Parkinson's disease and freezing of gait; FOG-: patients with Parkinson's disease without freezing of gait; PCA: principal component analysis.

OFF and ON patients was compared using different feature set sizes, different dimensionality reduction methods, and optimizing regression model parameters. Finally, the performance of the model in OFF and ON patients was compared.

Effect of FOG. The dataset was divided according to the clinical presence of FOG. Then, the Mann–Whitney U test was used to compare both clinical scores and engineering characteristics of FOG+ and FOG- patients. The analysis shown in Figure 3.17 was then performed, optimizing the model. The performance obtained on patients with and without FOG was compared using different feature set sizes, different dimensionality reduction methods, and optimizing the parameters of the regression model. The entire procedure was performed for each pharmacological condition. Finally, the effect of FOG on model performance was evaluated.

All the experiments were executed in Matlab R2020a, using a personal computer with Microsoft Windows 10, a 2.4 GHz Intel® Core Processor i5-6200, 8 GB RAM, and 4 GB GPU.

3.3.2 Prediction of self-perceived gait impairment

The objective of this work [279] was to establish the usefulness of motion data taken using a single inertial sensor for the prediction of axial impairment as self-reported

by the patients themselves during daily activities. These latter are related to the sum of MDS-UPDRS items 2.11-2.13 (in the following reported as axial impairment measure - AIM), taken as the reference metric to evaluate the performance of the proposed algorithm. Specifically, part II of the UPDRS is a questionnaire related to motor aspects of experiences of daily living. Items 2.11-2.13 refer to the following aspects: getting out of bed or a deep chair, walking and balance, and FOG. Data and preprocessing procedures are those used in Section 3.3.1. Starting from the features extracted in the previous study, an RF regression model was optimized to predict the AIM.

Preprocessing

The initial set of features was divided into a training set and a test set, based on the LOSO test. At each iteration, the following processing steps were performed. First, the training and test sets were normalized using range normalization (Equation 3.3). Then, the features X of the training and test sets were normalized according to the minimum and maximum value of these features in the training set.

$$X' = \frac{X - min(X_{train})}{max(X_{train}) - min(X_{train})} \tag{3.3}$$

Pearson's correlation coefficient r of features in the training set was used to discard redundant and non-significant features from both the training and test sets. Specifically, features that were highly correlated with other features ($r > 0.9$) and/or poorly correlated with the target ($r < 0.4$) were discarded. The F-test was used to rank the features in descending order of importance. The number of selected features n_f was adjusted in the range 1-N_f, where N_f is the total number of features.

Regression

In this work, an RF regression model was implemented. It is a supervised ML algorithm that averages the predictions of multiple decision trees to compute the final output. The number of learners n_l is an internal parameter to be optimized. Other parameters to be adjusted are the minimum leaf size m_{ls}, the maximum number of subdivisions m_{ns}, and the maximum parent size m_{ps}. During the optimization procedure, n_l was adjusted in the range 3–40, m_{ls} in the range 1–30, m_{ns} in the

range 1–20 and m_{ps} in the range 1–20. A k-fold CV was implemented. Data from a given PwPD were included in either the training or validation set to ensure patient independence. In this study, k was set to 10, resulting in data from 3 PwPD in each validation fold. A grid search approach was used to optimize the number of selected features and model parameters. For each combination of n_f, n_l, m_{ls}, $m_n s$, and $m_p s$, a 10-fold CV was performed and the RMSE was calculated using the predictions of the validation set. The optimal combination of parameters was identified by searching for the minimum value of RMSE. The performance of the model was evaluated using LOSO validation. First, since several observations (corresponding to different strides) of each PwPD were available, they were averaged to obtain a single measure for each PwPD. The results were reported in terms of r, RMSE, and MAE. In addition, the correlation plot was obtained from the PwPD OFF and ON treatment.

3.4 Bradykinesia

The aim of this work was to develop an automatic scoring system for the evaluation of lower limb bradykinesia. Data from a large cohort of PwPD were recorded using a single smartphone on the thigh, while patients performed the MDS-UPDRS leg agility (LA) task (item 3.8). The inter-rater variability was evaluated and the average raters' score was used to train and validate the computer scoring system, providing an objective and unbiased evaluation of bradykinesia.

Subjects

A total of 93 PwPD were recruited in the study. The inclusion criteria were a clinical diagnosis of PD with motor signs and symptoms, no major cognitive impairment or other conditions that prevented the patient from performing the task correctly. Because data acquisition was done during scheduled outpatient visits, most patients were in the *daily on* condition. In some special cases (scheduled late morning visits - about 4% of patients), some of them showed an end-dose effect. However, the number of these patients was not sufficient to perform a differential analysis, so they were not differentiated by motor condition. This is not likely to affect system

performance in any way, given the small number of subjects. The characteristics of the population are summarized in Table 3.17.

Table 3.17 Demographic and clinical characteristics of PD population. Measures are reported in terms of mean ± standard deviation.

# patients (male)	Age (years)	Disease duration (years)	H&Y	LA score
93 (65)	69 ± 10	9.0 ± 6.5	2.5 ± 0.8	2 ± 1

Data acquisition

During the examination, subjects were asked to sit in a straight-backed chair and place their feet on the floor in a comfortable position. Then, after being properly instructed by experienced neurologists as recommended by the MDS-UPDRS guidelines, they performed the LA with each leg separately. A simple Velcro armband equipped with a smartphone was placed around the patient's thigh, with the y-axis parallel to the direction of the femur. The smartphone recording application was started before and stopped after the task was performed, so each recording included a single LA performance. Experiments were videotaped to allow for multi-rater evaluation. Figure 3.18 shows the experimental setup. A total of 184 LA tests were recorded (2 patients were able to perform the test with only one leg).

The LA task was rated by four experienced neurologists according to the MDS-UPDRS scale, either directly or after inspection of the video sequences. The rounded mean ratings were used as class labels for the supervised classification algorithms. Figure 3.19 shows the distribution of assigned MDS-UPDRS scores. It is worth noting that the data set includes a few cases belonging to MDS-UPDRS-3 and 4 classes. In fact, the clinical conditions of patients belonging to these classes are severe and may even prevent them from performing the task. In particular, although five MDS-UPDRS-4 patients were tested, the usefulness of including them in a monitoring system is questionable. Therefore, as also suggested by medical experts, patients with MDS-UPDRS-4 were excluded from further analysis. Acceleration, angular velocity, and orientation data were collected via the SensorLog application, stored locally on an SD card, exported in CSV format, and processed offline with `MATLAB`, version 2018a for Windows 10.

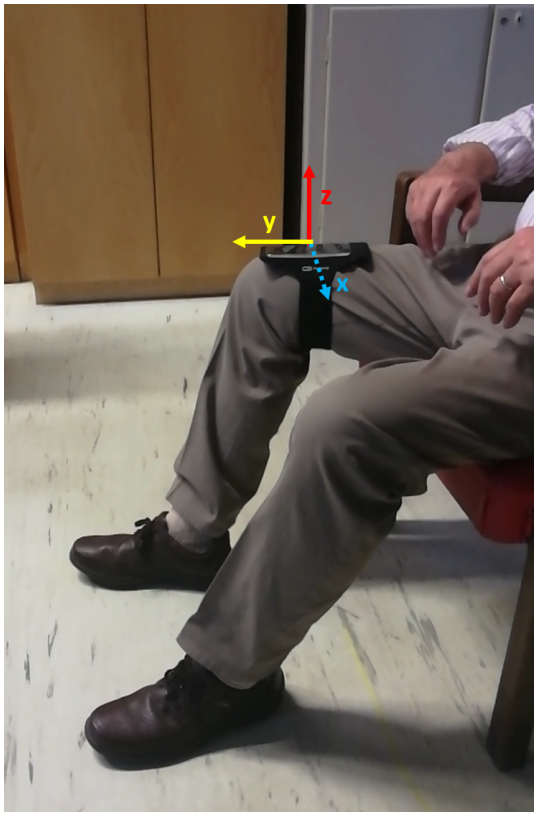


Fig. 3.18 Smartphone position adopted for the LA task scoring.

Preprocessing

The inertial signals were recalibrated to compensate for deviations from ideal positioning, i.e., gravity acting only on the vertical component (the z-axis of the accelerometer). The method proposed in [254], which consists of a 3-axis accelerometer orientation correction by applying a quaternion rotation transformation to the raw device data, was applied.

Filtering. After the removal of the mean value, the pitch, acceleration, and angular velocity data were filtered with a low-pass filter to remove high-frequency noise. A second-order zero-lag Butterworth low-pass filter with a cutoff frequency of 4 Hz was chosen to maintain at least 90% of the signal power (calculated using the FFT on all data).

Feature extraction. Then, a set of 36 kinematic features (shown in Table 3.18) was extracted from each signal, representative of the main features that distinguish movement in PwPD and unaffected controls. Indications from the literature on other similar studies were taken into account [185, 280–283]. In addition to cross-

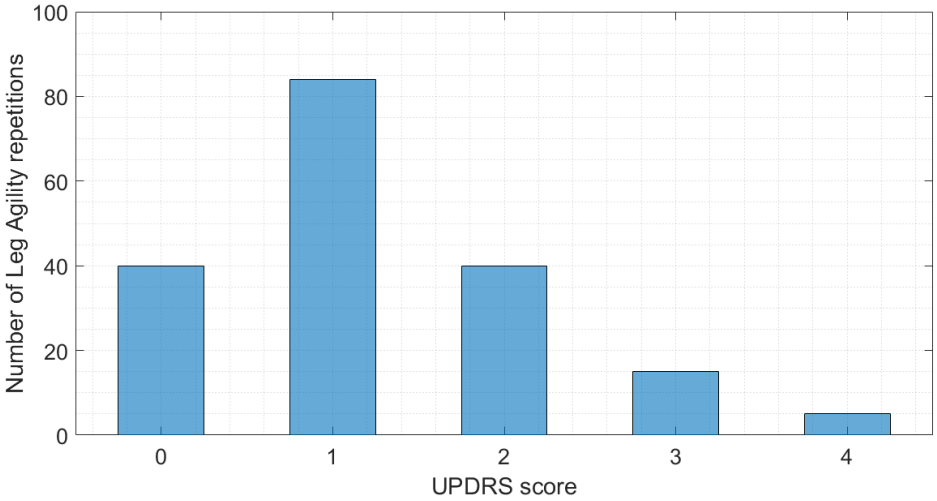


Fig. 3.19 Distribution of the MDS-UPDRS scores assigned to the LA tasks. 0: normal. 1: slight. 2: mild. 3: moderate. 4: severe.

correlation, which provides information in the spectral domain, the FFT of the signal was calculated and used to extract features such as the frequency, amplitude, and width of the dominant harmonic, the total number of harmonics, and the power ratio between the main harmonics and other harmonics. Some features were then further combined to increase their discriminatory power (for example, the average peak value feature takes into account the number and width of harmonics in the FFT).

Feature selection. To identify the most significant features, a feature selection procedure was performed, based on the correlation between feature values and the target. Figure 3.20 shows the Pearson correlation coefficient for each feature.

The optimal subset of features, i.e., the one containing the most informative features while maintaining a small size was obtained as follows. First, the features were sorted in descending order according to Pearson's correlation coefficient, and those with a correlation of less than 0.4 (i.e., weak correlation) were discarded. Then, three thresholds were set, shown in Figure 3.20 as *C1, C2* and *C3*, corresponding to a significant difference in the correlation value between adjacent features. The three resulting feature subgroups were used as inputs for common ML algorithms, including SVM, kNN, DT, and a new ANN approach. The final subgroup was the one that provided the best results in terms of accuracy. The optimal value turned out to be *C2*, which yielded 16 selected features (Table 3.18). A brief description

Table 3.18 List of features extracted, along with the selected components. θ_x: pitch signal around the x-axis. ω_x: Angular velocity around the x-axis. α_z: acceleration along the vertical direction. - indicates that none of the components were selected.

Selected component	Extracted features
$\theta_x, \omega_x, \alpha_z$	Dominant frequency
-	Entropy
-	Minimum
ω_x, α_z	Maximum
ω_x, α_z	Root Mean Square
ω_x, α_z	Range
ω_x, α_z	Spectral Entropy
ω_x, α_z	Mean amplitude
ω_x, α_z	Regularity
α_z	Dominant Ratio
-	Standard deviation
-	Mean peak value

of some features is given below. The dominant frequency is the frequency value corresponding to the highest peak of the FFT function. Spectral entropy is the Shannon entropy calculated on the FFT of the signal. Regularity is the amplitude of the first positive peak of the autocorrelation function, normalized to the maximum of the correlation function, as described in [284]. The dominant ratio is the ratio of the power of the dominant frequency band to the total power.

Classification

Common ML algorithms were used for comparison with other studies (e.g., [185, 241, 186]). Moreover, a novel ANN approach was implemented, providing a continuous output.

Optimization. To identify the best combination of ML model parameters, a Bayesian optimization algorithm was run. The CV error was used as the objective function to be minimized in a LOSO validation. The following hyperparameters were optimized: kernel function, kernel scale, and box constraint (cost) for one-vs-one SVM; number of neighbors, distance metric, and distance weight for kNN; maximum number of splits and split criterion for DT. As for ANN, stop conditions were set at a maximum of 2000 iterations, gradient value at 10^{-5}, initial learning rate

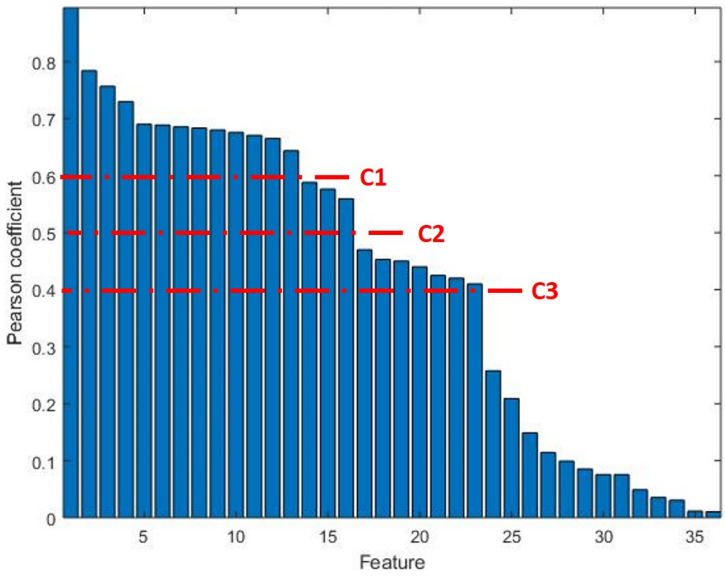


Fig. 3.20 Feature ranking based on Pearson's correlation coefficient r. *C1, C2, C3* identify significant differences in r-values of adjacent features.

at 0.01, increasing and decreasing values at 10 and 20%. Then, the best architecture was investigated. In particular, the number of hidden layers, the number of hidden neurons per layer, and the transfer function were tuned. Finally, the ANN architecture that provided the lowest classification error was selected.

Evaluation. The performance of ML algorithms was evaluated in terms of accuracy and AUC in LOSO validation. In this case, the target class was obtained by rounding the mean of the clinical scores. To better understand the behavior of the proposed classifiers, the cumulative distribution function (CDF) was calculated as a function of the absolute classification error, i.e., the absolute difference between the classifier algorithm and the rounded mean MDS-UPDRS class provided by neurologists. In addition, the continuous classifier result was compared with the mean clinical score using Pearson's correlation coefficient and RMSE. Finally, inter-rater variability in LA assessment was visually assessed by pie charts and computationally estimated by calculating the intra-class correlation coefficient. In addition, the Bland-Altman plot was generated to evaluate the significant differences between the clinical score and the output of the developed classification model.

Chapter 4

Results

4.1 Freezing of gait

4.1.1 Prediction of freezing of gait using lower limbs inertial sensors

Offline clinical evaluation of video recordings by two independent neurologists with expertise in movement disorders identified 41 episodes of FOG in on-treatment PwPD and 54 episodes of FOG in off-treatment PwPD. All episodes were used for the FOG detection task. On the other hand, 6 and 10 episodes were excluded from the pre-FOG analysis, for patients ON and OFF therapy, respectively, because they occurred during the gait initiation task, that is, during the transition between getting up and starting to walk.

FOG Detection

Table 4.1 summarizes the algorithm performance, in terms of sensitivity, specificity, accuracy, PPV, NPV, and F-score, in detecting FOG episodes in PwPD both ON and OFF therapy.

Table 4.1 Algorithm performance in FOG detection in patients under (ON) and not under (OFF) dopaminergic therapy. cv: cross-validation; NPV: negative predictive value; PPV: positive predictive value; LOSO: leave-one-subject-out.

Evaluation metric	Condition	10-fold CV	70-30 training-test	LOSO
Sensitivity	ON	0.959	0.939	0.937
	OFF	0.971	0.949	0.939
Specificity	ON	0.954	0.942	0.918
	OFF	0.935	0.906	0.850
Accuracy	ON	0.955	0.941	0.926
	OFF	0.963	0.931	0.920
PPV	ON	0.953	0.939	0.917
	OFF	0.942	0.935	0.868
NPV	ON	0.962	0.941	0.938
	OFF	0.957	0.927	0.914
F-score	ON	0.956	0.939	0.927
	OFF	0.956	0.942	0.902

All performance metrics showed a slight decrease from 10-fold CV to 70/30 training/test and LOSO validation, due to a progressively larger portion of the data used as the test set. It is worth noting that LOSO validation provides an estimate of the algorithm's performance under real working conditions, i.e., when testing the model on a new unknown subject. The classification algorithm achieved high sensitivity, accuracy, and F-score (Table 4.1), always above 90%. In addition, the training/test procedure demonstrated the absence of model over-fitting, as evident from the high performance obtained. Comparing patients ON and OFF therapy, the algorithm achieved similar values of sensitivity, accuracy, and F-score in FOG detection with LOSO, thus showing no significant dependence of performance on L-Dopa intake. However, the algorithm showed lower specificity in patients not on therapy compared with those on therapy, due to the increased number of FPs after discontinuation of L-Dopa.

Table 4.2 reports the performance of the algorithm in detecting FOG after training with data recorded from patients ON therapy and tested on patients OFF therapy

and vice versa. The performance is satisfactory in both cases, with accuracy and F-score above 89%. It can be observed that, in general, and especially with regard to sensitivity, the algorithm provided higher performance in FOG detection when trained with data from patients OFF therapy.

Table 4.2 Algorithm performance in FOG detection after training with patients under dopaminergic therapy (ON) and testing on patients not under dopaminergic therapy (OFF), and vice versa. NPV: negative predictive value; PPV: positive predictive value.

Train - test set	Sensitivity	Specificity	Accuracy	PPV	NPV	F-score
ON - OFF	0.880	0.903	0.890	0.919	0.858	0.899
OFF - ON	0.962	0.890	0.926	0.894	0.961	0.927

Pre-FOG Detection

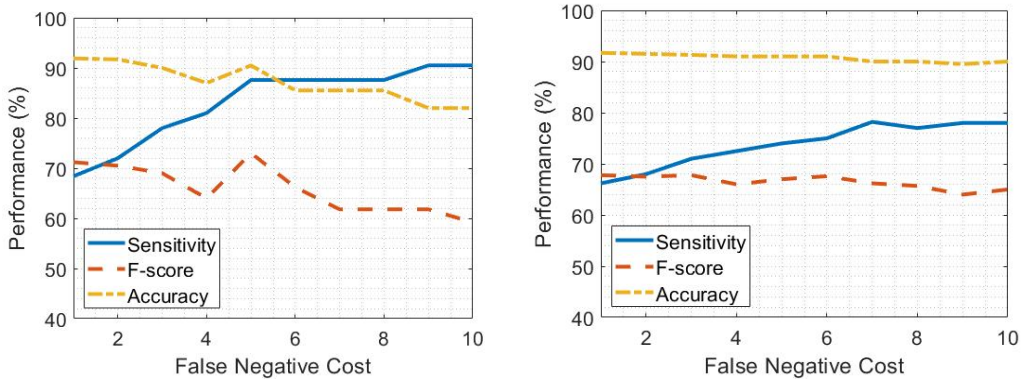
Table 4.3 summarizes the accuracy of different ML classifiers in identifying pre-FOG periods in patients ON and OFF therapy, also considering different sizes of the pre-FOG window (from 2 to 5 s).

Table 4.3 Accuracy of different classifiers in pre-FOG recognition by considering various pre-FOG window lengths. kNN: k-nearest neighbor; LDA: linear discriminant analysis; LR: linear regression; SVM: support vector machine.

Window length (s)	SVM		kNN		LDA		LR	
	ON	OFF	ON	OFF	ON	OFF	ON	OFF
2	0.913	0.921	0.847	0.898	0.917	0.947	0.890	0.906
3	0.861	0.887	0.802	0.847	0.856	0.864	0.844	0.852
4	0.778	0.846	0.694	0.804	0.786	0.826	0.752	0.818
5	0.649	0.746	0.589	0.793	0.658	0.754	0.449	0.711

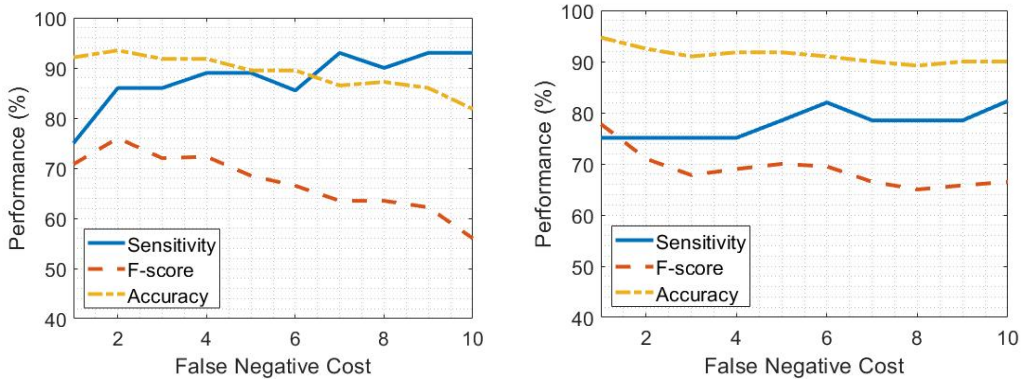
All models showed a progressive reduction in accuracy in pre-FOG recognition with increasing window length. In fact, as the length of the pre-FOG window doubled, an average reduction in accuracy of about 14% and 10% was observed for patients ON and OFF therapy, respectively. Overall, the accuracy in detecting pre-FOG in patients who are not on therapy was higher than that of patients on therapy. In addition, the SVM and LDA classifiers performed best in terms of accuracy, with a sensitivity of 0.684 and 0.662, respectively.

Figures 4.1 and 4.2 report the sensitivity, accuracy, and F-score of SVM and LDA classifiers when detecting pre-FOG periods in patients ON and OFF therapy, respectively, and considering different FN cost values.



(a) SVM (b) LDA

Fig. 4.1 False negative tuning in PD patients under dopaminergic therapy (ON) for support vector machine and linear discriminant analysis classifiers.



(a) SVM (b) LDA

Fig. 4.2 False negative tuning in PD patients not under dopaminergic therapy (OFF) for support vector machine and linear discriminant analysis classifiers.

As FN cost increases, sensitivity improves but, in turn, accuracy and F-score decrease due to the presence of FPs. This has an impact on all performance evaluation metrics other than sensitivity. Figures 4.1a and 4.2a show that considering the SVM classifier and setting an FN cost equal to 5 leads to a sensitivity of 0.875 and 0.892 in patients ON and OFF therapy, respectively while maintaining high accuracy and F-score values. In contrast, the increase in sensitivity using the LDA classifier is less satisfactory than SVM, as shown in Figure 4.1b and Figure 4.2b.

Table 4.4 reports the sensitivity, accuracy, and F-score in pre-FOG detection using SVM and LDA classifiers separately, with and without FN cost optimization, as well as the combination of SVM and LDA classifiers in PwPD ON and OFF therapies.

Table 4.4 Performance of support vector machine (SVM) and linear discriminant analysis (LDA) classifiers, separately and in combination, with and without optimized false negative cost, in pre-FOG detection in PD patients under (ON) and not under (OFF) dopaminergic therapy

Model	Sensitivity		Accuracy		F-score	
	ON	OFF	ON	OFF	ON	OFF
SVM	0.684	0.750	0.918	0.921	0.712	0.719
LDA	0.662	0.751	0.917	0.947	0.678	0.778
SVM + LDA	0.721	0.792	0.917	0.929	0.667	0.757
SVM (optimized cost)	0.875	0.892	0.902	0.894	0.723	0.681
LDA (optimized cost)	0.782	0.822	0.901	0.910	0.660	0.698

The combination of SVM and LDA classifiers resulted in higher sensitivity in detecting pre-FOG in both ON and OFF patients compared with the separate performance of SVM and LDA classifiers. The FN cost of 7 and 6 for the LDA classifier (Figure 4.1b and Figure 4.2b) showed satisfactory performance in patients in ON and OFF therapy, respectively. The SVM classifier with an FN cost of 5 achieved the highest performance in pre-FOG detection, especially in terms of sensitivity, with comparable accuracy values in both ON and OFF therapy patients (Figure 4.1a and Figure 4.2a).

Table 4.5 summarizes the sensitivity, specificity, accuracy, PPV, NPV, and F-score of the pre-FOG classification algorithm in a LOSO validation in patients on ON and OFF therapy.

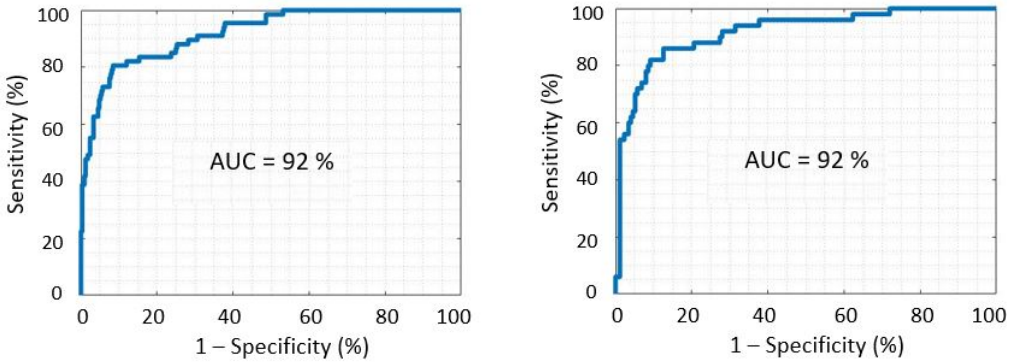
Table 4.5 Performance of the pre-FOG classification algorithm in the leave-one-subject-out validation in patients under (ON) and not under (OFF) dopaminergic therapy.

Therapy	Sensitivity	Specificity	Accuracy	PPV	NPV	F-score
ON	0.841	0.859	0.855	0.651	0.935	0.734
OFF	0.85.5	0.863	0.861	0.662	0.930	0.746

Comparing patients ON and OFF therapy, the classification algorithm demonstrated different performances in recognizing pre-FOG periods, with the highest

values of sensitivity, specificity, accuracy, and F-score after L-Dopa withdrawal (i.e., in OFF patients compared with those on ON therapy). In addition, the implemented models detected pre-FOG periods with different latencies in patients on OFF and ON therapy. More specifically, when considering patients ON therapy, pre-FOG periods were recognized 4 *pm* 1.1 steps before the onset of FOG. In contrast, pre-FOG periods were recognized 6 *pm* 1.3 steps before the onset of FOG in patients OFF therapy. This is probably because the pattern of rhythm degradation in patients OFF therapy is more evident than in those ON therapy, where it is (partially) corrected by L-Dopa.

Figure 4.3a and 4.3b report the ROC curve of the SVM classifier, in patients ON and OFF therapy, respectively.



(a) ON therapy (b) OFF therapy

Fig. 4.3 Receiver operating characteristic curves of the final classification model, for patients under (ON) and not under (OFF) dopaminergic therapy.

ROC curves show a similar pattern in both conditions and the AUC value is identical. For specificity values over 0.8, slightly higher values of sensitivity can be observed for patients OFF therapy, compared to those ON therapy.

Table 4.6 reports the algorithm performance in pre-FOG detection after training with data recorded from patients ON therapy and then tested on data from those OFF therapy, and vice versa.

As is evident, the different tests on patients ON and OFF therapy led to opposite results in terms of both sensitivity and specificity in detecting pre-FOG periods. In more detail, the algorithm trained with data from patients ON therapy and tested on data from patients OFF therapy showed significantly higher sensitivity and lower specificity than the algorithm trained with data from patients OFF therapy and tested

Table 4.6 Algorithm performance in pre-FOG detection after training with Parkinson's disease patients under dopaminergic therapy (ON) and testing on PwPD not under dopaminergic therapy (OFF), and vice versa.

Train- test set	Sensitivity	Specificity	Accuracy	PPV	NPV	F-score
ON - OFF	0.840	0.883	0.874	0.667	0.952	0.744
OFF - ON	0.566	0.925	0.863	0.779	0.882	0.656

on data from patients ON therapy. In fact, in the latter case, sensitivity was severely impaired and reached values below 0.6.

Finally, Figure 4.4 shows the Spearman correlation coefficient, calculated for the most frequently selected features in patients ON and OFF therapy during the pre-FOG periods. All corresponding p-values were <0.001.

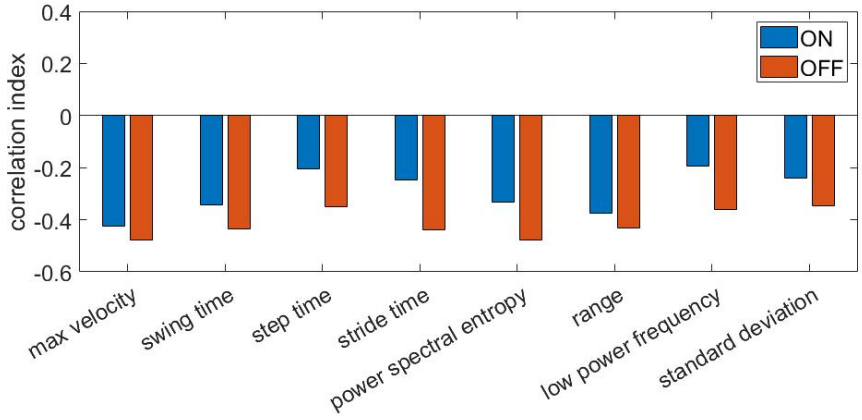


Fig. 4.4 Spearman correlation coefficient between selected features and class label (i.e., 0 and 1 for gait and pre-FOG respectively). A negative correlation denotes decreasing values of features during pre-FOG. ON: under dopaminergic therapy; OFF: not under dopaminergic therapy

4.1.2 Real-time detection of freezing of gait using a single accelerometer

CNN architecture and parameters

The optimization process aimed to find the best CNN architecture, model parameters, and learning settings led to the following results.

A batch size of 256, a learning rate of 0.001, and a maximum number of iterations of 120 epochs were selected for training, optimization, and testing procedures. The final CNN architecture is schematically reported in Figure 4.5. It consists of a three-head CNN block connected to classification (dense) layers. Each head is composed

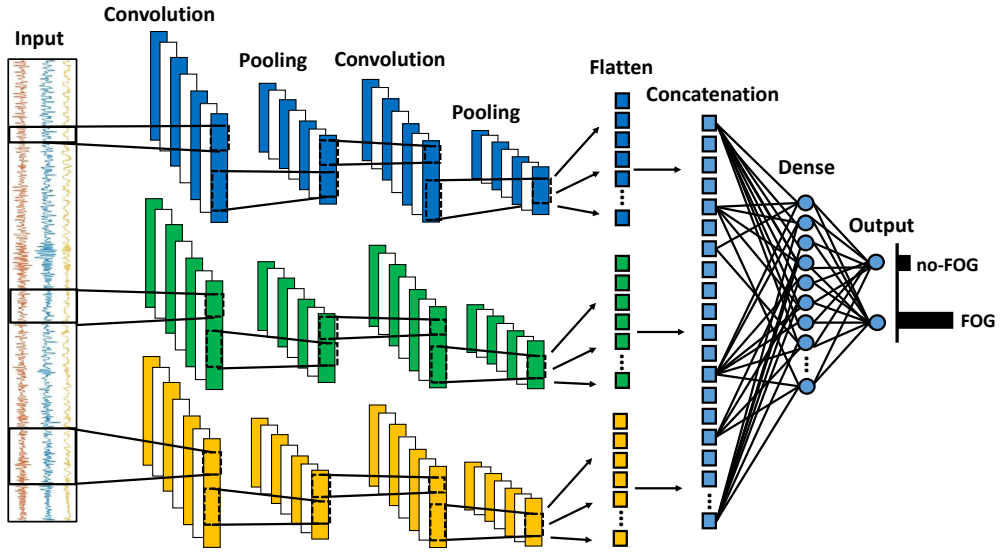


Fig. 4.5 Architecture of the optimized multi-head convolutional neural network model.

of two convolutional and two max-pooling layers. Each of these heads processes the input (80 time-steps * 3 channels) simultaneously using a different kernel size. The outputs of the CNN heads are flattened and concatenated to compose a vector that feeds a single dense layer (16 units and a dropout rate of 0.5), and a final output layer with two outputs corresponding to the probability of FOG or no-FOG, respectively. Convolutional layers have 16 filters and ReLU activations each, with different kernel sizes in each head of the CNN. Specifically, kernel sizes of 6 and 3 were selected in the two convolutional layers of the first head, 12 and 6 in the second head, and 18 and 9 in the third head. The stride length was set to 1 in all the convolution layers with no padding to gradually reduce the size of the input signal. In addition, a pool size of 3 and a dropout rate of 0.5 was used for all convolutional layers, while the regularization term was set to 0.1 in the softmax layer.

Table 4.7 Multi-head convolutional neural network layers, features, and parameters. n: number of filters; k: kernel size; d: dropout rate; p: pool size.

	Layer	Features	Shape	# param.
	Input	-	(80,3)	0
Head 1	conv	n = 16, k = 6	(75,16)	304
	dropout	d = 0.5	(75,16)	0
	pool	p = 3, s = 3	(25,16)	0
	conv	n = 16, k = 3	(23,16)	784
	dropout	d = 0.5	(23,16)	0
	pool	p = 3, s = 3	(7,16)	0
	flatten	-	112	0
Head 2	conv	n = 16, k = 12	(69,16)	592
	dropout	d = 0.5	(69,16)	0
	pool	p = 3, s = 3	(23,16)	0
	conv	n = 16, k = 6	(18,16)	1552
	dropout	d = 0.5	(18,16)	0
	pool	p = 3, s = 3	(6,16)	0
	flatten	-	96	0
Head 3	conv	n = 16, k = 18	(63,16)	880
	dropout	d = 0.5	(63,16)	0
	pool	p = 3, s = 3	(21,16)	0
	conv	n = 16, k = 9	(13,16)	2320
	dropout	d = 0.5	(13,16)	0
	pool	p = 3, s = 3	(4,16)	0
	flatten	-	64	0
	merge	-	272	0
	dense	-	16	4368
	dropout	d = 0.5	16	0
	softmax	-	2	34

Model layers and parameters are reported in Table 4.7, together with the features and output shape of each layer. The implemented model includes a total number of

10834 trainable parameters, 6432 of which from convolutional layers and 4402 from the densely connected layers.

Performance

Training, validation and test sets from the REMPARK data set were arranged as in Table 4.8. As it can be observed, participants' age, symptoms duration, disease progression (measured by the H&Y score), and motor impairment (measured by the total MDS-UPDRS-III) are similar across the subsets. Participants included in the test set have a slightly larger FOG-Q score compared to those included in the training set, and the difference is more evident if compared to the validation set.

Table 4.9 reports the performance of the model on the training, validation, and test set of the main data set. Model performance is similar in the three sets, with a negligible impairment when moving from the training to the test set. This proves the high generalization capability of the classification algorithm, providing good results also on the test set, which includes data from completely unknown patients.

Table 4.8 Demographic and clinical features of PD patients include in training, validation, and test set. H&Y: Hoehn and Yahr score; MMSE: mini-mental state examination; FOG-Q: freezing of gait questionnaire; MDS-UPDRS: unified Parkinson's disease rating scale; ON: under dopaminergic therapy; OFF: not under dopaminergic therapy.

Set (# subjects)	Train (12)	Val (4)	Test (5)
Age (years)	69.5	66.5	72.2
Symptoms duration (years)	11.0	15.5	12.6
H&Y	3.1	2.8	3.2
MMSE	27.9	27.5	27.4
FOG-Q	15.1	11.8	18.6
MDS-UPDRS-III ON	18.5	13.0	16.0
MDS-UPDRS-III OFF	36.3	37.5	33.6

As far as concerns the true FOG episodes detected in the REMPARK data set, 100% of episodes were correctly identified by the algorithm, with an average percentage of 84.8% of FOG detected in each episode. More specifically, the average proportion of FOG detected in each episode was found to be 76.4%, 87.5%, and 90.2% in FOG episodes of duration 0-5s, 5-10s, and $>$10s respectively (Table 4.10). Similar results were obtained for the 6MWT data set when considering episodes of

Table 4.9 Performance of the implemented classification model on training, validation, and test set of the main data set. EER : equal error rate; AUC : area under the curve.

Set	Train	Validation	Test
Sensitivity	0.884	0.879	0.877
Specificity	0.885	0.880	0.883
Geometric mean	0.885	0.880	0.880
F-score	0.886	0.838	0.830
EER (%)	11.5	11.9	11.9
AUC	0.955	0.947	0.946

duration larger than 5s, while a reduction in the detection rate was observed for FOG episodes of duration smaller than 5s.

Table 4.10 Percentage of FOG episodes detected.

Data set	FOG episodes detection rate (%)		
	0-5s	5-10s	>10s
REMPARK	100	100	100
6MWT	87	100	100

FPs duration was found to be 2.7 ± 1.5 s, with 37.7%, 61.1%, and 83.0% of false episodes less than 5s, 10s, and 20s far from the closest real FOG, respectively. This suggests that false FOG episodes are relatively short and distributed close to real FOG events.

Concerning the time resolution in FOG detection, 52.3% of the FOG episodes were predicted before the actual onset, with an average advance of 3.1s (SD = 2.6s, min = 0.5s, max = 11s), while 47.7% of FOG episodes were detected with an average delay of 0.8s (SD = 0.6s, min = 0.5s, max = 3s).

Testing the classification algorithm on the external 6MWT data set yielded sensitivity 0.796, specificity 0.933, geometric-mean 0.862, accuracy 0.929, and AUC 0.953.

Figure 4.6 reports the ROC curve of the implemented classification model tested on the main (REMPARK) and the external data set (6MWT). ROC curves are similar in the two data sets, with slightly better performance obtained in the external data set. This may be due to the different compositions of the two data sets. Specifically, while the REMPARK data set includes data from patients ON and OFF therapy

performing free walking activities, in the 6MWT data set participants were in daily ON condition and performed a simple 6MWT. However, even considering these differences, no performance impairment is observed when testing the classification algorithm on a different data set, with data collected from different patients, under different conditions, and using a different sensor setting.

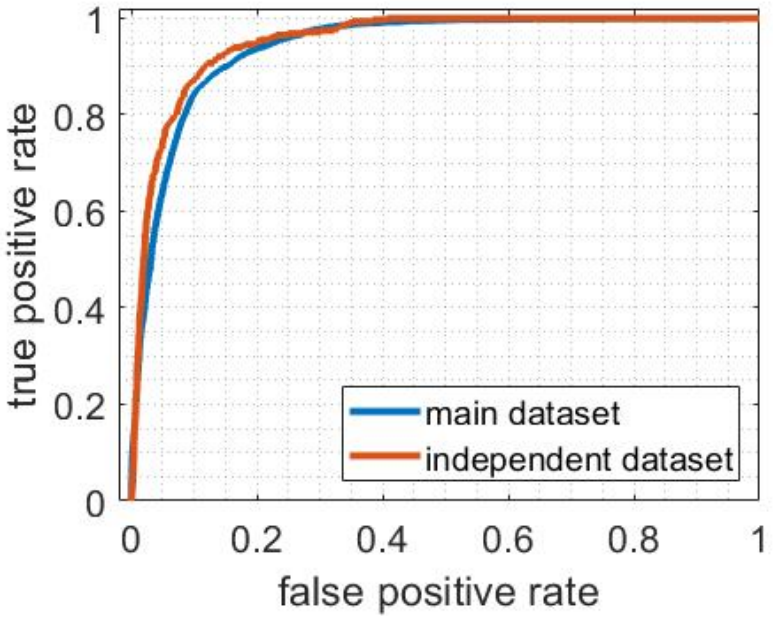


Fig. 4.6 Receiver operating characteristics of the classification model tested on the main and the external data set.

As far as concerns the true FOG episodes detected, 91.2% of episodes were correctly identified by the algorithm, with an average percentage of 68.7% of FOG detected in each episode. More specifically, 87%, 100%, and 100% of episodes of duration 0-5s, 5-10s, and >10s were identified by the system on the 6MWT data set, with an average proportion of 67.8%, 68.8%, and 74.5% of FOG detected in each episode (Table 4.10).

The false FOG episodes duration was found to be 2.5 $\pm$ 1.1 s, with 14.6%, 23.7%, and 33.9% of false episodes less than 5s, 10s, and 20s distant from the closest true FOG episode, respectively.

As for the time resolution in FOG detection, 32.5% of the FOG episodes were predicted before the actual onset, with an average advance of 1.3s (SD = 0.8s, min = 0.5s, max = 3.5s), while 50% of FOG episodes were detected with an average

delay of 1.1s (SD = 0.7s, min = 0.5s, max = 3s). The remaining 17.5% of FOG episodes were detected with a time delay larger than 3s. Analyzing the activities corresponding to false FOG episodes resulted in 6.7% of gait and 1.4% of stance classified as FOG. When testing the model on the control subjects of the 6MWT data set, a specificity equal to 1 was obtained, demonstrating excellent performance in rejecting false positives from elderly subjects without PD.

Specificity 0.977 was obtained when testing the model on the external ADL data set, which actually does not encompass true FOG episodes. The analysis of activities corresponding to false FOG episodes resulted in 4.7% of gait classified as FOG, 0.9% of stance, 3.4% of postural transitions (i.e., standing up, sitting down), 21% of the pull test, and 6.8% of the foot-tapping task, while the rest of the FPs were registered during unlabeled activities. However, it is worth noting that some activities (e.g., pull test, foot-tapping) are performed only during the MDS-UPDRS evaluation, and do not represent common ADLs.

Computational complexity

Figure 4.7 reports the testing time required by the model for different dimensions of the input data. More specifically, 43 ms are necessary for testing a single window, which represents the actual working condition for real-time applications. Moreover, 11 ms and 65 ms are required to classify 1000 (8.4 min of data) and 10000 windows (1.4 h of data) of data, respectively. Taking into consideration a FOG detection system receiving raw acceleration data from a single inertial sensor, the time required for the pre-processing steps should be taken into consideration. In fact, before undergoing the classification stage, mean removal should be performed on the raw signal, and the resulting data should be properly reshaped for input to the classification algorithm. However, the pre-processing time was found to be negligible compared to the classification time, with 0.07 ms required for mean removal and 0.003 ms for data reshaping. Finally, the total memory required by the model was 54.94 KB, with a single data window (i.e., 80 samples; Fs = 40 Hz) accounting for 44.10 KB, while the CNN parameters require only 10.84 KB.

The proposed model requires a total of 0.399M FLOPs to perform a prediction on a single (2s) window. This result is comparable with related DL methods such as that proposed in [217] (0.337M FLOPs), and significantly lower than those proposed in [156] (3.14M FLOPs) and [158] (4.76M FLOPs).

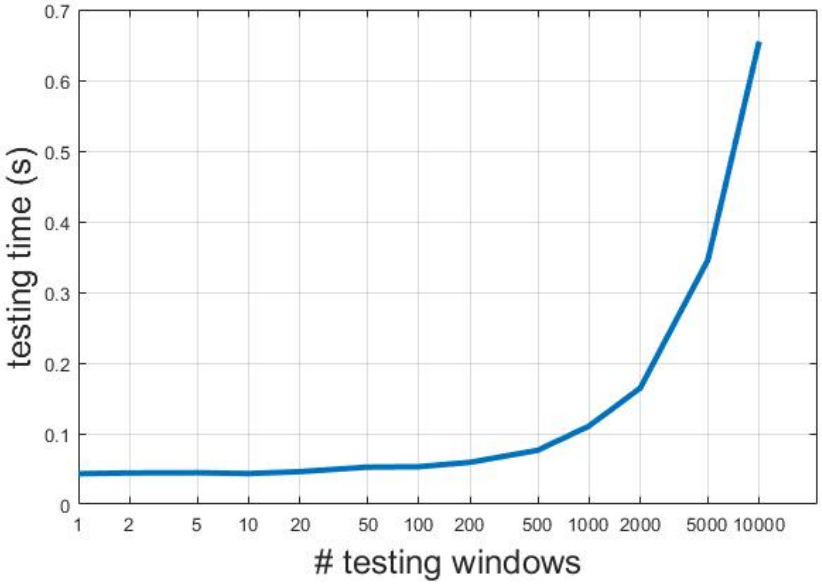


Fig. 4.7 Testing time for different input dimensions.

The STAT-ON device employs a lithium-polymer battery with 1200mAh capacity and a battery life of 7 days when continuously working for 8 hours [285]. The sensor has an average consumption of 4.1±4.2mA in current conditions, taking into account that the Bluetooth process is the most consuming process. In normal conditions, the Bluetooth system does not switch on and the consumption decreases to approximately 3.7mA. Currently, the total time spent by the microprocessor computing between samples is 0.279ms (25 ms available between samples) and the operations need 9.7ms every window to compute the window outcome plus 12.47ms to write the information to the Flash memory [135]. Given that the current algorithm needs 0.399 MFLOPS per window (2 seconds) and the processor executes approximately 210 Dhrystones Mega instructions per second (MIPS), the needed time in the current processor would be around extra 1.9ms in a period of 2 seconds. Thus, the change in battery life is practically insignificant, corresponding to an increase of 0.095% on-time extra that the battery needs to be active.

Activity threshold

Figure 4.8 reports the performance of the developed FOG detection algorithm, in terms of sensitivity and specificity, and the ratio of discarded windows while

increasing the threshold on the magnitude vector. The analyses were performed on the validation set. As can be observed, for threshold values smaller than 0.4, the

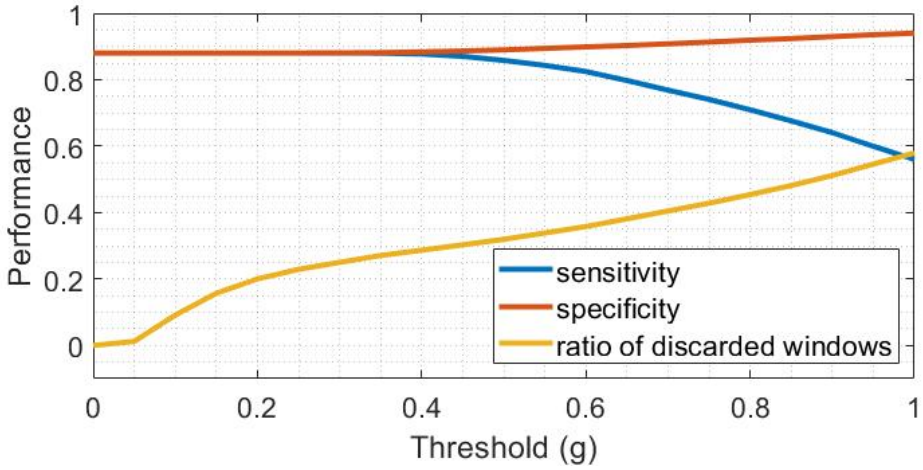


Fig. 4.8 The effect of the activity threshold tuning on the performance of the detection algorithm and on the ratio of discarded windows.

performance does not change significantly (-0.2% sensitivity and +0.4% specificity), while the ratio of discarded windows increases from 0 to 0.287. Further increasing the threshold value until 1g leads to a progressive improvement of specificity (from 0.884 to 0.941) and discard ratio (from 0.287 to 0.580), with an evident impairment of sensitivity (from 0.877 to 0.560). Table 4.11 reports the performance of the algorithm and the discard rate in absence of the activity threshold and for values equal to 0.4 and 0.7. Increasing the threshold from 0.4 to 0.7 produces a reduction of -10.8% in sensitivity and an increment of +2.6% in specificity and +11.8% in the rate of discarded windows.

Table 4.11 The effect of different activity thresholds on the performance of the detection algorithm and on the ratio of discarded windows.

Activity threshold	Sensitivity	Specificity	Discard rate
0	0.879	0.880	0
0.4	0.877	0.884	0.287
0.7	0.769	0.910	0.405

The effect of the activity threshold on the test set of the REMPARK data set, and on the two independent data sets is reported in Table 4.12. In general, the effect of the activity threshold is a progressive sensitivity reduction and slight specificity

improvement, as expected. Such an effect is more evident in the REMPARK data set than in the external data sets. Specifically, setting the threshold value to 0.4g produces a -3.1% reduction in sensitivity and a +2.5% increment in specificity in the REMPARK data set, while performance is not affected in the two external data sets. On the other hand, 27% and 20% of data were discarded prior to classification in the REMPARK and 6MWT data sets, respectively, while up to 63.2% of windows were discarded in the ADL data set.

Table 4.12 The effect of the activity threshold on the main test set and on the two independent data sets.

Data set	Performance	Activity threshold		
		0	0.4	0.7
REMPARK	sensitivity	0.877	0.846	0.801
	specificity	0.883	0.908	0.911
	discard rate	0	0.270	0.376
6MWT	sensitivity	0.796	0.796	0.744
	specificity	0.933	0.934	0.938
	discard rate	0	0.201	0.227
ADL	specificity	0.977	0.978	0.981
	discard rate	0	0.632	0.692

4.2 Postural stability

4.2.1 Turn quality and postural stability assessment using smartphones

Participants in G-PD1 performed an average of 3.6 turns per patient, with a total of 213 turns recorded. In G-PD2 subjects, 126 turns were recorded, with an average of 9.7 turns per patient. Specifically, 59 turns were recorded from patients with moderate impairment and 67 from the severe impairment class. The final set of selected features is shown in Table 4.13. Interestingly, 5 out of 9 features are related to vertical angular velocity, which provides a measure of turning speed and has previously been reported to be sensitive to motor symptom severity in [286, 256].

Table 4.13 List of features selected for binary classification. α and ω represent acceleration and angular velocity, respectively.

Feature	Component	r	p-value
1	$\omega_x, \alpha_z, \alpha_x$	0.66, 0.54, 0.49	< 0.001
2	ω_x	0.63	< 0.001
3	ω_x, α_z	0.63, 0.54	< 0.001
6	ω_x	0.55	< 0.001
9	ω_x, α_z	0.58	< 0.001

The classification performance obtained by the ML algorithms is shown in terms of accuracy and AUC in Table 4.14. LDA provided the best accuracy and AUC, with sensitivity, specificity, and precision of 0.955, 0.881, and 0.901, respectively. The confusion matrix is shown in Figure 4.9. LDA detects 95.5% of the 180 rounds that, with a probability of 90.1%, were performed by patients belonging to the *S* class. Of the 67 (59) turns performed by patients belonging to class *S* (*M*), 64 (52) were correctly classified and only 3 (7) were incorrectly classified.

Table 4.14 Performance of different machine learning models for the binary (moderate vs severe motor impairment) classification task.

Classifier	Parameters	Accuracy	AUC
KNN	Number of neighbors: 33. Distance metric: Euclidean. Distance weight: equal	0.881	0.93
Linear SVM	Kernel function: linear. Boxconstraint: 0.13	0.913	0.95
LDA	Delta: 1×10^{-6}. Gamma: 4×10^{-4}	0.921	0.97

LDA also achieved the best results in LOSO validation. More specifically, it correctly classified all 7 subjects belonging to the *S* class and misclassified only one subject with motor conditions *M*. Therefore, it achieved a unit sensitivity in identifying patients with severe motor conditions and an accuracy of 0.880. This demonstrates a good generalization ability of the model. On the other hand, the number of misclassified subjects was 2 for SVM and KNN and 3 for DT.

The LDA classifier was then tested on the 59 PwPD belonging to the G-PD1 subgroup. The Pearson correlation coefficient between the QoM index and clinical

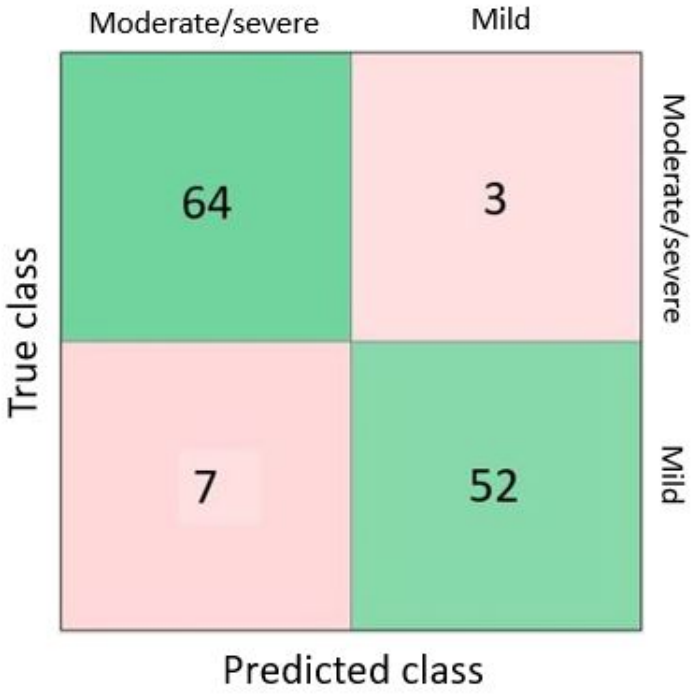


Fig. 4.9 Confusion matrix related to LDA classifier.

scores is given in Table 4.15. It can be observed that the QoM index has a strong correlation ($r > 0.7$) with MDS-UPDRS item 3.12 "postural stability" and with the mean motor score of the MDS-UPDRS Part III. The boxplot of QoM values as a function of the patient's PS score is also shown in Figure 4.10.

It is worth noting that the QoM index evaluation involved 59 PwPD not engaged in the training phase, with different disease progression and gait/postural impairment. This demonstrates the significance of the QoM index and the reliability of the proposed method. In addition, these patients performed a simple walking protocol (i.e., the task required for clinical assessment of MDS-UPDRS item 3.10 "gait"), which is representative of walking during daily activities. Therefore, the proposed method may be suitable for implementation in an unsupervised home setting.

Table 4.15 Pearson correlation coefficient r between the QoM index and some clinical MDS-UPDRS part III scores.

Clinical score	r	p-value
3.9 *arising from chair*	0.5	< 0.001
3.10 *gait*	0.61	< 0.001
3.12 *postural stability*	0.73	< 0.001
H&Y	0.66	< 0.001
Average Score	0.75	< 0.001

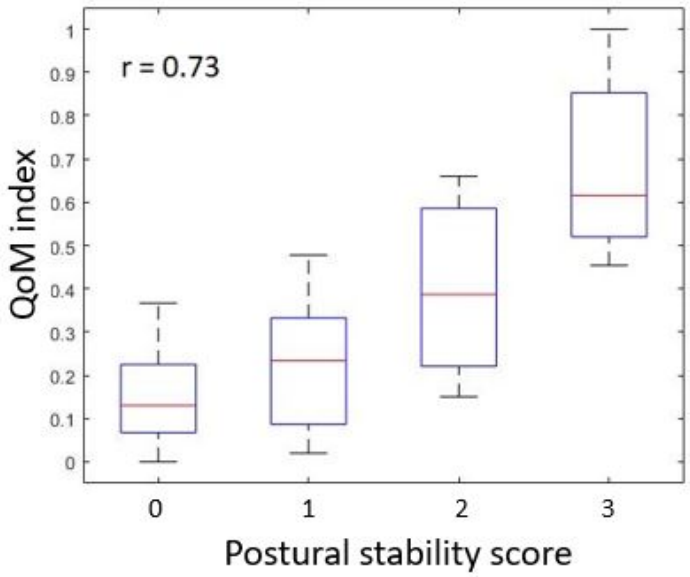


Fig. 4.10 Boxplot of QoM index vs Item 3.12 "postural stability" score.

4.2.2 Postural stability assessment during quiet stance using smartphones

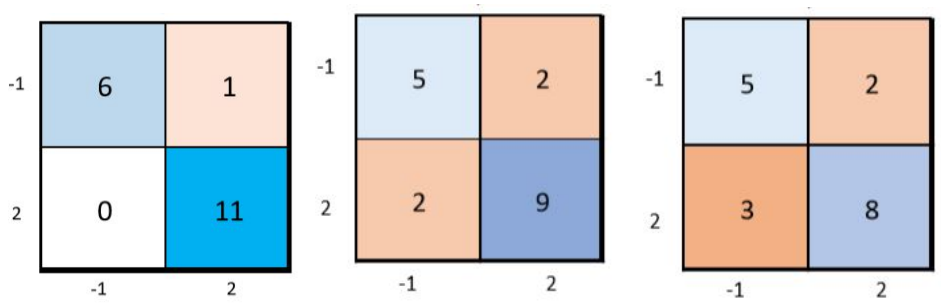
The final feature subset obtained from the feature selection algorithm is reported in Table 4.16.

The performance of the models in the multi-class classification task is reported below. In the LOSO validation, SVM with linear kernel provided the best results, achieving an accuracy of 0.701 and an average F-score of 0.673. The RF model provided an accuracy of 0.650 and an average F-score of 0.486. The results obtained with the proposed approach are given in the following. Regarding the first level of

Table 4.16 The most significant and non-redundant features. All the correlation coefficients reported were associated with a p-value < 0.05.

ID	Study	Feature	Component	r
3	Present Study	Variance	z	0.4
21	Present Study	Spectral Entropy	y, z	0.44, 0.53
22	[267],[173]	f_{50}	x, z	0.52, 0.55
25	Present Study	RAPP $\frac{(2-3)Hz}{(0-2)Hz}$	x,y	0.69, 0.54
25	Present Study	RAPP $\frac{(1-2Hz)}{(0-1)Hz}$	x	0.58

classification, which aims to classify subjects with very different postural control, the results obtained for the different ML models are shown in Figure 4.11, in terms of confusion matrices.



(a) *Linear-SVM* (b) *kNN* (c) *DT*

Fig. 4.11 Confusion matrices of different ML models. Label -1 and 2 indicate control and class 2 subjects, respectively.

The results were obtained using LOSO validation. As can be observed, SVM provided the best performance in differentiating controls from PwPD with severe impairment of postural control. Therefore, this model was used in the subsequent processing steps. Figure 4.17 shows the soft output of the SVM model, along with the line of best fit, i.e., the line that minimizes the MSE. There is a significant gap between controls and class 0 subjects, as well as between class 0 and class 2 subjects. In addition, class 1 data partially overlap with class 0 and 2 data. However, a high Pearson correlation coefficient was achieved between soft SVM output and clinical score, i.e. $r = 0.76 with p < 0.0001$. These results prove the potential effectiveness of the proposed approach.

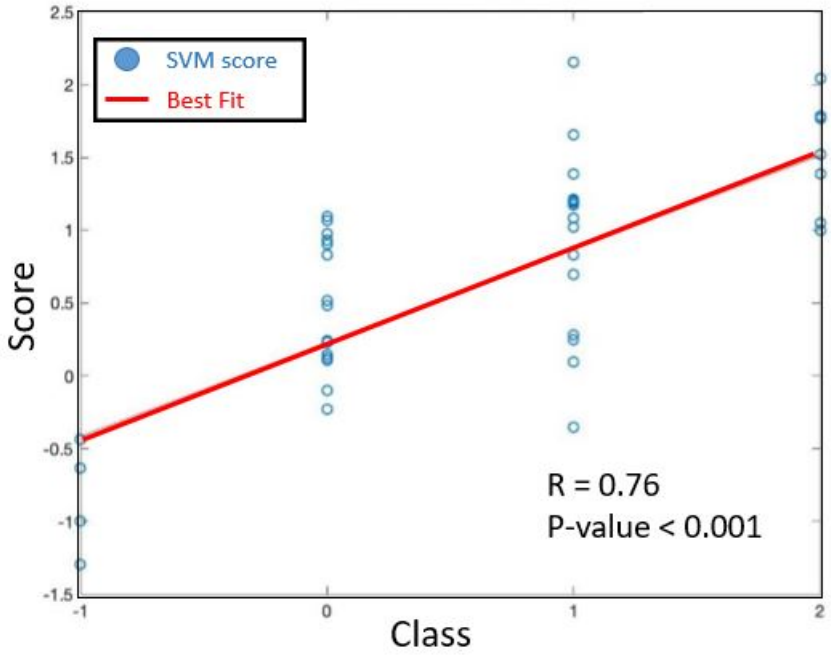
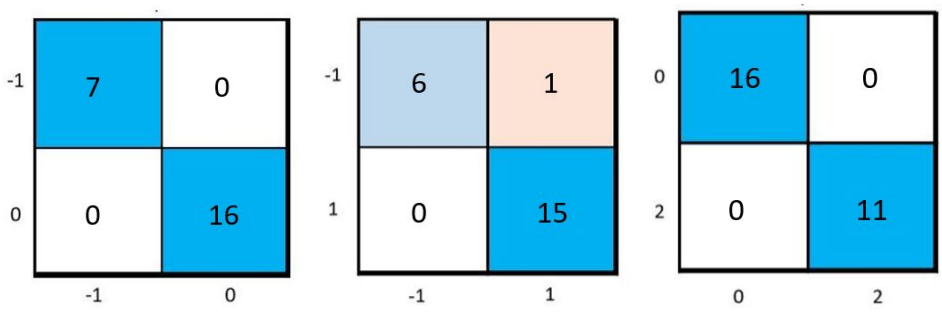


Fig. 4.12 Distribution of the continuous output provided by the SVM model, along with the best-fit line.

The performance of the three binary SVM models (second layer) is reported in Figure 4.13, in terms of confusion matrices, and in Table 4.17, in terms of accuracy, sensitivity, specificity, precision, and F-score.



(a) *Control vs Class 0* (b) *Control vs Class 1* (c) *Class 0 vs Class 2*

Fig. 4.13 Confusion matrices obtained with three binary SVM classifiers. Label -1 and 0 and 2 denote control, class 1, and class 2 subjects respectively.

As can be observed from Figure 4.13 and Table 4.17, the performance obtained is very satisfactory, with high accuracy in three of the five classification tasks. However,

Table 4.17 Classification performance achieved using the three binary SVM models.

Task	Accuracy	Sensitivity	Specificity	Precision	F-score
Control vs Class 0	1	1	1	1	1
Control vs Class 1	0.950	1	0.860	0.940	0.970
Class 0 vs Class 1	0.828	0.857	0.800	0.800	0.828
Class 0 vs Class 2	1	1	1	1	1
Class 1 vs Class 2	0.720	0.636	0.786	0.700	0.712

the classification task gives lower performance when trying to distinguish subjects belonging to adjacent classes. This may be due to both a more difficult classification task and the intra- and inter-rater variability, which represents an inherent uncertainty of one class in the clinical assessment [98, 29, 70]. The latter consideration suggests that one should not rely on such fine grading between adjacent classes, as it may provide misleading results. It is worth noting that these results were obtained using only 8 features and three very simple SVM models.

Satisfactory performance was obtained for multiple binary classifications, allowing discrimination between control subjects and patients with mild, slight, and severe postural instability. Regarding the practical use of the algorithm, a first approach may be to use the regression model described in Figure 4.17 to obtain an initial indication of the extent of postural control impairment (e.g., mild). Then, the appropriate binary SVM model can be applied to perform finer classification. This approach could be reasonable and efficient. For completeness, since the input is a short inertial signal and only 8 features and simple SVM were employed, the processing time was extremely short (5 ms for data loading, 50 ms for feature extraction, and 2 ms for classification). This makes possible a real-time on-board implementation of the algorithm.

4.3 Gait impairment

4.3.1 Prediction of postural stability and gait difficulty using a single inertial measurement unit

Clinical-Behavioural Correlations

Pearson's correlation analysis showed that most of the time and frequency domain features were significantly correlated with PIGD scores. In more detail, as axial motor control worsened, the minimum value of inertial signals increased, while the maximum and mean square value of signals, mean peak height in the time domain, and dominant harmonic height decreased. Table 4.18 summarizes the Pearson correlation coefficients and respective p-values for the different feature–PIGD pairs. Only the most informative features for both treatment conditions, that is, those with a Pearson correlation coefficient with a PIGD score greater than 0.5, were included in the table. Figure 4.14 shows the scatter plots of the mean height of the dominant harmonic (mean DH height) versus the PIGD OFF and ON scores.

Table 4.18 Correlation between engineered features and PIGD score. Results are reported in terms of Pearson correlation coefficient and relative p-value. For each feature, significant components are reported. α: acceleration; ω: angular velocity; θ: orientation.

Feature	Component	Pearson Correlation Coefficient (p-value)	
		PIGD OFF	PIGD ON
Min	α_y	0.54 (0.002)	0.58 (<0.001)
Mean	ω_x	0.64 (<0.001)	0.54 (0.002)
RMS	α_y	−0.67 (<0.001)	−0.57 (0.002)
	ω_x	−0.74 (<0.001)	−0.72 (<0.001)
hPeaks	α_y	−0.70 (<0.001)	−0.58 (<0.001)
	ω_x	−0.74 (<0.001)	−0.60 (<0.001)
DH height	α_y	−0.59 (<0.001)	−0.57 (<0.001)
	ω_x	−0.71 (<0.001)	−0.70 (<0.001)
	θ_x	−0.69 (<0.001)	−0.65 (<0.001)

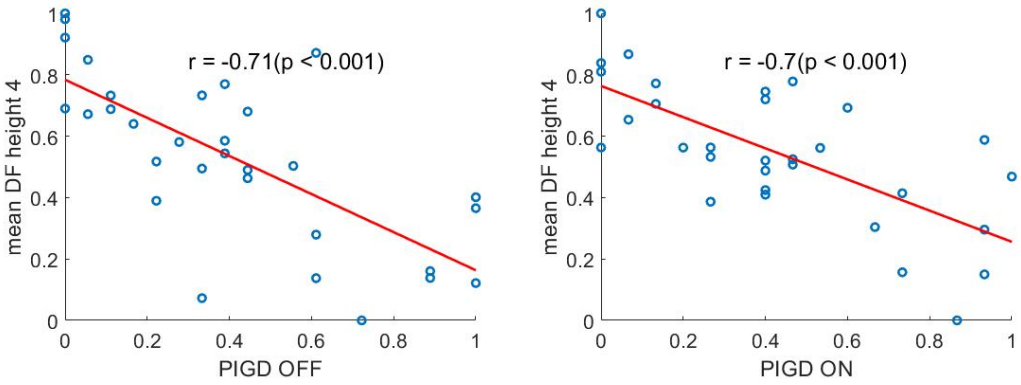


Fig. 4.14 Correlation plot between the average principal harmonic height for the x-axis angular velocity (DH height 4) and PIGD score. Data and PIGD score refer to patients OFF (left) and ON (right) therapy. PIGD: postural instability and gait difficulty.

PIGD Score Regression

This section reports the results of the optimized SVR model in LOSO validation. Specifically, the effect of L-Dopa was evaluated by comparing regression models in PwPD OFF and ON therapy. For each pharmacological condition, the best model configuration was identified and the performance of the regression models was compared. The effect of FOG was evaluated by comparing the regression models in *FOG+* and *FOG-*. The best model configuration was extracted for each subgroup of patients, and the performance of the regression models was compared.

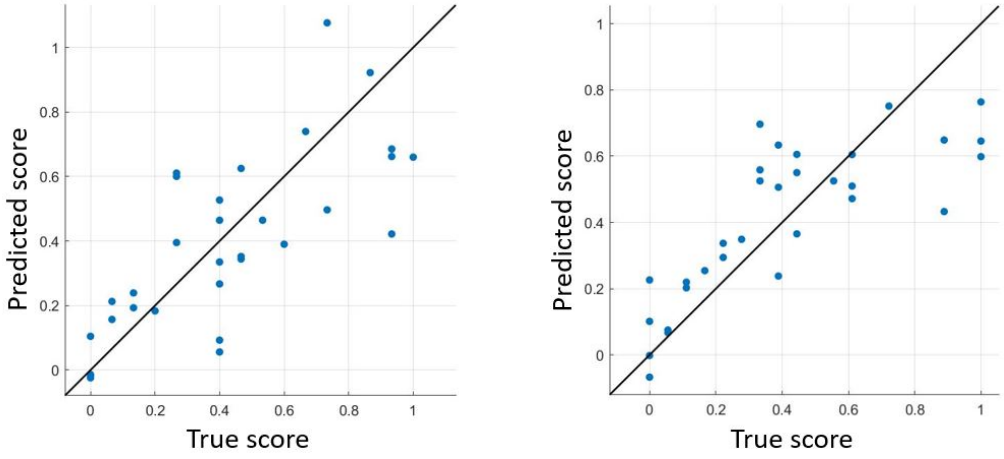
The Effect of L-Dopa

Wilcoxon's test showed that both MDS-UPDRS part III score and PIGD score were different in OFF and ON therapy patients ($p < 0.001$). Table 4.19 summarizes the performance of the regression model in terms of Pearson correlation coefficient, RMSE, and MAE, in PwPD OFF and ON therapy. Results are reported for different dimensions of the feature set and different methods of dimensionality reduction. Based on the results in Table 4.19, the following considerations can be made. SVR with linear kernel was selected in % of cases; the best performance was obtained with linear kernel and small values of the box-constraint parameter (i.e., < 0.009). Increasing the size of the feature set did not provide progressively better performance. The best results were obtained with n = 15 features, both on patients OFF and ON therapy. For larger feature sets (i.e., # features > 15), PCA-based dimensional-

ity reduction always provided better results, compared with those obtained with correlation-based feature selection. PCA-based dimensionality reduction led to the best results for both PwPD OFF and ON therapy. Regression models provided better performance in OFF patients than in ON patients. Accordingly, the best regression model parameters were identified for each therapy condition. Then, these models were trained on patients ON (OFF) therapy and tested on patients OFF (ON) therapy. This procedure resulted in r = 0.70 (0.67), RMSE = 0.57 (0.42), and MAE = 0.47 (0.15). When the model was tested using LOSO on all available data, regardless of therapy, r = 0.64, RMSE = 0.22 and MAE = 0.17 were obtained from an SVR with linear kernel and box-constraint = 0.07. Figure 4.15 shows the scatter plot of the true score versus the predicted score, together with the best-fit line.

Table 4.19 Performance of regression models in different pharmacological conditions using leave-one-subject-out validation. Results were obtained for different sizes of the feature set and different dimensionality reduction methods. Best performance are marked with bold type.

# Features	Therapy	Dimensionality Reduction		Model parameters		Performance		
		Method	Value	Kernel	Cost	r	RMSE	MAE
5	ON	r (min-max)	0.65–0.72	linear	10.9	0.71	0.22	0.18
		var (%)	82.9	linear	0.09	0.71	0.25	0.20
	OFF	r (min–max)	0.76–0.77	gaussian	2.67	0.76	0.21	0.18
		var (%)	77.3	linear	0.006	0.77	0.21	0.16
10	ON	r (min–max)	0.58–0.72	linear	0.54	0.27	0.22	
		var (%)	93.0	linear	0.07	0.55	0.26	0.20
	OFF	r (min–max)	0.74–0.77	linear	0.35	0.69	0.23	0.19
		var (%)	88.7	linear	1.91	0.51	0.28	0.23
15	ON	r (min–max)	0.56–0.72	linear	0.003	0.67	0.23	0.19
		var (%)	97.5	linear	0.009	**0.75**	**0.20**	**0.16**
	OFF	r (min–max)	0.68–0.77	gaussian	253.51	0.69	0.23	0.19
		var (%)	94.7	linear	0.001	**0.79**	**0.19**	**0.15**
20	ON	r (min–max)	0.55–0.72	linear	0.002	0.71	0.22	0.16
		var (%)	99.2	linear	641.6	0.5	0.28	0.24
	OFF	r (min–max)	0.66–0.77	linear	0.004	0.79	0.20	0.15
		var (%)	97.9	gaussian	0.87	0.76	0.21	0.15
25	ON	r (min–max)	0.52–0.72	linear	0.005	0.66	0.24	0.19
		var (%)	99.8	linear	0.003	0.71	0.22	0.16
	OFF	r (min–max)	0.62–0.72	linear	0.004	0.78	0.20	0.15
		var (%)	99.5	cubic	0.19	0.75	0.21	0.16



(a) (b)

Fig. 4.15 Regression results for patients under different pharmacological conditions. Data are plotted using a scatter plot and the regression line is reported as the best-fit line. (**a**) Patients under dopaminergic therapy. (**b**) Patients not under dopaminergic therapy.

The Effect of Freezing of Gait

Table 4.20 reports the demographic and clinical characteristics of PwPD with and without FOG, along with the significance level calculated with the Mann–Whitney U test. In PwPD with FOG, the duration of FOG was 4.8 *pm* 2.7 and the total FOG-Q score was 15.4 *pm* 4.4. Regarding engineering characteristics, Mann–Whitney's U test showed that patients with FOG had higher *Min* and lower *Mean*, *RMS*, *DHheight*, and E_{tot} than those without FOG ($p < 0.001$).

Table 4.20 Demographic and clinical features of patients enrolled in the present study (mean $\pm$ standard deviation).

Group	# Patients (Male)	Age (Years)	MDS-UPDRS-III OFF (ON)	PIGD OFF (ON)
FOG+	17 (13)	72.0 $\pm$ 7.6	40.9 $\pm$ 13.2 (32.9 $\pm$ 14.1)	11.2 $\pm$ 4.5 (9.6 $\pm$ 3.4)
FOG-	14 (10)	71.8 $\pm$ 6.4	29.7 $\pm$ 12.3 (21.9 $\pm$ 10.8)	2.6 $\pm$ 2.5 (2.4 $\pm$ 2.3)
p	0.353 (0.531)	0.811	0.054 (0.030)	<0.001 (<0.001)

Table 4.21 reports the performance of the optimized regression models for *FOG+* and *FOG-* patients ON state of therapy.

When the feature set size increased, the regression models provided comparable performance in *FOG+* and *FOG-*. This is particularly evident for n = 25 features, for which *r* and RMSE were very similar in the two populations, regardless of the

Table 4.21 Performance of regression models in patients with (FOG+) and without (FOG-) freezing of gait, both under dopaminergic therapy. Results were obtained for different feature set sizes and different dimensionality reduction methods. Best performances are marked with bold type.

# Features	Group	Dimensionality Reduction		SVM parameters		Performance		
		Method	Value	Kernel	Cost	r	RMSE	MAE
5	FOG+	r (min–max)	0.58–0.69	linear	0.007	0.7	0.28	0.22
		var (%)	85.2	linear	0.02	0.63	0.30	0.25
	FOG-	r (min–max)	0.77–0.84	linear	0.07	**0.85**	**0.19**	**0.13**
		var (%)	82.1	linear	0.007	0.77	0.22	0.15
10	FOG+	r (min–max)	0.55–0.69	linear	1.69	0.5	0.34	0.26
		var (%)	96.6	linear	0.03	0.47	0.34	0.29
	FOG-	r (min–max)	0.69–0.84	linear	0.02	0.83	0.19	0.15
		var (%)	96.5	linear	0.004	0.83	0.19	0.15
15	FOG+	r (min–max)	0.53–0.69	linear	0.51	**0.71**	**0.27**	**0.21**
		var (%)	99.8	linear	0.06	0.64	0.30	0.24
	FOG-	r (min–max)	0.66–0.84	quadratic	0.019	0.7	0.25	0.20
		var (%)	99.8	linear	0.15	0.76	0.22	0.18
20	FOG+	r (min–max)	0.50–0.69	linear	0.007	0.58	0.32	0.28
		var (%)	99.9	linear	0.46	0.67	0.29	0.23
	FOG-	r (min–max)	0.64–0.84	linear	0.004	0.81	0.20	0.25
		var (%)	99.9	linear	0.01	0.76	0.22	0.17
25	FOG+	r (min–max)	0.46–0.69	linear	0.014	0.7	0.28	0.24
		var (%)	99.9	linear	4.02	0.71	0.27	0.22
	FOG-	r (min–max)	0.62–0.84	linear	0.02	0.77	0.22	0.19
		var (%)	99.95	linear	85.9	0.76	0.22	0.18

dimensionality reduction method. From the above results, the following observations can be made. SVR with linear kernel was selected in 95% of cases; the best performance was obtained with linear kernel and small values of the box-constraint parameter (i.e., < 0.51). Increasing the size of the feature set did not provide progressively better performance. The best results were obtained with n = 15 (n = 5) features in patients with (without) FOG. Correlation-based and PCA-based dimensionality reduction methods provided similar results regardless of feature set size. The correlation-based dimensionality reduction method yielded the best results for both *FOG+* and *FOG-* patients. Regression models provided better performance in *FOG-* patients, as reflected by the higher values of *r* and lower values of *RMSE*.

Based on the above considerations, the feature set size was set at 15 (5) for *FOG+* (*FOG-*) patients, and the dimensionality reduction method was based on correlation for both populations. Then, the best regression model was trained on *FOG+* (*FOG-*) patients and tested on *FOG-* (*FOG+*). This procedure produced r = 0.34 (0.40), RMSE = 0.43 (0.39), and MAE = 0.37 (0.35), respectively.

The performance gap between PwPD with and without FOG may be due to the different discrimination power of some features in the two populations. From Table 4.21, it appears that for each dimension of the feature set, the correlation between top-ranked features and PIGD score is higher in patients without FOG. The top-ranked features for these patients were found to be *Min* (r = 0.79, p = 0.001), *vPeaks* ($r = -0.75$, p = 0.004), *RMS* ($r = -0.72$, p = 0.006), *hPeaks* ($r = -0.72$, p = 0.006), *hPeaks* (r = -0.70, p = 0.008) from the orientation signal of the x axis, and E_{tot} ($r = -0.79$, p = 0.001) from the angular velocity signal of the x axis. As far as concerns PwPD with FOG, top-ranked features included *Min* (r = 0.66, p =0.004), *DHheight* ($r = -0.62$, p =0.008), *RMS* (r = -0.58, p =0.015), *hPeaks* ($r = -0.58$, p = 0.016) from the y-axis acceleration signal, and *DHheight* ($r = -0.65$, p =0.005) from the x-axis orientation signal.

Table 4.22 reports the performance of the optimized regression models for *FOG+* and *FOG-* OFF state of therapy.

Based on the above results, the following considerations can be made. SVR with linear kernel was selected in 95% of cases; the best performance was obtained with linear kernel both in patients with and without gait freezing. Increasing the feature set size did not provide progressively better performance. The best results were obtained with n = 25 (n = 15) features in patients with (without) gait freezing. PCA-based dimensionality reduction (correlation) was chosen for patients with (without) FOG. Regression models provided slightly better performance in patients without FOG, in terms of RMSE, regardless of model configuration; performance in terms of r depended on regression model parameters, with better results in patients with FOG (Table 4.22).

Based on the above considerations, the feature set size was set to 25 (15), and the PCA-based dimensionality reduction method (correlation) for *FOG+* (*FOG-*). Then, the best regression models were trained on emphFOG+ (emphFOG-) patients and tested on emphFOG- (*FOG+*). This procedure yielded r= 0.73 (0.69), RMSE = 0.36 (0.33), and MAE = 0.25 (0.25), respectively.

Table 4.22 Performance of regression models in patients with (FOG+) and without (FOG-) freezing of gait, both not under dopaminergic therapy. Results were obtained for different feature set sizes and different dimensionality reduction methods. Best performances are marked with bold type.

# Features	Group	Dimensionality reduction		SVM parameters		Performance		
		Method	Value	Kernel	Cost	r	RMSE	MAE
5	FOG+	r (min–max)	0.58–0.65	linear	2.7	0.54	0.33	0.29
		var (%)	83.1	gaussian	2.3	0.65	0.30	0.25
	FOG-	r (min–max)	0.70–0.76	linear	0.01	0.74	0.23	0.16
		var (%)	83.5	linear	0.06	0.65	0.26	0.20
10	FOG+	r (min–max)	0.55–0.65	linear	0.93	0.76	0.25	0.18
		var (%)	95.7	linear	0.004	0.78	0.25	0.22
	FOG-	r (min–max)	0.69–0.76	linear	83.9	0.75	0.23	0.18
		var (%)	97.2	linear	0.003	0.75	0.23	0.15
15	FOG+	r (min–max)	0.52–0.65	linear	0.03	0.67	0.29	0.24
		var (%)	99.7	linear	118.5	0.63	0.30	0.26
	FOG-	r (min–max)	0.63–0.76	linear	0.006	**0.79**	**0.21**	**0.15**
		var (%)	99.6	linear	0.002	0.69	0.25	0.16
20	FOG+	r (min–max)	0.50–0.65	linear	0.37	0.82	0.22	0.19
		var (%)	99.8	linear	0.05	0.79	0.24	0.21
	FOG-	r (min–max)	0.61–0.76	linear	0.009	0.78	0.22	0.14
		var (%)	99.8	linear	0.002	0.71	0.24	0.16
25	FOG+	r (min–max)	0.48–0.65	linear	621.2	0.81	0.23	0.19
		var (%)	99.9	linear	24.2	**0.83**	**0.22**	**0.19**
	FOG-	r (min–max)	0.59–0.76	linear	0.69	0.75	0.23	0.16
		var (%)	99.9	linear	0.12	0.69	0.25	0.17

Table 4.23 shows all the results obtained for each test population and for each pharmacological condition; the results were obtained using the LOSO test. As is evident from the table, the performance of the model improves when patients with different pharmacological conditions are considered separately. Regarding the effect of FOG, if the model is specifically trained on *FOG+* and *FOG-* separately, the performance improves significantly in patients without FOG in the ON condition and in patients with FOG in the OFF condition. Table 4.24 reports the results obtained by training and testing the regression model on different populations (i.e., therapy ON vs. therapy OFF, FOG+ vs. FOG-). The prediction errors provided by the global model (i.e., the regression model trained and validated on all subjects, regardless of therapy condition and gait freeze) were compared with those obtained using different models

for each pharmacological condition separately. The Wilcoxon test showed that the difference in prediction errors was not statistically significant ($p = 0.074$); therefore, a single model can be used to estimate the PIGD score. On the other hand, training the model on specific subgroups (e.g., patients with FOG, patients ON therapy) and testing on different subgroups resulted in a large reduction in performance, as evident from the RMSE values shown in Table 4.24. Summarizing these results, it is possible to implement a very general algorithm, but care must be taken to collect a very general dataset, including patients in different therapy conditions as well as patients with and without FOG.

Table 4.23 Performance of regression models on different populations under different pharmacological conditions.

Therapy	FOG	Performance		
		r	RMSE	MAE
All	All	0.64	0.22	0.17
ON	All	0.75	0.20	0.16
	FOG+	0.71	0.27	0.21
	FOG-	0.85	0.19	0.13
OFF	All	0.79	0.19	0.15
	FOG+	0.83	0.22	0.19
	FOG-	0.79	0.21	0.15

Table 4.24 Performance of the regression model for different combinations of training and test samples.

Training sample	Testing sample	Performance		
		r	RMSE	MAE
ON	OFF	0.70	0.57	0.47
OFF	ON	0.67	0.42	0.15
FOG+ (ON)	FOG- (ON)	0.34	0.43	0.37
FOG- (ON)	FOG+ (ON)	0.40	0.39	0.35
FOG+ (OFF)	FOG- (OFF)	0.73	0.36	0.25
FOG- (OFF)	FOG+ (OFF)	0.69	0.33	0.25

The large prediction errors observed during training and testing of the model on different populations may be due to the different discrimination power of some

features. As can be seen in Figure 4.16, the sensitivity of some features to changes in the PIGD score depends on the medication condition. Some features have a strong correlation with PIGD in patients ON therapy but not in those OFF therapy, and vice versa. The same behavior is observed when training and testing the regression model on patients with and without FOG while on therapy. As already reported in Table 4.21, the highest-rated features were different in FOG+ and FOG- patients, so the performance of the prediction model deteriorates when it is trained and tested on different populations.

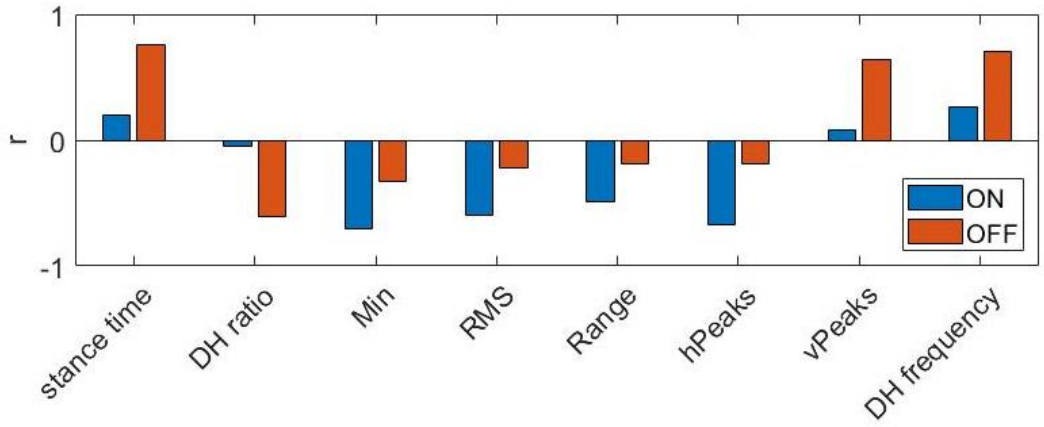


Fig. 4.16 Pearson correlation coefficient between features and PIGD score. DH ratio and DH frequency refer to the component θ_x; Min and vPeaks refer to the component ω_x; RMSE and Range refer to the component α_y.

Finally, from Table 4.24, it can be observed that performance was not significantly impaired when the model was trained with data from FOG+ (FOG-) patients and tested on data from FOG- (FOG+) patients while OFF therapy. In this case, the most common features classified were the minimum value, mean square value, and mean peak height of the orientation signal; the mean value of the angular velocity signal; and the height of the dominant harmonic of the acceleration signal along the y axis.

4.3.2 Prediction of self-perceived gait impairment

Table 4.25 reports the performance of the proposed predictive model, in terms of Pearson's correlation coefficient with AIM (i.e., the sum of items 2.11-2.13 of the MDS-UPDRS), in patients on OFF and ON therapy. The model provided better performance in PwPD in the absence of therapy, as reflected by all performance evaluation metrics. Wilcoxon's test showed that predictions were significantly different

in PwPD OFF and ON therapy ($p = 0.014$). Figure 4.17 shows the correlation plot for PwPD OFF and ON therapy. The best-fit line equation, Pearson's correlation coefficient, and sample size are given. In PwPD OFF therapy, higher correlations and smaller prediction errors are observed.

Table 4.25 Pearson correlation coefficient of the prediction model with AIM (sum of MDS-UPDRS items 2.11-2.13), in PDPs under (ON) and not under (OFF) dopaminergic therapy.

Therapy	r (p)	RMSE	MAE
OFF	0.86 (<0.001)	1.72	1.52
ON	0.76 (<0.001)	2.30	1.70

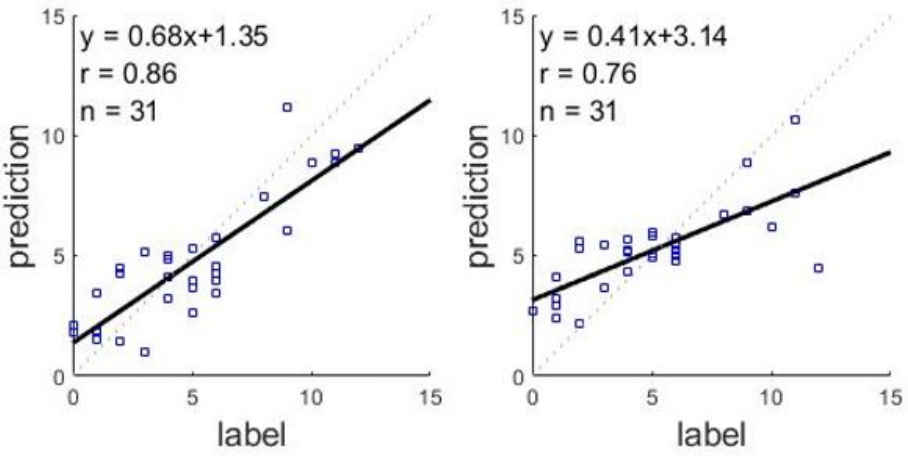


Fig. 4.17 Correlation plot for PDPs OFF (left) and ON (right) therapy.

Table 4.26 reports the Pearson correlation coefficient and respective p-value for other prediction-clinical score pairs, in PwPD OFF and ON therapy. The predicted score showed a moderate to strong correlation with disease progression (as measured by H&Y score), the clinically assessed level of motor impairment (as measured by MDS-UPDRS total score part III), and axial impairment (as measured by PIGD score). Again, higher correlations are observed for off-therapy PwPD.

Table 4.27 shows the final set of features used by the prediction model for PwPD OFF and ON therapy. As can be observed, the features selected for PwPD ON therapy are a subset of those used for PwPD OFF therapy. As AIM worsens, the minimum and average values of the inertial signals increase, while the motion

Table 4.26 Correlation between prediction and various clinical scores, for PDPs OFF and ON therapy.

Clinical score	r (p-value)	
	OFF	ON
H&Y	0.68 (<0.001)	0.48 (0.007)
MDS-UPDRS-III	0.58 (<0.001)	0.53 (0.002)
PIGD	0.79 (<0.001)	0.69 (<0.001)

intensity (measured by the standard deviation) and signal energy (measured by the total energy and the amplitude of the dominant harmonic) decrease.

Table 4.27 Feature selected in the optimization procedure, for patients OFF and ON therapy. α : acceleration; ω : angular velocity; θ : orientation

Feature	Component		Trend
	OFF	ON	
Min	ω_x, θ_x	θ_x	increase
Mean	ω_x	-	increase
Std	$\alpha_y, \alpha_z, \omega_x, \omega_y, \theta_x$	ω_x	decrease
DH_h	$\alpha_y, \alpha_z, \omega_y, \omega_x$	ω_x	decrease
E_{tot}	α_z, ω_x	ω_x	decrease

It is worth noting that the regression model was trained and validated to predict the sum of MDS-UPDRS items 2.11-2.13. In addition to demonstrating excellent predictive ability on that metric, which represents the subjective perception of motor impairment in the preceding days, the output of the algorithm was found to be correlated with important clinical items. Comparing the present results with that of [239], a higher correlation with the MDS-UPDRS part-III total score ($r = 0.58$ and $r = 0.53$ in PwPD OFF and ON therapy vs $r = 0.48$) and PIGD score ($r = 0.79$ and $r = 0.69$ in PwPD OFF and ON therapy vs $r = 0.61$) was obtained in this study. Furthermore, whereas in [239] data from 3 sensors were recorded during different activities (i.e. gait, turn, stance), in this study, only gait data recorded with a single sensor were analyzed. Compared with [216], the correlation with PIGD is the same in patients not under therapy, while a lower correlation is observed in patients ON

therapy ($r = 0.69$ vs $r = 0.75$). However, it is worth noting that the proposed algorithm was not developed for the prediction of PIGD, as this is not the main objective of this study.

4.4 Bradykinesia

The final parameters selected for each addressed classification algorithm were as follows. Linear kernel function and cost equal to 36 for SVM. Number of neighbors equal to 5, euclidean distance, and equal distance weight for kNN. Gini-Simpson diversity index [204] split criterion and maximum number of split equal to 4 for DT. Number of hidden layers equal to 2, number of neurons per layer of 16, and hyperbolic tangent sigmoid transfer function for ANN.

4.4.1 Classification results

Table 4.28 summarizes the performance of each model in terms of AUC.

Table 4.28 Performance of several ML methods in case of discrete output.

Method	Accuracy	AUC
DT	0.591	0.53
kNN	0.603	0.82
SVM	0.609	0.80
ANN	0.777	0.92

The ANN model shows the best performance among the implemented ML classifiers (accuracy 0.777), exceeding the results reported in the literature (e.g., in [241, 185] an accuracy of 0.430 is reported). All methods, with the exception of DT, misclassify by a maximum of one step on the MDS-UPDRS scale, i.e., the misclassification error is 1 in 100% of cases. This error is comparable to inter-rater variability, as also discussed in [185].

When using the continuous ANN output, the error histogram shown in Figure 4.18 reports $leq0.5$ error in more than 80% of instances. This suggests that a finer discretization (i.e., 0.5 instead of unit steps) could certainly improve the performance

of the algorithm and provide a more accurate score. For completeness, the Bland-Altman diagram shown in Figure 4.19 shows the difference between the results of the implemented ANN model versus the average scores of the four physicians. Finally, in Table 4.29 the results of the present study are compared with those of similar works.

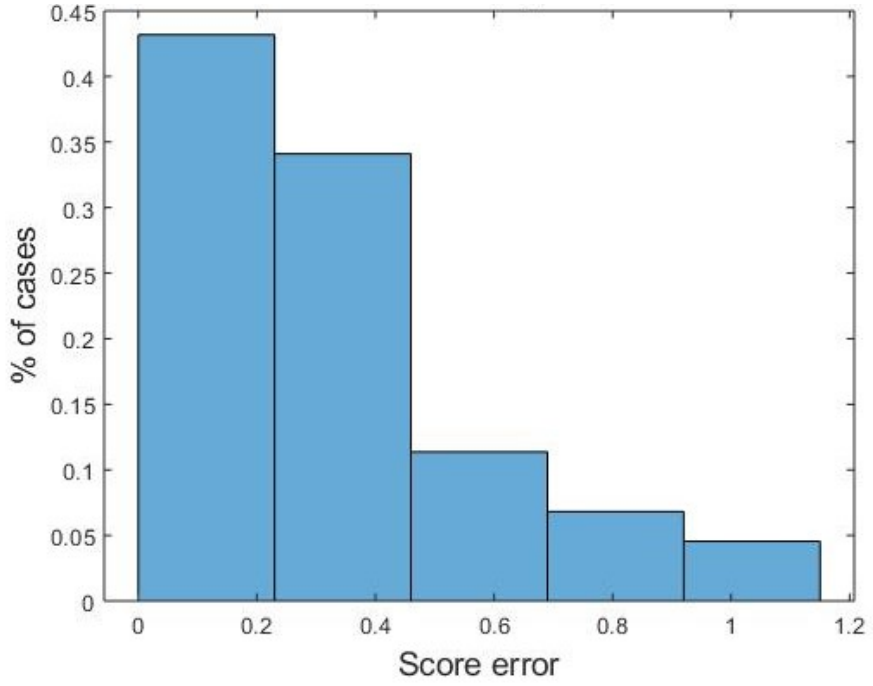


Fig. 4.18 Histogram of the distance between mean clinicians score and ANN outcomes. Continuous values were taken into account for assessment.

Table 4.29 Comparison with state-of-the-art algorithms. *Dedicated Application-performed evaluation. ** Values not explicitly reported within the text.

Study	Device	# subjects	# sensors	# raters	*r*	*RMSE*	*ICC*
[282]	IMU	42	5	3	0.79	0.46	–
[185]	IMU	34	3	3	0.79	–	–
[242]	Smartphone	14	1	0*	0.5	–	–
[188]	Smartphone	44	1	0*	–	–	0.88
[186]	IMU	19	2	3	0.83	0.53	0.89
[187]	IMU	50	2	2	n.r. **	n.r. **	n.r. **
Proposed	Smartphone	93	1	4	0.92	0.42	0.88

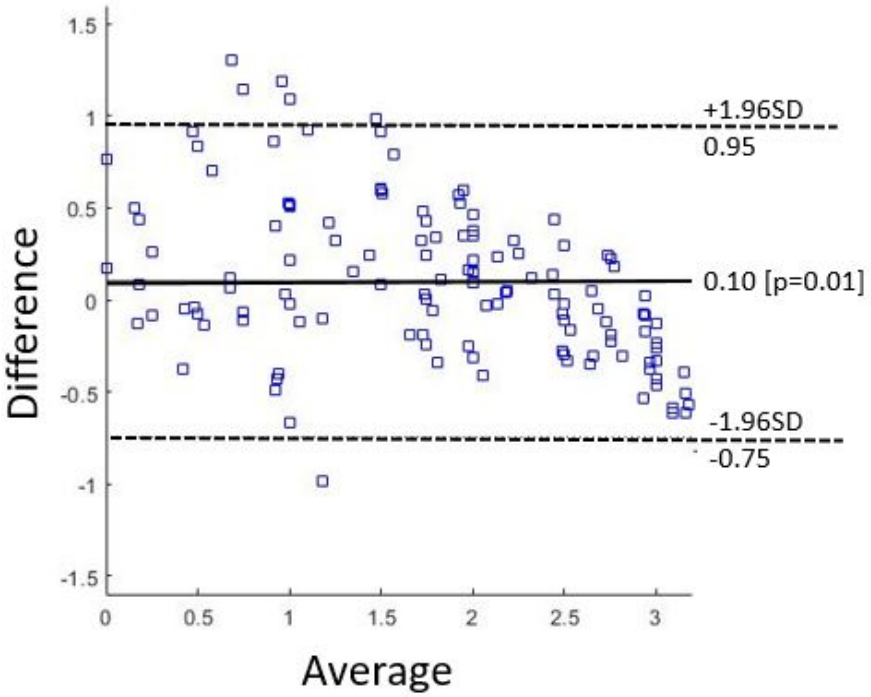
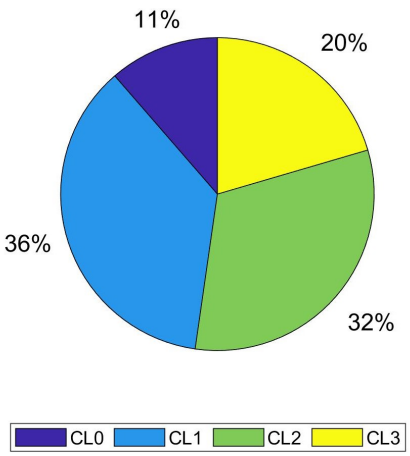


Fig. 4.19 Bland-Altman Plot of mean clinicians score and ANN outcomes.

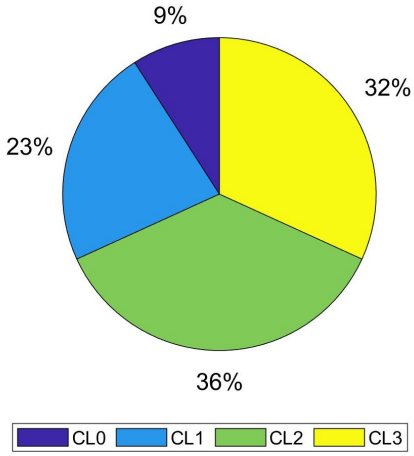
From Table 4.29 it can be observed that the present work employs the largest cohort of patients, as well as the largest number of clinicians; this ensures greater significance of the results. In addition, two of the three performance metrics (i.e., *r* and *RMSE*) outperform studies in the literature, while the *ICC* coefficient is in line with other studies. In particular, a higher Pearson coefficient ensures a higher correlation with the mean clinical score, and a lower *RMSE* is indicative of a better concentration of data around the line of best fit. Finally, it is worth noting that the correlation between the soft output and the mean clinical score is higher than the best agreement between clinicians (0.92 vs. 0.88). The issue of inter-rater variability is discussed in the next subsection. Currently, data acquisition was done during scheduled outpatient visits. Most patients had taken their usual dose of medication, although variable intervals of time had elapsed since then and the next scheduled dose was not forthcoming.

4.4.2 Inter-rater variability

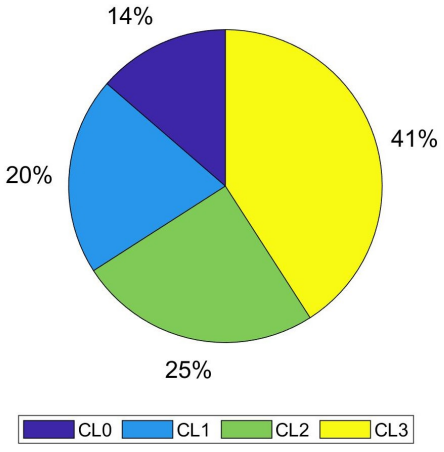
MDS-UPDRS assessment performed by multiple neurologists may not be homogeneous, mainly because of the difficulty in discriminating between adjacent classes in cases of intermediate severity. In this study, the inter-rater agreement index was between 0.74 and 0.88.



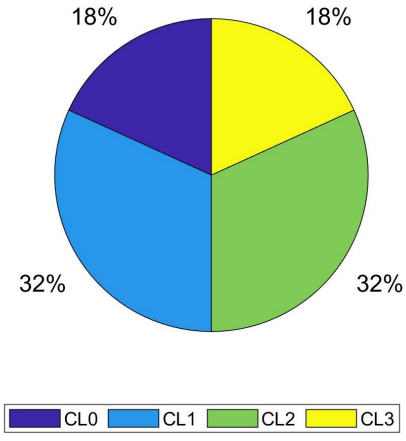
(a) Clinician 1



(b) Clinician 2



(c) Clinician 3



(d) Clinician 4

Fig. 4.20 For each evaluating clinician, score distribution among MDS-UPDRS part III Leg-Agility score (CLx stands for score x, x = LA score)

The possible disagreement in MDS-UPDRS scoring may be justified by the complexity of discriminating between adjacent classes, as physicians have to pay attention to different aspects (e.g., movement amplitude, velocity, and regularity). It is therefore worth asking whether differences between clinical and automatic scores are due to incorrect selection of features or are inherent in the data. This dilemma is also addressed in [99], where the authors speculate that automatic methods may prove more reliable than clinicians themselves for this reason. To this end, the agreement of neurologists by MDS-UPDRS class was analyzed. The pie charts shown in Figure 4.20 show the distribution of scores for each clinician. The correlation between the continuous ANN score and the average clinical score provided by four neurologists examining each patient directly or by videotaped leg movement was higher than the agreement among the best clinicians (0.92 vs. 0.88). A continuous score provided by an ML-based instrument can be used as a digital tool to overcome intra- and inter-rater variability.

Chapter 5

Discussion

The monitoring of PD is very complex, due to several factors. The severity of symptoms is assessed during the pre-scheduled outpatient visits, carried out once or twice a year. Moreover, inter-rater and intra-rater variability affect the evaluation of PD motor aspects, leading to poor objectivity. Finally, some motor signs of the disease (e.g., FOG) may not manifest during the clinical examination, and their fluctuation is commonly assessed only using subjective diaries.

In order to provide objective measures of motor dysfunction, a large variety of technologies have been used for motion analysis in PD. However, some of them (e.g., motion capture systems, instrumented walkways, and balance boards) represent costly and obtrusive instrumentation commonly used in laboratory settings. Such technology is capable of providing accurate movement analysis but it does not allow to perform continuous monitoring in free-living settings.

In this context, wearable inertial sensors represents an ecological solution for the collection of motion data in non-supervised environments. They are capable of accurately measuring human movement during daily life, providing a large amount of data that can be used in several applications, including human activity recognition, mobility assessment, and monitoring of PD.

However, collecting data using ecological solutions (i.e., smartwatch, smartphone) requires proper data analysis techniques, the complexity of which depends on a large number of factors, including the number and type of sensors, their position, the specific task, and the objective. To this end, signal processing and data mining, especially with the use of AI methodologies such as ML, have represented essential tools for the accurate analysis of wearable sensor data. ML and, more recently, DL

algorithms have been proposed to objectively and continuously evaluate PD motor symptoms, their fluctuations, and their response to therapy.

The combination of WMS and AI has been successfully employed for the assessment of PD. However, the cost, size, number, and location of sensors may result in costly and obtrusive solutions, and this limits the suitability of the designed systems for continuous real-life monitoring. Moreover, the clinical validity of the developed processing tools is often limited by the small sample size and/or the non-representative population enrolled in the study. Indeed, the heterogeneity of the movement patterns and symptoms manifestation in PD makes it challenging to design digital solutions that can be generalized to the entire PD population. Finally, information extraction is a critical process in the development of digital solutions for computer-aided diagnosis or monitoring. The identification of specific characteristics that fully describe movement and discriminate the different patterns in different populations is of pivotal importance, thus, it should be carefully and exhaustively performed.

In this study, a single smartphone was used to collect motion data in different settings. A single smartphone on the thigh was used to assess lower limb bradykinesia [98] and gait impairment [279, 216]. Such location resembles that of the smartphone placed in the front pocket, which may represent the usual position of smartphones in daily life. A smartphone on lower back was employed for the evaluation of postural stability [140], gait [176], and FOG [157, 287] in semi-supervised and unsupervised settings. In total, data were recorded from more than 200 PwPD during different activities (e.g., walk, turn, stand, postural transitions). The choice of simple ADLs is due to different factors. First, they are non-invasive and easy to perform by patients with different levels of motor impairment. Moreover, they can reflect the patient's mobility and the severity of motor signs. Finally, they are among the most performed activities in daily life, thus allowing the collection and analysis of a large amount of data. The developed algorithms were validated on a large patient cohort, and this reinforces the clinical validity of the obtained results. The experiments conducted in the present study were designed to answer different practical needs. First, the clinical need for more objective measures of symptom severity and disease progression. Second, the limitations of the current literature works, including the small patient cohort, the lack of classification tools, the performance of the system, and the complexity of the processing algorithms. Specifically, the main results and

contributions of this study are discussed in the following, together with their clinical utility, the comparison with similar literature works, and the limitations.

5.1 Freezing of gait

FOG detection and prediction have been addressed in [157, 219, 287]. A wide range of wearable solutions and a variety of processing algorithms have been proposed in the literature. However, few studies have addressed FOG prediction, and none of them have considered the effect of FOG on detection and prediction performance. Regarding FOG detection, the proposed algorithm [219] provided high performance in detecting FOG episodes, comparable to those previously described [288, 289]. Furthermore, in line with previous research [145], the developed algorithm detected FOG episodes in PwPD ON and OFF therapy with similar sensitivity, thus suggesting that L-Dopa does not significantly change FOG-related features, but only affects the frequency and duration of FOG episodes. Despite comparable sensitivity, the algorithm recognized FOG episodes with lower specificity in patients OFF than in those ON therapy. This could reflect an increased number of FPs due to the difficulty in differentiating abnormal spatial and temporal gait parameters, which are important in patients OFF therapy, from FOG events. Finally, training the algorithm in patients OFF therapy resulted in better performance in detecting FOG episodes in patients ON therapy than vice-versa. This likely reflects the higher frequency and duration of FOG episodes in OFF patients compared to ON patients, thus providing more data for training [145]. Regarding pre-FOG detection, performance was in line with recent works, confirming the possibility of predicting FOG in PwPD using ML algorithms [227, 228]. Pre-FOG window length crucially affected the accuracy of pre-FOG recognition, with a progressive decrease in performance as the window length increased. Consistent with previous studies [290], this result suggests that the degradation of the walking pattern that commonly precedes FOG episodes becomes increasingly evident as the FOG episode approaches. Therefore, to improve the overall accuracy of FOG prediction, windows of short duration before FOG (e.g., 2-3 s) should be used. Specifically, during pre-FOG periods, leg movement slows down (i.e., decrease in range, standard deviation, and maximum velocity), strides become faster and shorter (i.e., decrease in step, stride, and swing time), stride frequency content becomes more variable (i.e., decrease in power spectral entropy), and stride

frequency content shows a shift toward higher frequencies (i.e., decrease in low power frequency).

When assessing the effects of L-Dopa on ML performance, the accuracy in detecting pre-FOG was higher in OFF PwPD than in those on therapy, fully in line with the hypothesis that L-Dopa has an impact on FOG prediction, since the rhythm degradation pattern in ON therapy is partially corrected by L-Dopa itself. In further support of the influence of L-Dopa on the prediction of FOG, the classification algorithm recognized pre-FOG periods earlier in patients in OFF therapy than in those in ON therapy (6 *pm* 1.3 and 4 *pm* 1.1 steps before the onset of FOG, respectively). In addition, training the algorithm with data from patients ON and OFF therapy significantly changed the ability to detect pre-FOG. These results agree with the observation that L-Dopa improves spatio-temporal gait parameters outside FOG episodes [169, 145]. As a result, L-Dopa modifies the typical degradation pattern preceding FOG episodes, attenuating pre-FOG periods in patients ON therapy. When considering the present results, some limitations should be taken into account. First, although individuals with disabilities were selected based on strict clinical criteria, lending homogeneity to the cohort, the cohort consisted of a limited number of subjects. Consequently, to further increase the statistical significance of the results, future studies should enroll a larger sample of patients. In addition, the use of independent test sets would increase the validity of the results. The present study, like most of the works in the literature, made use of laboratory settings and supervised experiments for data acquisition, and this limits the generalizability of the results to free-living settings. Finally, as in most proposed FOG detection systems, the proposed approach relies on offline data processing and is not suitable for real-time applications.

To overcome these limitations, a real-time FOG detection system was proposed in [287]. Acceleration data from a large number of subjects (118 PD patients and 21 healthy elderly subjects) were included in the study, with more than a thousand FOG episodes registered. The analyses were performed on three different data sets, two of which were employed as independent test sets. This allowed testing the generalization capability of the detection algorithm when processing data including a large variety of gait patterns and activities. Data were collected in non-supervised conditions, during activities similar to those of daily living. The availability of labeled activities made it possible to get an insight into the context in which false positive events were detected. The detection algorithm was conceived for use in real-time. It employed

a single time window, with no pre-processing performed on the entire signal, and no information regarding past or future data. Moreover, computational complexity, testing time, memory requirements, and detection latency were carefully estimated. In-depth post-processing procedures were performed to provide a comprehensive analysis of the efficacy in detecting FOG and discarding other activities. The effect of activity thresholding on the performance of the detection system and on the percentage of discarded windows was evaluated. Actually, an effective tool for excluding data from the classification process would help reduce the computational burden and increase the battery life. Performance was found to be stable in the main data set, demonstrating the good generalization capability of the detection algorithm. When testing the model on the external data set, a reduction in sensitivity and an improvement in specificity were observed. This may be due to the different clinical characteristics and therapy conditions of patients in the two data sets, with the main corpus including subjects with more severe gait impairment and FOG manifestations. Moreover, results should be interpreted considering that data from the main and independent data sets were recorded using a different device positioned in a different location (left side of the waist and lower back, respectively). Testing the model on the external ADL data set, encompassing 59 PwPD performing several activities, resulted in very high specificity, with few FPs registered during common ADLs. Moreover, false FOG episodes were found to be short and located close to real FOG episodes, suggesting a degraded walking pattern before and/or after FOG. Finally, unitary specificity was obtained when testing the algorithm on elderly control subjects, proving the high capability of rejecting FPs from elderly subjects without PD. As far as concerns the FOG detection rate, 52.3% (32.5%) of FOG episodes were predicted 3.1s (1.3s) before their actual onset and 47.7% (50.0%) detected after 0.8s (1.1s), in the main (independent) test set. The results of a previous study [219] suggested that while it is possible to develop a robust classifier for the detection of FOG, an evident performance impairment was observed when implementing a FOG prediction system. A similar performance reduction was observed in [221]. Moreover, data were registered in laboratory settings in both studies, during pre-defined walking tasks. Data recorded in home environments in non-supervised conditions provide even more challenges for the prediction of FOG. However, this study demonstrated that implementing an algorithm for the timely detection of FOG provides the opportunity to predict up to 50% FOG episodes before their actual occurrence. Concerning the computation complexity and testing time, minimal

pre-processing, together with a memory requirement of less than 55 KB and testing time of 43 ms, makes the algorithm suitable for real-time implementation in a stand-alone device. Finally, the analysis performed on the activity threshold demonstrated that it is possible to reduce the computation burden of the detection algorithm without a significant performance impairment. However, the present work has some limitations. First, the REMPARK data set included most of the recorded FOG episodes, but the activity label was not available. On the other hand, the ADL data set included several annotations of the performed activities, but any FOG episodes were registered. A large data set including both a large number of patients and a large number of FOG episodes is necessary, together with a careful annotation of the most informative activities (e.g., gait, stand, sit, postural transitions). Moreover, raw input data was used in this study to limit the computation burden and reduce the processing time. However, this may not represent the solution providing the best performance. Novel implementations of time-frequency transforms may provide performance improvement without significantly increasing the computation burden. Finally, despite computational complexity, memory requirement, and testing time having been estimated, the developed detection algorithm has not yet been embedded in a stand-alone device for real-time use in home environments.

5.2 Postural stability

Postural stability represents one of the most significant symptoms of PD, increasing the risk of injuries. The clinical evaluation of postural stability is conducted following the MDS-UPDRS guidelines. Specifically, the item 3.12 - "Postural Stability" consists in a retropulsion test, where patients are pulled back by neurologists to assess their level of postural response. Such a task can be dangerous both for patients and for clinicians, as the lack of postural control may lead to falling backward. Moreover, this task does not represent an activity usually performed in daily living, and this precludes the possibility of continuous monitoring in real-life scenarios. In this study, [140], postural stability was evaluated during quiet stance, which represents a safe, easy-to-perform, and common activity of normal living. The results suggest that it is possible to provide objective measures of postural stability using data recorded during a different task, with respect to that used for the clinical evaluation. This aspect, along with the use of a single commercial smartphone, paves the way for

remote passive monitoring of PwPD in the home environment. However, the obtained results should be considered with caution, due to the use of a reduced data set (42 PwPD, 7 control subjects). Due to the cardinality of classes and the simple protocol selected, the data and algorithm were not adequate to perform a fine classification between adjacent classes. As for a possible implementation of the algorithm for early detection of postural impairment in PwPD, a much larger cohort of PwPD in the early stage of the disease, together with an age-matched control population should be employed. As for a finer classification between adjacent classes, the cardinality of each PS class has to be increased, in order to provide much more statistical meaningfulness to the results. Furthermore, due to the intrinsic intra- and inter-rater variability, the evaluation should be performed by several neurologists, in order to employ the average clinical evaluation as the ground truth of the classification task [98]. Finally, the implementation of the developed algorithm requires the automatic identification of the static upright position, in order to properly process the data.

5.3 Gait impairment

In this study, gait was analyzed to estimate gait impairment and postural stability in PD [176, 157, 279]. The walking pattern of PwPD was analyzed to predict the clinical axial impairment score [157, 279]. To maximize prediction performance, many time and frequency domain features were computed in addition to the classical spatio-temporal parameters routinely used in gait analysis studies. The prediction performance was higher than that reported in previous studies [215, 239]. In addition, whereas in [239] different tasks (e.g., gait, turn, and stance) were analyzed to provide the final output, in the present study only features extracted from the walking bouts were used to predict the PIDG score. Finally, in contrast to the only previous study that predicted PIGD scores in PD using three sensing devices [239], here a single wearable, small, and lightweight inertial sensor placed on the thigh was used, offering a low-cost and minimally invasive solution for daily applications in free-living contexts. Analyses were performed in PwPD both ON and OFF therapy, also considering the effect of FOG. A comprehensive statistical analysis of both clinical scores and engineered features was performed to provide more insights into the ability of the features to measure axial motor disability in PwPD. Significant correlations were found between specific sensor-based variables in the

time and frequency domains and PIGD scores, suggesting that higher PIGD scores are associated with greater kinematic abnormalities during walking. More specifically, greater axial motor impairment as measured by PIGD was associated with greater abnormalities in amplitude, intensity, and regularity of movement. Consistent with these findings, the authors of [239] found significant associations between PIGD scores and sensor-based measures, including the number of steps, gait speed, and sway area. These results also agree with previous studies that have shown greater impairment of spatio-temporal gait parameters in PwPD who have a PIGD phenotype with more severe axial dysfunction than those with a tremor-dominant phenotype [291, 292]. When considering the prediction of PIGD in relation to L-Dopa intake, the regression models performed better in PwPD in the OFF state than in the ON state ($p = 0.002$). This is in line with previous results reporting variable accuracy of ML algorithms in PwPD under different pharmacological conditions [219]. Indeed, L-Dopa significantly modifies the spatio-temporal parameters of gait and, consequently, affects the performance of ML in measuring gait in PD [219, 293]. Since the PIGD score is composed of several items reflecting postural and gait skills, one possible explanation for the different ML performances in OFF and ON patients is based on the heterogeneous L-Dopa sensitivity of balance and gait. In fact, unlike gait, L-Dopa usually does not have a substantial impact on balance in PD [294–296]. Therefore, PIGD scores are more accurately predicted in patients in the OFF state than in those in the ON state, due to a more similar pattern of postural and walking abilities in patients not on dopaminergic therapy. Different time- and frequency-domain features were found to have different sensitivity in PwPD with and without FOG, a finding fully in agreement with previous studies that have shown worse continuous gait abnormalities in PwPD with FOG than those without FOG, even outside episodes of FOG [297, 298, 145]. In addition, the different time and frequency domain characteristics explain another relevant finding of this study, namely that models trained on patients without therapy do not perform well in patients on therapy and vice versa. It is likely that in addition to changes in continuous gait parameters and pharmacological conditions, the unpredictable and sudden onset of FOG affects the walking pattern in PwPD, worsening ML performance. Furthermore, despite its direct impact on walking, FOG is not included in the calculation of the PIGD score and, consequently, is not considered for the assessment of axial impairment when using this standardized clinical index. The results suggest that it is possible to implement a single regression model that can predict PIGD in PwPD, regardless of

therapeutic condition and the presence of FOG. However, data should be collected from a heterogeneous cohort of PwPD under different pharmacological conditions. When models were trained on a subgroup of patients (e.g., patients with FOG, patients ON therapy), impaired performance was observed when tests were performed on a different subgroup. Future studies are needed to clarify the technical feasibility of applying the proposed ML algorithms to data recorded using smartphones in unsupervised environments. In addition, when considering the results of this study, the lack of validation on an independent test set is a possible limitation to be taken into account [240]. Consequently, further studies are needed to reproduce these results in larger cohorts of patients.

5.4 Bradykinesia

When implementing an automatic tool for the evaluation of bradykinesia [98], results suggested that a single smartphone can be used to provide objective measures of movement velocity and regularity in a simple activity as the leg agility task. The proposed method outperformed the related literature studies [185, 241, 242, 188, 186], despite the use of a single commercial smartphone for the data acquisition task. Moreover, the clinical scores were obtained from four different raters, allowing to compute the inter-rater agreement and providing a more robust ground truth. When considering the continuous ML model output and the average clinical score (instead of the discrete score provided by the MDS-UPDRS), the model performance further improved, suggesting that a finer classification would be more suitable for assessing small changes in patients' motor performance. Finally, the correlation between the automatic scoring system and the average clinical score was larger than the best inter-rater agreement. This suggests that, for very specific motor tasks, computer-aided scoring systems may provide a more objective and precise evaluation of motor performance. This result can be justified by the difficulty of discriminating between adjacent classes, given that clinicians are required to pay attention to several different aspects of movement, including frequency, velocity, amplitude, and regularity. Given that the MDS-UPDRS does not provide any suggestion on the relative importance of one aspect over the others, different neurologists can give different weights to these different aspects, thus promoting one over the others (e.g., amplitude over velocity,

regularity over amplitude). Instead, the ML algorithm trained to predict the average score of many raters results less biased, providing a more objective evaluation.

5.5 Limitations

Overall, the results obtained in the present study provide good opportunities for the remote monitoring of PD in real-life scenarios. However, while the algorithms for FOG detection and turn analysis do not require any prior knowledge of the activity context, the automatic evaluation of bradykinesia and postural stability (either during stance or walking) requires the proper detection of such activities. To this end, an accurate HAR algorithm should be implemented, in order to identify the different activities to be analyzed by the developed prediction models. Actually, many HAR systems have been proposed with the use of inertial sensors [189], and even using smartphones [199, 136]. However, most of the studies performed the analysis on data from healthy young controls, which may not be as accurate when applied to data from elderly people [299]. Considering the high heterogeneity of movement patterns produced by PwPD, it turns clear that HAR algorithms should be trained and validated on a large number of PwPD in real-life settings. Moreover, besides walking, turning, and static positions, postural transitions (e.g., sit to stand, stand to sit) represents common ADLs, which were not analyzed in this study. However, the analysis of such activities using WS could bring important information regarding mobility and postural stability [300] in daily life, especially in PwPD [232, 301]. Moreover, postural transitions can be detected and analyzed using a single sensor on the lower back [302, 301], in line with the methods of the present study [140, 176, 157, 287].

The major limitation of this study is represented by the offline processing performed in all studies. Specifically, while the data acquisition procedures were conducted either using smartphones or IMUs, the data processing tasks were performed on the personal computer. Despite most of the developed works can be easily implemented in stand-alone devices, due to their low memory requirement and fast processing time [287, 176, 140], the data processing and transmission capabilities of the employed hardware were not fully exploited. This latter aspect is fundamental for the development of a stand-alone wearable solution for the passive monitoring of PD in daily life. Moreover, while some processing tools [287, 176]

are ready to passively monitor PD symptoms in real-life settings, the generalization capability of other digital tools [98, 140, 216] should be validated in unsupervised environments. Finally, the ultimate objective of this work is the generation of a *daily report* of patients' mobility, symptoms severity, and postural stability. To this end, the extracted measures should be summarized into a few indices representing the patients' motor performance. These metrics should be robust and significant, in order to provide clinicians with accurate information regarding disease progression, response to therapy, and risk of falls. In this context, the information gathered from the analysis of different ADLs (i.e., walk, turn, stance, postural transitions) should be combined and summarized to provide an overall estimation of the motor and postural impairment.

Chapter 6

Conclusion and future works

In this study, the combination of wearable sensors and ML algorithms was used for assessing different motor aspects of PD. Simple experimental procedures and unobtrusive technologies were used for data acquisition, while exhaustive information extraction, classification, and prediction tasks were addressed by properly exploiting the potentiality of signal processing, ML, and DL methods. Data were recorded from more than 200 PD subjects, in different settings, and during different activities. The results suggest that a single inertial sensor is capable of providing sufficient information for assessing the presence and severity of motor symptoms. Overall, this study can provide useful information regarding wearable sensors and processing algorithms for assessing PD.

The very recent development of smart textiles [303] provides a new opportunity for a continuous and less invasive large-scale monitoring of physical activity and physiological parameters. Indeed, wearable sensors embedded in everyday clothing increase the subjects' comfort and compliance, fundamental for the long-term monitoring of chronic diseases.

Future studies will be in the direction of an even more ecological data collection procedure, possibly using smart clothes. An accurate and efficient activity recognition will represent the basis for both the collection of mobility information and for triggering the different algorithms developed in this study. Experiments will be conducted in non-supervised free-living conditions.

Finally, the clinical validity and usefulness of the tools developed (i.e. hardware and software) should be carefully analyzed. To this end, *Goldsack et al.* [304]

provides guidance on determining whether a metric derived from a digital medicine product is reliable and, by extension, whether that digital medicine product is fit for purpose. The proposed approach (termed V3) involves several steps, including verification, analytical validation, and clinical validation. The former involves the evaluation of sensor results at the sample level and is usually the responsibility of the manufacturer. Analytical validation evaluates the data processing algorithms that convert sample-level sensor measurements into physiological metrics, and is the responsibility of the researchers or developers. Clinical validation serves to demonstrate that the digital medicine product acceptably identifies, measures or predicts clinical, biological, physical, functional status or experience in the defined context of use (which includes population definition). The clinical utility ensures that the use of the digital medicine product will lead to improved health outcomes or provide useful information for the diagnosis, treatment, management or prevention of a disease. Other criteria, such as cost, accessibility, compatibility, burden, and ease of use, should be considered for determining fitness for purpose.

Institutional Review Board Statement. The studies have been reviewed and approved by the Intercompany Ethics Committee A.O.U. Città della Salute e della Scienza di Torino- A.O. Ordine Mauriziano di Torino - A.S.L. Città di Torino, procols number 1534–2019 and 0041378–2021. The research has been designed in accordance with the principles of the Declaration of Helsinki and all applicable regulations. The research protocol, informed consent form, and any relevant materials have been reviewed and approved by the Ethics Committee.

Informed Consent Statement. Participants received detailed information on the purpose of the study and execution, and written informed consent was obtained from all subjects involved in the study.

Data Availability Statement. The data that support the findings of this research are available upon request from the corresponding author. Interested researchers may request access to the data by contacting the corresponding author at luigi.borzi@polito.it.

Conflicts of Interest. The author declares no conflict of interest.

Funding. This research received no external funding.

List of Figures

List of Tables

References

[1] L. Hirsch, N. Jette, A. Frolkis, T. Steeves, and T. Pringsheim. The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Neuroepidemiology*, 46(4):292–300, 2016. doi: 10.1159/000445751.

[2] J. Jankovic. Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery and Psychiatry*, 79(4):368–376, 2008. doi: 10.1136/jnnp.2007.131045.

[3] J.M. Savitt, V.L. Dawson, and T.M. Dawson. Diagnosis and treatment of Parkinson disease: molecules to medicine. *The Journal of clinical investigation*, 116(7):1744–54, 2006. doi: 10.1172/JCI29178.

[4] J.G.G. Hou and E.C. Lai. Non-motor Symptoms of Parkinson's Disease. *International Journal of Gerontology*, 1(2):53–64, 2007. doi: 10.1016/S1873-9598(08)70024-3.

[5] C.A. Davie. A review of Parkinson's disease. *British Medical Bulletin*, 86(1):109–127, 2008. doi: 10.1093/bmb/ldn013.

[6] A. Samà, C. Pérez-López, D. Rodríguez-Martín, A. Català, J.M. Moreno-Aróstegui, et al. Estimating bradykinesia severity in Parkinson's disease by analysing gait through a waist-worn sensor. *Computers in Biology and Medicine*, 84:114–123, 2017. doi: 10.1016/j.compbiomed.2017.03.020.

[7] A.E. Gomez Ayala. Enfermedad de pakrinson: abordaje terapéutico y farmacológico. *Offarm: farmacia y sociedad*, 26(5):70–78, 2007.

[8] J.A. Driver, G. Logroscino, J.M. Gaziano, and T. Kurth. Incidence and remaining lifetime risk of Parkinson disease in advanced age. *Neurology*, 72 (5):432–438, 2009. doi: 10.1212/01.wnl.0000341769.50075.bb.

[9] W. Dauer and S. Przedborski. Parkinson's disease: Mechanisms and models. *Neuron*, 39(6):889–909, 2003. doi: 10.1016/S0896-6273(03)00568-3.

[10] S.K. Van Den Eeden, C.M. Tanner, A.L. Bernstein, R.D. Fross, A. Leimpeter, et al. Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity. *American Journal of Epidemiology*, 157(11):1015–1022, 2003. doi: 10.1093/aje/kwg068.

[11] A. Elbaz, J.H. Bower, D.M. Maraganore, S.K. McDonnell, B.J. Peterson, et al. Risk tables for parkinsonism and Parkinson's disease. *Journal of Clinical Epidemiology*, 55(1):25–31, 2002. doi: 10.1016/S0895-4356(01)00425-5.

[12] R. Savica, B.R. Grossardt, J.H. Bower, J.E. Ahlskog, and W.A. Rocca. Time Trends in the Incidence of Parkinson Disease. *JAMA Neurol*, 73(8):981–989, 2016. doi: 10.1001/jamaneurol.2016.0947.

[13] L. Breydo, J.W. Wu, and V.N. Uversky. α-Synuclein misfolding and Parkinson's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1822(2):261–285, 2012. doi: 10.1016/j.bbadis.2011.10.002.

[14] K.F. Winklhofer and C. Haass. Mitochondrial dysfunction in Parkinson's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1802 (1):29–44, 2010. doi: 10.1016/j.bbadis.2009.08.013.

[15] K.K.K. Chung, Y. Zhang, K.L. Lim, Y. Tanaka, H. Huang, et al. Parkin ubiquitinates the α-synuclein-interacting protein, synphilin-1: Implications for Lewy-body formation in Parkinson disease. *Nature Medicine*, 7(10): 1144–1150, 2001. doi: 10.1038/nm1001-1144.

[16] Lucinda Hampton. Parkinsonism, 2022. URL https://www.physio-pedia.com/Parkinsonism.

[17] K.R. Chaudhuri, D.G. Healy, and A.H.V. Schapira. Non-motor symptoms of Parkinson's disease: Diagnosis and management. *Lancet Neurology*, 5(3): 235–245, 2006. doi: 10.1016/S1474-4422(06)70373-8.

[18] R.C. Helmich, M. Hallett, G. Deuschl, I Toni, and B.R. Bloem. Cerebral causes and consequences of parkinsonian resting tremor: A tale of two circuits? *Brain*, 135(11):3206–3226, 2012. doi: 10.1093/brain/aws023.

[19] G. Deuschl, P. Bain, and M. Brin. Consensus Statement of the Movement Disorder Society on Tremor. *Movement Disorders*, 13(S3):2–23, 2008. doi: 10.1002/mds.870131303.

[20] J. Jankovic, K.S. Schwartz, and W. Ondo. Re-emergent tremor of Parkinson's disease. *J Neurol Neurosurg Psychiatry*, 67(5):646, 1999. doi: 10.1136/jnnp.67.5.646.

[21] M.M. Hoehn and M.D. Yahr. Parkinsonism : onset, progression, and mortality Parkinsonism: onset, progression, and mortality. *Neurology*, 17(5):427–442, 1967. doi: 10.1212/WNL.17.5.427.

[22] N. Baradaran, S.N. Tan, A. Liu, A. Ashoori, S.J. Palmer, et al. Parkinson's disease rigidity: Relation to brain connectivity and motor performance. *Frontiers in Neurology*, 4:67, 2013. doi: 10.3389/fneur.2013.00067.

[23] M.C. Rodriguez-Oroz, M. Jahanshahi, P. Krack, I. Litvan, R. Macias, et al. Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *The Lancet Neurology*, 8(12):1128–1139, 2009. doi: 10.1016/S1474-4422(09)70293-5.

[24] C.J. Andrews, D. Burke, and J.W. Lance. The response to muscle stretch and shortening in parkinsonian rigidity. *Brain*, 95(4):795–812, 1972. doi: 10.1093/brain/95.4.795.

[25] 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. doi: 10.1136/jnnp.46.1.35.

[26] F.J.G. Vingerhoets, M. Schulzer, D.B. Calne, and B.J. Snow. Which clinical sign of Parkinson's disease best reflects the nigrostriatal lesion? *Annals of Neurology*, 41(1):58–64, 1997. doi: 10.1002/ana.410410111.

[27] A. Berardelli. Pathophysiology of bradykinesia in Parkinson's disease. *Brain*, 124(11):2131–2146, 2001. doi: 10.1093/brain/124.11.2131.

[28] E.V. Evarts, H. Teräväinen, and D.B. Calne. Reaction time in Parkinson's disease. *Brain*, 104(1):167–186, 1981. doi: 10.1093/brain/104.1.167.

[29] C.G. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, et al. 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. doi: 10.1002/mds.21198.

[30] H.M. Bronte-Stewart. Postural instability in idiopathic Parkinson's disease: the role of medication and unilateral pallidotomy. *Brain*, 125(9):2100–2114, 2002. doi: 10.1093/brain/awf207.

[31] D.R. Williams, H.C. Watt, and A.J. Lees. Predictors of falls and fractures in bradykinetic rigid syndromes: A retrospective study. *Journal of Neurology, Neurosurgery and Psychiatry*, 77(4):468–473, 2006. doi: 10.1136/jnnp.2005.074070.

[32] K.M. Doherty, B.P. van de Warrenburg, M.C. Peralta, L. Silveira-Moriyama, J.P. Azulay, et al. Postural deformities in Parkinson's disease. *The Lancet Neurology*, 10(6):538–549, 2011. doi: 10.1016/S1474-4422(11)70067-9.

[33] A. Castrioto, C. Piscicelli, D. Pérennou, P. Krack, and B. Debû. The pathogenesis of Pisa syndrome in Parkinson's disease. *Movement Disorders*, 29(9):1100–1107, 2014. doi: 10.1002/mds.25925.

[34] P. Martínez-Martín, A. Gil-Nagel, L. M. Gracia, J. B. Gómez, J. Martínez-Sarriés, et al. Unified Parkinson's disease rating scale characteristics and structure. *Movement Disorders*, 9(1):76–83, 1994. doi: 10.1002/mds.870090112.

[35] C. Maurer, T. Mergner, J. Xie, M. Faist, P. Pollak, et al. Effect of chronic bilateral subthalamic nucleus (STN) stimulation on postural control in Parkinson's disease. *Brain*, 126(5):1146–1163, 2003. doi: 10.1093/brain/awg100.

[36] A. Weiss, T. Herman, N. Giladi, and J.M. Hausdorff. New evidence for gait abnormalities among Parkinson's disease patients who suffer from freezing of gait: insights using a body-fixed sensor worn for 3 days. *Journal of Neural Transmission*, 122(3):403–410, 2015. doi: 10.1007/s00702-014-1279-y.

[37] J.G. Nutt, B.R. Bloem, N. Giladi, M. Hallett, F.B. Horak, et al. Freezing of gait: Moving forward on a mysterious clinical phenomenon. *The Lancet Neurology*, 10(8):734–744, 2011. doi: 10.1016/S1474-4422(11)70143-0.

[38] J.D. Schaafsma, Y. Balash, T. Gurevich, A.L. Bartels, J.M. Hausdorff, et al. 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. doi: 10.1046/j.1468-1331.2003.00611.x.

[39] S. Mazilu, A. Calatroni, E. Gazit, D. Roggen, J.M. Hausdorff, et al. Feature learning for detection and prediction of freezing of gait in Parkinson's disease. *Machine Learning and Data Mining in Pattern Recognition. MLDM 2013. Lecture Notes in Computer Science*, 7988:144, 2013. doi: 10.1007/978-3-642-39712-7_11.

[40] F. Demrozi, V. Bragoi, F. Tramarin, and G. Pravadelli. An indoor localization system to detect areas causing the freezing of gait in Parkinsonians. In *Design, Automation & Test in Europe Conference Exhibition (DATE)*, pages 952–955, 2019. doi: 10.23919/DATE.2019.8715093.

[41] N. Giladi, T.A. Treves, E.S. Simon, H. Shabtai, Y. Orlov, et al. Freezing of gait in patients with advanced Parkinson's disease. *Journal of Neural Transmission*, 108(1):53–61, 2001. doi: 10.1007/s007020170096.

[42] N. Giladi, R. Kao, and S. Fahn. Freezing phenomenon in patients with parkinsonian syndromes. *Movement Disorders*, 12(3):302–305, 1997. doi: 10.1002/mds.870120307.

[43] B.R. Bloem, J.M. Hausdorff, J.E. Visser, and N. Giladi. Falls and freezing of Gait in Parkinson's disease: A review of two interconnected, episodic phenomena. *Movement Disorders*, 19(8):871–884, 2004. doi: 10.1002/mds.20115.

[44] N. Giladi and J.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. doi: 10.1016/j.jns.2006.05.015.

[45] N. Giladi, M.P. McDermott, S. Fahn, S. Przedborski, J. Jankovic, et al. Freezing of gait in PD: Prospective assessment in the DATATOP cohort. *Neurology*, 56(12):1712–1721, 2001. doi: 10.1212/WNL.56.12.1712.

[46] J. Nonnekes, A.H. Snijders, J.G. Nutt, G. Deuschl, N. Giladi, et al. Freezing of gait: A practical approach to management. *The Lancet Neurology*, 14(7): 768–778, 2015. doi: 10.1016/S1474-4422(15)00041-1.

[47] D. Bäckström, G. Granåsen, M.E. Domellöf, J. Linder, and S.J. andothers Mo. Early predictors of mortality in parkinsonism and Parkinson's disease A population-based study. *Neurology*, 91(22):E2045–E2056, 2018. doi: 10.1212/WNL.0000000000006576.

[48] B. Habermann-Little. An analysis of the prevalence and etiology of depression in Parkinson's disease. *The Journal of neuroscience nursing: journal of the American Association of Neuroscience Nurses*, 23(3):165–169, 1991. doi: 10.1097/01376517-199106000-00006.

[49] P. Hantz, G. Caradoc-Davies, T. Caradoc-Davies, M. Weatherall, and G. Dixon. Depression in Parkinson's disease. *Am J Psychiatry*, 151(7):1010–1014, 1994. doi: 10.1176/ajp.151.7.1010.

[50] K. Walsh and G. Bennett. Parkinson's disease and anxiety. *Postgraduate Medical Journal*, 77(904):89–93, 2001. doi: 10.1136/pmj.77.904.89.

[51] Matthew A. Menza, Jacob Sage, Eric Marshall, Ronald Cody, and Roger Duvoisin. Mood changes and "on-off" phenomena in Parkinson's disease. *Movement Disorders*, 5(2):148–151, 1990. doi: 10.1002/mds.870050210.

[52] J.L.W. Bosboom, D. Stoffers, and E.C. Wolters. Cognitive dysfunction and dementia in Parkinson?s disease. *Journal of Neural Transmission*, 111(10): 1303–1315, 2004. doi: 10.1007/s00702-004-0168-1.

[53] N.J. Diederich, C.G. Goetz, and G.T. Stebbins. Repeated visual hallucinations in Parkinson's disease as disturbed external/internal perceptions: Focused review and a new integrative model. *Movement Disorders*, 20(2):130–140, 2005. doi: 10.1002/mds.20308.

[54] I. Rechichi, M. Zibetti, L. Borzì, G. Olmo, and L. Lopiano. Single-channel EEG classification of sleep stages based on REM microstructure. *Healthcare technology letters*, 8:58–65, 2021. doi: 10.1049/htl2.12007.

[55] Mark Stacy. Sleep disorders in Parkinson's disease: epidemiology and management. *Drugs & aging*, 19(10):733–739, 2002. doi: 10.2165/00002512-200219100-00002.

[56] David S. Goldstein. Orthostatic hypotension as an early finding in Parkinson's disease. 16(1):46–54, 2006. doi: 10.1007/s10286-006-0317-8.

[57] K. Kujawa, S. Leurgans, R. Raman, L. Blasucci, and C.G. Goetz. Acute orthostatic hypotension when starting dopamine agonists in Parkinson's disease. *Archives of neurology*, 57(10):1461–3, 2000. doi: 10.1001/archneur.57.10.1461.

[58] F. Amato, L. Borzì, G. Olmo, et al. An algorithm for Parkinson's disease speech classification based on isolated words analysis. *Health information science and systems*, 9:32, 2022. doi: 10.1007/s13755-021-00162-8.

[59] F. Amato, L. Borzì, G. Olmo, C.A. Artusi, G. Imbalzano, et al. Speech Impairment in Parkinson's Disease: Acoustic Analysis of Unvoiced Consonants in Italian Native Speakers. *IEEE Access*, 9:166370–166381, 2021. doi: 10.1109/ACCESS.2021.3135626.

[60] N.L.L Khan, E. Graham, P. Critchley, A.E. Schrag, N.W. Wood, et al. Parkin disease: A phenotypic study of a large case series. *Brain*, 126(6):1279–1292, 2003. doi: 10.1093/brain/awg142.

[61] R.F. Pfeiffer. Gastrointestinal dysfunction in Parkinson's disease. *Lancet Neurology*, 2(2):107–116, 2003. doi: 10.1016/S1474-4422(03)00307-7.

[62] J. Kaye, H. Gage, A. Kimber, L. Storey, and P. Trend. Excess burden of constipation in Parkinson's disease: A pilot study. *Movement Disorders*, 21 (8):1270–1273, 2006. doi: 10.1002/mds.20942.

[63] K. Winge, A.M. Skau, H. Stimpel, K.K. Nielsen, and L. Werdelin. Prevalence of bladder dysfunction in Parkinsons disease. *Neurourology and urodynamics*, 25(2):116–122, 2006. doi: 10.1002/nau.20193.

[64] A. Kilzheimer, T. Hentrich, and S. Burkhardt. The challenge and opportunity to diagnose parkinson's disease in midlife. *Frontiers in Neurology*, 10:1328, 2019. doi: 10.3389/fneur.2019.01328.

[65] D.J. Gelb, E. Oliver, and S. Gilman. Diagnostic criteria for parkinson disease. *Arch Neurol.*, 56(1):33–39, 1999. doi: 10.1001/archneur.56.1.33.

[66] L.V. Kalia and A.E. Lang. Parkinson's disease. *Lancet*, 386(9996):896–912, 2015. doi: 10.1016/S0140-6736(14)61393-3.

[67] M.J. Armstrong and M.S. Okun. Diagnosis and treatment of parkinson disease: A review. *JAMA*, 323(6):548–560, 2020. doi: 10.1001/jama.2019.22360.

[68] C.G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G.T. Stebbins, et al. Movement disorder society task force report on the hoehn and yahr staging scale: status and recommendations. *Mov Disord.*, 19(9):1020–1028, 2004. doi: 10.1002/mds.20213.

[69] Petrie Raddock. Parkinson's.....why wait? URL https://www.connectneurophysiotherapy.com/parkinsons-disease/parkinsons-why-wait/.

[70] C. Ramaker, J. Marinus, A.M. Stiggelbout, and B.J. van Hilten. Systematic evaluation of rating scales for impairment and disability in parkinson's disease. *Movement Disorders*, 17(5):867–876, 2002. doi: 10.1002/mds.10248.

[71] C.C. Goetz. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders*, 18(7):738–750, 2003. doi: 10.1002/mds.10473.

[72] A. Mittur, S. Gupta, and N.B. Modi. Pharmacokinetics of Rytary, An Extended-Release Capsule Formulation of Carbidopa–Levodopa. *Clinical Pharmacokinetics*, 56(9):999–1014, 2017. doi: 10.1007/s40262-017-0511-y.

[73] M.A. Hely, J.G.L. Morris, W.G.J. Reid, and R. Trafficante. Sydney Multicenter Study of Parkinson's disease: Non-L-dopa-responsive problems dominate at 15 years. *Movement Disorders*, 20(2):190–199, 2005. doi: 10.1002/mds.20324.

[74] S. Fahn. Parkinson disease, the effect of levodopa, and the ELLDOPA trial. *Archives of Neurology*, 56(5):529–535, 1999. doi: 10.1001/archneur.56.5.529.

[75] Sethi K. Olanow C.W., Stern M.B. The scientific and clinical basis for the treatment of Parkinson's disease. *Neurology*, 72(21):S1–136, 2009. doi: 10.1212/WNL.0b013e3181a1d44c.

[76] T.N. Chase, M.M. Mouradian, and T.M. Engber. Motor response complications and the function of striatal efferent systems. *Neurology*, 43(12):S23–7, 1993.

[77] B.A. Wright and C.H. Waters. Continuous dopaminergic delivery to minimize motor complications in Parkinson's disease. *Expert Review of Neurotherapeutics*, 13(6):719–729, 2013. doi: 10.1586/ern.13.47.

[78] A. Schrag and N. Quinn. Dyskinesias and motor fluctuations in Parkinson's disease. *Brain*, 123(11):2297–2305, 2000. doi: 10.1093/brain/123.11.2297.

[79] Anthony H. V. Schapira. Treatment options in the modern management of Parkinson disease. *Archives of Neurology*, 64(8):1083–1088, 2007. doi: 10.1001/archneur.64.8.1083.

[80] N.J. Ives, R.L. Stowe, J. Marro, C. Counsell, A. Macleod, et al. 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. doi: 10.1136/bmj.38184.606169.AE.

[81] Parkinson Study Group. Pramipexole vs Levodopa as Initial Treatment for Parkinson Disease: A Randomized Controlled Trial. *Jama*, 284(15):1931–1938, 2000. doi: 10.1001/jama.284.15.1931.

[82] H. Braak, K. Del Tredici, U. Rüb, R.A.I. De Vos, E.N.H. Jansen Steur, et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*, 24(2):197–211, 2003. doi: 10.1016/S0197-4580(02)00065-9.

[83] E. Faggiani and A. Benazzouz. Deep brain stimulation of the subthalamic nucleus in Parkinson's disease: From history to the interaction with the monoaminergic systems. *Progress in Neurobiology*, 151:139–156, 2017. doi: 10.1016/j.pneurobio.2016.07.003.

[84] A. Kaelin-Lang and A. Stibal. Chirurgische therapie des morbus Parkinson. *Therapeutische Umschau*, 64(1):21–27, 2007. doi: 10.1024/0040-5930.64.1.21.

[85] M.S. Okun, H.H. Fernandez, S.S. Wu, L. Kirsch-Darrow, D. Bowers, et al. Cognition and mood in Parkinson's disease in subthalamic nucleus versus globus pallidus interna deep brain stimulation: The COMPARE trial. *Annals of Neurology*, 65(5):586–595, 2009. doi: 10.1002/ana.21596.

[86] T. Wichmann and M.R. DeLong. Deep Brain Stimulation for Movement Disorders of Basal Ganglia Origin: Restoring Function or Functionality? *Neurotherapeutics*, 13(2):264–283, 2016. doi: 10.1007/s13311-016-0426-6.

[87] M.S. Okun. Deep-brain stimulation for Parkinson's disease. *N Engl J Med.*, 367(16):1529–38, 2012. doi: 10.1056/NEJMct1208070.

[88] R. Di Marco, F. Pistonesi, V. Cianci, R. Biundo, L. Weis, et al. Effect of Intensive Rehabilitation Program in Thermal Water on a Group of People with Parkinson's Disease: A Retrospective Longitudinal Study. *Healthcare*, 10(2):368, 2022. doi: 10.3390/healthcare10020368.

[89] F. Magrinelli, A. Picelli, P. Tocco, A. Federico, L. Roncari, et al. Pathophysiology of Motor Dysfunction in Parkinson's Disease as the Rationale for Drug Treatment and Rehabilitation. *Parkinson's Disease*, 2016:9832839, 2016. doi: 10.1155/2016/9832839.

[90] D. Goldberg. Epidemiology of mental disorders in primary care settings. *Epidemiol Rev*, 17:182–190, 1995. doi: 10.1093/oxfordjournals.epirev.a036174.

[91] D.P. Breen, J.R. Evans, K. Farrell, et al. Determinants of delayed diagnosis in parkinson's disease. *J Neurol*, 260:1978—-1981, 2013. doi: 10.1007/s00415-013-6905-3.

[92] C. Barthel, E. Mallia, B. Debû, B.R. Bloem, and M.U. Ferraye. The Practicalities of Assessing Freezing of Gait. *Journal of Parkinson's Disease*, 6(4):667–674, 2016. doi: 10.3233/JPD-160927.

[93] E. Heremans, A. Nieuwboer, J. Spildooren, J. Vandenbossche, N. Deroost, et al. Cognitive aspects of freezing of gait in Parkinson's disease: A challenge for rehabilitation. *Journal of Neural Transmission*, 120(4):543–557, 2013. doi: 10.1007/s00702-012-0964-y.

[94] K.A. Ehgoetz Martens, E.L. Lukasik, M.J. Georgiades, M. Gilat, J.M. Hall, et al. Predicting the onset of freezing of gait: A longitudinal study. *Movement Disorders*, 33(1):128–135, 2018. doi: 10.1002/mds.27208.

[95] J. Klucken, R. Krüger, P. Schmidt, and B.R. Bloem. Management of Parkinson's Disease 20 Years from Now: Towards Digital Health Pathways. *Journal of Parkinsons's Disease*, 8(1):S85–S94, 2018. doi: 10.3233/JPD-181519.

[96] N. Giladi, J. Tal, T. Azulay, O. Rascol, D.J. Brooks, et al. Validation of the Freezing of Gait Questionnaire in patients with Parkinson's disease. *Movement Disorders*, 24(5):655–661, 2009. doi: 10.1002/mds.21745.

[97] A. Nieuwboer, L. Rochester, T. Herman, W. Vandenberghe, G.E. Emil, et al. 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, 2009. doi: 10.1016/j.gaitpost.2009.07.108.

[98] L. Borzì, M. Varrecchia, S. Sibille, G. Olmo, C.A. Artusi, et al. Smartphone-Based Estimation of Item 3.8 of the MDS-UPDRS-III for Assessing Leg Agility in People With Parkinson's Disease. *IEEE Open Journal of Engineering in Medicine and Biology*, 1:140–147, 2020. doi: 10.1109/ojemb.2020.2993463.

[99] A.J Espay, J.M. Hausdorff, A. Sanchez-Ferro, J. Klucken, A. Merola, et al. A Roadmap for Implementation of Patient-Centered Digital Outcome Measures in Parkinson's disease Obtained Using Mobile Health Technologies. 34(5): 657–663, 2019. doi: 10.1002/mds.27671.

[100] C Mazzà, L. Alcock, K. Aminian, C. Becker, S. Bertuletti, et al. Technical validation of real-world monitoring of gait: a multicentric observational study. *BMJ Open*, 11:e050785, 2022. doi: 10.1136/bmjopen-2021-050785.

[101] S. Del Din, C. Kirk, A.J. Yarnall, L. Rochester, and J.M. Hausdorff. Body-Worn Sensors for Remote Monitoring of Parkinson's Disease Motor Symptoms: Vision, State of the Art, and Challenges Ahead. *Journal of Parkinson's Disease*, 11(1):35–47, 2021. doi: 10.3233/JPD-202471.

[102] R. Matias, V. Paixao, et al. Perspective on wearable sensor measurements and data science for parkinson's disease. *Front. Neurol.*, 8:677, 2017. doi: 10.3389/fneur.2017.00677.

[103] C. Pei-Hao, W. Rong-Long, L. De-Jyun, and S. Jin-Siang. Gait disorders in parkinson's disease: Assessment and management. *International Journal of Gerontology*, 7(4):189–193, 2013. doi: 10.1016/j.ijge.2013.03.005.

[104] E. Mirek, J.L. Kubica, J. Szymura, S. Pasiut, M. Rudzinska, and W. Chwala. Gait therapy effectiveness in patients with parkinson's disease on the basis of three-dimensional movement analysis. *Front. Neurol.*, 7:102, 2016. doi: 10.3389/fneur.2016.00102.

[105] R.P. Hubble, G.A. Naughton, P.A. Silburn, and Michael H.C. Wearable sensor use for assessing standing balance and walking stability in people with parkinson's disease: A systematic review. *PLOS ONE*, 10(4), 2015. doi: 10.1371/journal.pone.0123705.

[106] M.A. Hobert, W. Maetzler, K. Aminian, and Chiari L. Technical and clinical view on ambulatory assessment in parkinson's disease. *Acta Neurol Scand.*, 130(3):139–47, 2014. doi: 10.1111/ane.12248.

[107] OptiTrack. Optitrack system, 2022. URL https://optitrack.com/applications/movement-sciences/.

[108] G. Chen, C. Patten, D.H. Kothari, and F.E. Zajac. Gait differences between individuals with post-stroke hemiparesis and non-disabled controls at matched speeds. *Gait Posture*, 22:51–56, 2005. doi: 10.1016/j.gaitpost.2004.06.009.

[109] C. Ferraris, V. Cimolin, L. Vismara, V. Votta, G. Amprimo, et al. Monitoring of gait parameters in post-stroke individuals: A feasibility study using rgb-d sensors. *Sensors*, 21(17):594, 2021. doi: 10.3390/s21175945.

[110] E. Salonini, S. Gambazza, I. Meneghelli, G. Tridello, M. Sanguanini, et al. Active video game playing in children and adolescents with cystic fibrosis: Exercise or just fun? *Respir. Care*, 60:1172—-1179, 2015. doi: 10.4187/respcare.03576.

[111] R.A. Clark, K.J. Bower, B.F. Mentiplay, K. Paterson, and Y.H. Pua. Concurrent validity of the microsoft kinect for assessment of spatiotemporal gait variables. *J. Biomech.*, 46:2722—-2725, 2013. doi: 10.1016/j.jbiomech.2013.08.011.

[112] C. Ferraris, G. Amprimo, G. Masi, L. Vismara, R. Cremascoli, et al. Evaluation of Arm Swing Features and Asymmetry during Gait in Parkinson's Disease Using the Azure Kinect Sensor. *Sensors*, 22:6282, 2022. doi: 10.3390/s22166282.

[113] G. Amprimo, G. Masi, L. Priano, C. Azzaro, F. Galli, et al. Assessment Tasks and Virtual Exergames for Remote Monitoring of Parkinson's Disease: An Integrated Approach Based on Azure Kinect. *Sensors*, 22:8173, 2022. doi: 10.3390/s22218173.

[114] StereoLabs. Body tracking overview, 2022. URL https://www.stereolabs.com/docs/body-tracking/.

[115] A. Muro-de-la Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla. Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications. *Sensors*, 14(2):3362–3394, 2014. ISSN 1424-8220. doi: 10.3390/s140203362.

[116] P.H. Chen, R.L. Wang, D.J. Liou, and J.S. Shaw. Gait Disorders in Parkinson's Disease: Assessment and Management. *International Journal of Gerontology*, 7(4):189–193, 2013. doi: 10.1016/j.ijge.2013.03.005.

[117] J. Ghika, A. Wiegner, J. Fang, L. Davies, R. Young, et al. Portable system for quantifying motorabnormalities in parkinson's disease. *IEEE Trans. Biomed. Eng.*, 40:276–283, 1993. doi: 10.1109/10.216411.

[118] H.W. Huang, M.S. Ju, and C.C.K. Lin. Flexor and extensor muscle tone evaluated using the quantitativependulum test in stroke and parkinsonian patients. *J. Clin. Neurosci.*, 27:48—-52, 2016. doi: 10.1016/j.jocn.2015.07.031.

[119] F. Di Nardo, C. Morbidoni, A. Cucchiarelli, and S. Fioretti. Recognition of gait phases with a single knee electrogoniometer: A deep learning approach. *Electronics*, 9(2):355, 2020. doi: 10.3390/electronics9020355.

[120] V. Ruonala, A. Meigal, S.M. Rissanen, O. Airaksinen, M. Kankaanpää, et al. EMG signal morphology and kinematic parameters in essential tremor and Parkinson's disease patients. *Journal of Electromyography and Kinesiology*, 24(2):300–306, 2014. doi: 10.1016/j.jelekin.2013.12.007.

[121] S. Spasojević, T.V. Ilić, I. Stojković, V. Potkonjak, A. Rodić, et al. Quantitative assessment of the arm/hand movements in Parkinson's disease using a wireless armband device. *Frontiers in Neurology*, 8:388, 2017. doi: 10.3389/fneur.2017.00388.

[122] J.A. Robichaud, K.D. Pfann, C.L. Comella, and D.M. Corcos. Effect of medication on EMG patterns in individuals with Parkinson's disease. *Movement Disorders*, 17(5):950–960, 2002. doi: 10.1002/mds.10218.

[123] I. Mazzetta, A. Zampogna, A. Suppa, A. Gumiero, M. Pessione, and F. Irrera. Wearable sensors system for an improved analysis of freezing of gait in parkinson's disease using electromyography and inertial signals. *Sensors*, 19 (4):948, 2019. doi: 10.3390/s19040948.

[124] G. Shalin, S. Pardoel, E.D. Lemaire, J Nantel, and J. Kofman. Prediction and detection of freezing of gait in Parkinson ' s disease from plantar pressure data using long short - term memory neural - networks. *Journal of NeuroEngineering and Rehabilitation*, 5:1–15, 2021. doi: 10.1186/s12984-021-00958-5.

[125] A. Marcante, R. Di Marco, G. Gentile, C. Pellicano, F. Assogna, et al. Foot Pressure Wearable Sensors for Freezing of Gait Detection in Parkinson's Disease. *Sensors*, 21(1):128, 2021. doi: 10.3390/s21010128.

[126] A. Channa, N. Popescu, and V. Ciobanu. Wearable Solutions for Patients with Parkinson's Disease and Neurocognitive Disorder: A Systematic Review. *Sensors*, 20(9):2713. doi: 10.3390/s20092713.

[127] L. Sigcha, I. Pavón, N. Costa, S. Costa, M. Gago, et al. Automatic resting tremor assessment in parkinson's disease using smartwatches and multitask convolutional neural networks. *Sensors*, 21(1):291, 2021. doi: 10.3390/s21010291.

[128] C. Pérez-López, A. Samà, D. Rodríguez-Martín, J.M. Moreno-Aróstegui, J. Cabestany, et al. Dopaminergic-induced dyskinesia assessment based on a single belt-worn accelerometer. *Artificial Intelligence in Medicine*, 67:47–56, 2016. doi: 10.1016/j.artmed.2016.01.001.

[129] A. Emmanouil, E. Rousanoglou, A. Georgaki, and K. Boudolos. Concurrent Validity of Inertially Sensed Measures during Voluntary Body Sway in Silence and while Exposed to a Rhythmic Acoustic Stimulus: A Pilot Study. *Digital Biomarkers*, 5(1):65–73, 2021. doi: 10.1159/000514325.

[130] B. Shi, E.B.A.r Tay, W.L. Au, D.M.L. Tan, N.S.Y. Chia, et al. Detection of Freezing of Gait using Convolutional Neural Networks and Data from Lower Limb Motion Sensors. *IEEE Transactions on Biomedical Engineering*, 69(7): 2256–2267, 2022. doi: 10.1109/TBME.2022.3140258.

[131] M. Mancini, P. Carlson-Kuhta, C. Zampieri, J.G. Nutt, L. Chiari, et al. Postural sway as a marker of progression in Parkinson's disease: A pilot longitudinal study. *Gait and Posture*, 36(3):471–476, 2012. doi: 10.1016/j.gaitpost.2012.04.010.

[132] L Borzì, M. Varrecchia, G. Olmo, C.A. Artusi, M. Fabbri, et al. Home monitoring of motor fluctuations in Parkinson's disease patients. *Journal of Reliable Intelligent Environments*, 5(3):145–162, 2019. doi: 10.1007/s40860-019-00086-x.

[133] M. Ghislieri, L. Gastaldi, S. Pastorelli, S. Tadano, and V. Agostini. Wearable inertial sensors to assess standing balance: a systematic review. *Sensors*, 19 (19):1–25, 2019. doi: 10.3390/s19194075.

[134] A. Vienne, R.P. Barrois, S. Buffat, D. Ricard, and P.P. Vidal. Inertial sensors to assess gait quality in patients with neurological disorders: A systematic review of technical and analytical challenges. *Frontiers in Psychology*, 8:817, 2017. doi: 10.3389/fpsyg.2017.00817.

[135] D. Rodríguez-Martín, A. Samà, C. Pérez-López, A. Català, J.M.M. Arostegui, et al. Home detection of freezing of gait using Support Vector Machines through a single waist-worn triaxial accelerometer. *PLoS ONE*, 12(2): e0171764, 2017. doi: 10.1371/journal.pone.0171764.

[136] W. Sousa Lima, E. Souto, K. El-Khatib, R. Jalali, and J. Gama. Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview. *Sensors*, 19(14):3213, 2019. doi: 10.3390/s19143213.

[137] F. Demrozi, C. Turetta, and G. Pravadelli. B-HAR: an open-source baseline framework for in depth study of human activity recognition datasets and workflows. *arXiv preprint*, 2021. doi: arXiv:2101.10870.

[138] T Stampfler, M Elgendi, R Fletcher, and C Menon. Fall detection using accelerometer-based smartphones: Where do we go from here? *Frontiers in Public Health*, 10:996021, 2022. doi: 10.3389/fpubh.2022.996021.

[139] L. Pepa, L. Ciabattoni, F. Verdini, and M. Capecci. Smartphone Based Fuzzy Logic Freezing of Gait Detection in Parkinson's Disease. In *IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, pages 1–6, 2014. doi: 10.1109/MESA.2014.6935630.

[140] L. Borzì, S. Fornara, F. Amato, G. Olmo, C.A. Artusi, et al. Smartphone-based evaluation of postural stability in Parkinson's disease patients during quiet stance. *Electronics*, 9(6):1–14, 2020. doi: 10.3390/electronics9060919.

[141] H.B. Kim, H.J. Lee, W.W. Lee, S.K. Kim, H.S. Jeon, et al. Validation of Freezing-of-Gait Monitoring Using Smartphone. *Telemedicine and e-Health*, 24(11):899–907, 2018. doi: 10.1089/tmj.2017.0215.

[142] A. Godfrey, R. Conway, D. Meagher, and G. OLaighin. Direct measurement of human movement by accelerometry. *Medical eng. & physics*, 30(10): 1364–86, 2008.

[143] V. Grigas, V. Eidukynas, and A. Domeika. Acceleration-based evaluation of the human walking and running parameters. *Journal of Vibroengineering*, 11 (3):506–510, 2014.

[144] T. Iluz, E. Gazit, T: Herman, E. Sprecher, M. Brozgol, et al. Automated detection of missteps during community ambulation in patients with Parkinson's disease: a new approach for quantifying fall risk in the community setting. *Journal of NeuroEngineering and Rehabilitation*, 11:48, 2014. doi: 10.1186/1743-0003-11-48.

[145] A. Suppa, A. Kita, G. Leodori, A. Zampogna, E. Nicolini, et al. L-Dopa and freezing of gait in Parkinson's disease: Objective assessment through a wearable wireless system. *Frontiers in Neurology*, 8:406, 2017. doi: 10.3389/fneur.2017.00406.

[146] K. Bettecken, F. Bernhard, J. Sartor, M.A. Hobert, M. Hofmann, et al. No relevant association of kinematic gait parameters with Health-related Quality of Life in Parkinson's disease. *PLoS ONE*, 12(5):1–11, 2017. doi: 10.1371/journal.pone.0176816.

[147] G. Pacini Panebianco, M.C. Bisi, R. Stagni, and S. Fantozzi. Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements. *Gait and Posture*, 66:76–82, 2018. doi: 10.1016/j.gaitpost.2018.08.025.

[148] D. Lukšys, D. Jatužis, G. Jonaitis, and J. Griškevičius. Application of continuous relative phase analysis for differentiation of gait in neurodegenerative disease. *Biomedical Signal Processing and Control*, 67:102558, 2021. doi: 10.1016/j.bspc.2021.102558.

[149] F.A. Storm, C.J. Buckley, and C. Mazzà. Gait event detection in laboratory and real life settings: Accuracy of ankle and waist sensor based methods. *Gait and Posture*, 50:42–46, 2016. doi: 10.1016/j.gaitpost.2016.08.012.

[150] A. Köse, A. Cereatti, and U. Della Croce. Bilateral step length estimation using a single inertial measurement unit attached to the pelvis. *Journal of NeuroEngineering and Rehabilitation*, 9(1):9, 2012. doi: 10.1186/1743-0003-9-9.

[151] J.F. Daneault, S.I. Lee, F.N. Golabchi, S. Patel, L.C. Shih, et al. Estimating Bradykinesia in Parkinson's Disease with a Minimum Number of Wearable Sensors. In *IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 264–265, 2017. doi: 10.1109/CHASE.2017.94.

[152] Hopkins Medicine. Parkinson's disease risk factors and causes. URL https://www.hopkinsmedicine.org/health/conditions-and-diseases/parkinsons-disease/parkinsons-disease-risk-factors-and-causes.

[153] H. Dai, H. Lin, and T.C. Lueth. Quantitative assessment of parkinsonian bradykinesia based on an inertial measurement unit. *BioMedical Engineering Online*, 14(1):1–13, 2015. doi: 10.1186/s12938-015-0067-8.

[154] I. Griffiths, K. Kotschet, S. Arfon, Z.M. Xu, W. Johnson, et al. Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *Journal of Parkinson's Disease*, 2(1):47–55, 2012. doi: 10.3233/JPD-2012-11071.

[155] A. Sama, C. Perez-Lopez, J. Romagosa, D. Rodriguez-Martin, A. Catala, et al. Dyskinesia and motor state detection in Parkinson's Disease patients with a single movement sensor. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 1194–1197, 2012. doi: 10.1109/EMBC.2012.6346150.

[156] T. Bikias, D. Iakovakis, S. Hadjidimitriou, V. Charisis, and L. J. Hadjileontiadis. DeepFoG: An IMU-Based Detection of Freezing of Gait Episodes in Parkinson's Disease Patients via Deep Learning. *Frontiers in Robotics and AI*, 8:537384, 2021. doi: 10.3389/frobt.2021.537384.

[157] L. Borzì, G. Olmo, C.A. Artusi, and L. Lopiano. Detection of Freezing of Gait in People with Parkinson's Disease using Smartphones. In *IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 625–635, 2020. doi: 10.1109/COMPSAC48688.2020.0-186.

[158] L. Sigcha, N. Costa, I. Pavón, S. Costa, P. Arezes, et al. Deep Learning Approaches for Detecting Freezing of Gait in Parkinson's Disease Patients through On-Body Acceleration Sensors. *Sensors*, 20(7):1895, 2020. doi: 10.3390/s20071895.

[159] L. Sigcha, L. Borzì, I. Pavón, N. Costa, S. Costa, et al. Improvement of Performance in Freezing of Gait detection in Parkinson's Disease using Transformer networks and a single waist-worn triaxial accelerometer. *Engineering Applications of Artificial Intelligence*, 116:105482, 2022. doi: 10.1016/j.engappai.2022.105482.

[160] M.H. Noor, A. Nazir, M.N. Wahab, and J.O. Ling. Detection of Freezing of Gait Using Unsupervised Convolutional Denoising Autoencoder. *IEEE Access*, 9(11):115700–115709, 2021. doi: 10.1109/ACCESS.2021.3104975.

[161] N. Naghavi and E. Wade. Towards Real-time Prediction of Freezing of Gait in Patients with Parkinson's Disease: A Novel Deep One-class Classifier. *IEEE J Biomed Health Inform*, 26(4):1726–1736, 2021. doi: 10.1109/JBHI.2021.3103071.

[162] F. Demrozi, R. Bacchin, S. Tamburin, M. Cristani, and G. Pravadelli. Toward a Wearable System for Predicting Freezing of Gait in People Affected by Parkinson's Disease. *IEEE J Biomed Health Inform*, 24(9):2444–2451, 2020. doi: 10.1109/JBHI.2019.2952618.

[163] F. Irrera, J. Cabestany, and A. Suppa. New advanced Wireless technologies for objective Monitoring of Motor Symptoms in Parkinson's disease. *Frontiers in neurology*, 9:216, 2018. doi: 10.3389/fneur.2018.00216.

[164] P. Tahafchi, R. Molina, J.A. Roper, K. Sowalsky, C.J. Hass, et al. Freezing-of-Gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 2867–2870, 2017. doi: 10.1109/EMBC.2017.8037455.

[165] A.T. Tzallas, M.G. Tsipouras, G. Rigas, D.G. Tsalikakis, E.C. Karvounis, et al. PERFORM: a system for monitoring, assessment and management of patients with Parkinson's disease. *Sensors*, 14(11):21329–21357, 2014. doi: 10.3390/s141121329.

[166] M. Capecci, L. Pepa, F. Verdini, and M.G. Ceravolo. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease. *Gait and Posture*, 50:28–33, 2016. doi: 10.1016/j.gaitpost.2016.08.018.

[167] S. Pardoel, J. Kofman, J. Nantel, and E.D. Lemaire. Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: A review. *Sensors (Switzerland)*, 19(23), 2019. doi: 10.3390/s19235141.

[168] A. Samà, D. Rodríguez-Martín, C. Pérez-López, A. Català, S. Alcaine, et al. Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments. *Pattern Recognition Letters*, 105:135–143, 2018. doi: 10.1016/j.patrec.2017.05.009.

[169] K. Smulders, M.L. Dale, P. Carlson-Kuhta, J.G. Nutt, and F.B. Horak. Pharmacological treatment in Parkinson's disease: Effects on gait. *Parkinsonism Related Disorders*, 31:3–13, 2016. ISSN 1353–8020. doi: https://doi.org/10.1016/j.parkreldis.2016.07.006.

[170] P. Ginis, E. Nackaerts, A. Nieuwboer, and E. Heremans. Cueing for people with Parkinson's disease with freezing of gait: A narrative review of the state-of-the-art and novel perspectives. *Annals of Physical and Rehabilitation Medicine*, 61(6):407–416, 2017. doi: 10.1016/j.rehab.2017.08.002.

[171] F.B. Horak, D. Dimitrova, and J.G. Nutt. Direction-specific postural instability in subjects with Parkinson's disease. *Experimental Neurology*, 193(2):504–521, 2005. doi: 10.1016/j.expneurol.2004.12.008.

[172] M. Mancini, L. Rocchi, F.B. Horak, and L. Chiari. Effects of Parkinson's disease and levodopa on functional limits of stability. *Clinical Biomechanics*, 23(4):450–458, 2008. doi: 10.1016/j.clinbiomech.2007.11.007.

[173] M. Mancini, F.B. Horak, C. Zampieri, P. Carlson-Kuhta, J.G. Nutt, et al. Trunk accelerometry reveals postural instability in untreated Parkinson's disease. *Parkinsonism and Related Disorders*, 17(7):557–562, 2011. doi: 10.1016/j.parkreldis.2011.05.010.

[174] P.M. Deshmukh, C M. Russell, L.E. Lucarino, and S.N. Robinovitch. Enhancing clinical measures of postural stability with wearable sensors. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 4521–4524, 2012. doi: 10.1109/EMBC.2012.6346972.

[175] M. O'sullivan, C. Blake, C. Cunningham, G. Boyle, and F. Ciarán. Correlation of accelerometry with clinical balance tests in older fallers and non-fallers. *Age and Ageing*, 38(3):308–313, 2009. doi: 10.1093/ageing/afp009.

[176] L. Borzì, G. Olmo, C.A. Artusi, M. Fabbri, M.G. Rizzone, et al. A new index to assess turning quality and postural stability in patients with Parkinson's disease. *Biomedical Signal Processing and Control*, 62:102059, 2020. doi: 10.1016/j.bspc.2020.102059.

[177] S. Rezvanian, T. Lockhart, C. Frames, R. Soangra, and A. Lieberman. Motor Subtypes of Parkinson's Disease Can Be Identified by Frequency Component of Postural Stability. *Sensors*, 18(4), 2018. doi: 10.3390/s18041102.

[178] J. Shah, L. Pillai, D.K. Williams, S.M. Doerhoff, L. Larson-Prior, et al. Increased foot strike variability in Parkinson's disease patients with freezing of gait. *Parkinsonism and Related Disorders*, 53:58–63, 2018. doi: 10.1016/j.parkreldis.2018.04.032.

[179] L. Bizovska, Z. Svoboda, M. Janura, M.C. Bisi, and N. Vuillerme. Local dynamic stability during gait for predicting falls in elderly people: A one-year prospective study. *PLoS ONE*, 13(5):1–7, 2018. doi: 10.1371/journal.pone.0197091.

[180] M. Ghahramani, F. Naghdy, D. Stirling, G. Naghdy, J. Potter, et al. Assessing the Risk of Fall in Older People through Turning Test. In *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2018. doi: 10.1109/MeMeA.2018.8438710.

[181] S. Mellone, M. Mancini, L.A. King, F.B. Horak, and L. Chiari. The quality of turning in Parkinson's disease: a compensatory strategy to prevent postural instability? *Journal of NeuroEngineering and Rehabilitation*, 13:39, 2016. doi: 10.1186/S12984-016-0147-4.

[182] A. Sánchez-Ferro, M. Elshehabi, C. Godinho, D: Salkovic, M.A. Hobert, et al. New methods for the assessment of Parkinson's disease: A systematic review. *Movement Disorders*, 31(9):1283–1292, 2016. doi: 10.1002/mds.26723.

[183] M. Son, C. Youm, S. Cheon, J. Kim, M. Lee, et al. Evaluation of the turning characteristics according to the severity of Parkinson's disease during the timed up and go test. *Aging Clinical and Experimental Research*, 29(6): 1191–1199, 2017. doi: 10.1007/s40520-016-0719-y.

[184] M.H. Pham, M. Elshehabi, L. Haertner, S. Del Din, K. Srulijes, et al. 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*, 8:457, 2017. doi: 10.3389/fneur.2017.00457.

[185] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro. Body-Sensor-Network-Based Kinematic Characterization and Comparative Outlook of UPDRS Scoring in Leg Agility, Sit-to-Stand, and Gait Tasks in Parkinson's Disease. *IEEE Journal on Biomedical and Health Informatics*, 19(6):1777–1793, 2015.

[186] S. Aghanavesi, F. Bergquist, D. Nyholm, M. Senek, and M. Memedi. Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge. *IEEE Journal of Biomedical and Health Informatics*, 24(1):111–119, 2019. doi: 10.1109/JBHI.2019.2898332.

[187] C. Ornelas-vences, L.P. Sánchez-fernández, .A. Sánchez-pérez, and J.M. Martínez-hernández. Computer model for leg agility quantification and assessment for Parkinson ' s disease patients. *Medical & Biological Engineering & Computing*, 57:463–476, 2019.

[188] F. Lipsmeier, K.I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Movement Disorders*, 33(8): 1287–1297, 2018. doi: 10.1002/mds.27376.

[189] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi. Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey. *IEEE Access*, 8:210816–210836, 2020. doi: 10.1109/ACCESS.2020.3037715.

[190] R. Lu, Y. Xu, X. Li, Y Fan, W. Zeng, et al. Evaluation of Wearable Sensor Devices in Parkinson's Disease: A Review of Current Status and Future Prospects. *Parkinson's Disease*, 2020:4693019, 2020. doi: 10.1155/2020/4693019.

[191] H.W. Loh, W. Hong, C.P. Ooi, S. Chakraborty, P.D. Barua, et al. Application of Deep Learning Models for Automated Identification of Parkinson's disease: A Review. *Sensors*, 21(21):7034, 2021. doi: 10.3390/s21217034.

[192] B. Mirza, W. Wang, J. Wang, H. Choi, N. C. Chung, et al. Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes*, 10(2):87, 2019. doi: 10.3390/genes10020087.

[193] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53, 2021.

[194] J. Mei, C. Desrosiers, and J. Frasnelli. Machine learning for the diagnosis of parkinson's disease: A review of literature. *Frontiers in Aging Neuroscience*, 13, 2021. doi: 10.3389/fnagi.2021.633752.

[195] I.H. Sarker. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):1–20, 2021.

[196] P. Barralon, N. Vuillerme, and N. Noury. Walk detection with a kinematic sensor: Frequency and wavelet comparison. In *Annual Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1711–1714, 2006. doi: 10.1109/IEMBS.2006.260770.

[197] E. Rovini, C. Maremmani, and F. Cavallo. How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review. *Frontiers in Neuroscience*, 11:555, 2017.

[198] S. Mazilu, A. Calatroni, E. Gazit, A. Mirelman, J.M. Hausdorff, et al. Prediction of freezing of gait in Parkinson's from physiological wearables: An exploratory study. *IEEE Journal of Biomedical and Health Informatics*, 19 (6):1843–1854, 2015. doi: 10.1109/JBHI.2015.2465134.

[199] E. Ramanujam, T. Perumal, and S. Padmavathi. Human Activity Recognition with Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review. *IEEE Sensors Journal*, 21(12):1309–13040, 2021. doi: 10.1109/JSEN.2021.3069927.

[200] N. Haji Ghassemi, J. Hannink, C.F. Martindale, H. Gaßner, M. Müller, et al. Segmentation of gait sequences in sensor-based movement analysis: A comparison of methods in Parkinson's disease. *Sensors*, 18(1):1–15, 2018. doi: 10.3390/s18010145.

[201] M. Radovic, M. Ghalwash, N. Filipovic, et al. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1):9, 2017. doi: 10.1186/s12859-016-1423-9.

[202] Iryna Sydorenko. How to choose the right machine learning algorithm: A pragmatic approach, 2021. URL https://labelyourdata.com/articles/how-to-choose-a-machine-learning-algorithm.

[203] R. ElShawi, Y. Sherif, M. Al-Mallah, and S. Sakr. Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*, 37:1633–1650, 2021. doi: 10.1111/coin.12410.

[204] L. Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006. doi: 10.1111/j.2006.0030-1299.14714.x.

[205] Italo José. K-nearest neighbors, 2018. URL https://towardsdatascience.com/knn-k-nearest-neighbors-1-a4707b24bd1d.

[206] Aditya Rohilla. A brief introduction to support vector machine. URL https://adityarohilla.com/2018/11/02/a-brief-introduction-to-support-vector-machine/.

[207] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015. doi: 10.1038/nature14539.

[208] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019. doi: 10.1016/j.patrec.2018.02.010.

[209] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv*, 1412.6980, 2014. doi: 10.48550/arXiv.1412.6980.

[210] N.V. Chawla, K.V. Bowyer, L.O. Hall, and W.P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, 2002. doi: 10.1613/jair.953.

[211] H. Dai, G. Cai, Z. Lin, Z. Wang, and Q. Ye. Validation of inertial sensing-based wearable device for tremor and bradykinesia quantification. *IEEE Journal of Biomedical and Health Informatics*, pages 997–1005, 2021. doi: 10.1109/JBHI.2020.3009319.

[212] A. Rodríguez-Molinero, C. Pérez-López, A. Samà, D. Rodríguez-Martín, et al. Estimating dyskinesia severity in Parkinson's disease by using a waist-worn sensor: concurrent validity study. *Scientific Reports*, 9(1):1–7, 2019. doi: 10.1038/s41598-019-49798-3.

[213] M.D. Hssayeni, J. Jimenez-Shahed, M.A. Burack, and B. Ghoraani. Dyskinesia severity estimation in patients with Parkinson's disease using wearable sensors and a deep LSTM network. In *42th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 6001–6004, 2020. doi: 10.1109/EMBC44109.2020.9176847.

[214] L. Mesin, P. Porcu, D. Russu, G. Farina, and L. Borzì. A multi-modal analysis of the freezing of gait phenomenon in parkinson's disease. *Sensors*, 22(7):2613, 2022. doi: 10.3390/s22072613.

[215] A. Rodríguez-Molinero, A. Samà, C. Pérez-López, Rodríguez-Martín D., et al. Analysis of correlation between an accelerometer-based algorithm for detecting parkinsonian gait and UPDRS subscales. *Frontiers in Neurology*, 8:431, 2017. doi: 10.3389/fneur.2017.00431.

[216] L. Borzì, I. Mazzetta, A. Zampogna, A. Suppa, F. Irrera, et al. Predicting Axial Impairment in Parkinson's Disease through a Single Inertial Sensor. *Sensors*, 22:412, 2022. doi: 10.3390/s22020412.

[217] J. Camps, A. Samà, M. Martín, D. Rodríguez-Martín, C. Pérez-López, et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, 139:119–131, 2018. doi: 10.1016/j.knosys.2017.10.017.

[218] J. O'Day, M. Lee, K. Seagers, S. Hoffman, A. Jih-Schiff, L. Kidziński, and et al. Assessing inertial measurement unit locations for freezing of gait detection and patient preference. *J NeuroEngineering Rehabil*, 19(1):20, 2022. doi: 10.1186/s12984-022-00992-x.

[219] L. Borzì, I. Mazzetta, A. Zampogna, A. Suppa, G. Olmo, et al. Prediction of freezing of gait in Parkinson's disease using wearables and machine learning. *Sensors*, 21(2):1–19, 2021. doi: 10.3390/s21020614.

[220] I. Mazzetta, A. Zampogna, A. Suppa, A. Gumiero, M. Pessione, et al. Wearable sensors system for an improved analysis of freezing of gait in Parkinson's disease using electromyography and inertial signals. *Sensors*, 19(4):948, 2019. doi: 10.3390/s19040948.

[221] S. Pardoel, G. Shalin, J. Nantel, E.D. Lemaire, and J. Kofman. Early detection of freezing of gait during walking using inertial measurement unit and plantar pressure distribution data. *Sensors*, 21(6):1–14, 2021. doi: 10.3390/s21062246.

[222] S.T. Moore, H.G. MacDougall, and W.G. Ondo. Ambulatory monitoring of freezing of gait in Parkinson's disease. *Journal of Neuroscience Methods*, 167 (2):340–348, 2008. doi: 10.1016/j.jneumeth.2007.08.023.

[223] M. Bächlin, J.M. Hausdorff, D. Roggen, N. Giladi, M. Plotnik, et al. Online detection of freezing of gait in parkinson's disease patients: A performance characterization. In *Fourth International Conference on Body Area Networks*, pages 1–8, 2009. doi: 10.4108/ICST.BODYNETS2009.5852.

[224] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Tröster, et al. Online Detection of Freezing of Gait with Smartphones and Machine Learning Techniques. In *6th International ICST Conference on Pervasive Computing Technologies for Healthcare*, number 3, pages 123–130, 2012. doi: 10.4108/icst.pervasivehealth.2012.248680.

[225] A. Arami, A. Poulakakis-Daktylidis, Y.F. Tai, and E. Burdet. Prediction of Gait Freezing in Parkinsonian Patients: A Binary Classification Augmented With Time Series Prediction. *IEEE transactions on neural systems and rehabilitation engineering*, 27(9):1909–1919, 2019. doi: 10.1109/TNSRE.2019.2933626.

[226] N. Naghavi and E. Wade. Prediction of freezing of gait in Parkinson's disease using statistical inference and lower-limb acceleration data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(5):947–955, 2019. doi: 10.1109/TNSRE.2019.2910165.

[227] Y. Zhang, W. Yan, Y. Yao, J.B. Ahmed, Y. Tan, and D. Gu. Prediction of Freezing of Gait in Patients with Parkinson's Disease by Identifying Impaired Gait Patterns. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(3):591–600, 2020. doi: 10.1109/TNSRE.2020.2969649.

[228] L. Palmerini, L. Rocchi, S. Mazilu, E. Gazit, J.M. Hausdorff, et al. Identification of characteristic motor patterns preceding freezing of gait in Parkinson's disease using wearable sensors. *Frontiers in Neurology*, 8:1–12, 2017. doi: 10.3389/fneur.2017.00394.

[229] N.K. Orphanidou, A. Hussain, R. Keight, P. Lishoa, J. Hind, et al. Predicting Freezing of Gait in Parkinsons Disease Patients Using Machine Learning. In *IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, 2018. doi: 10.1109/CEC.2018.8477909.

[230] F.B. Horak and M. Mancini. Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors. *Movement Disorders*, 28(11): 1544–1551, 2013. doi: 10.1002/mds.25684.

[231] A. Zampogna, I. Mileti, E. Palermo, C. Celletti, M. Paoloni, et al. Fifteen years of wireless sensors for balance assessment in neurological disorders. *Sensors*, 20(11):1–32, 2020. doi: 10.3390/s20113247.

[232] A. Weiss, T. Herman, A. Mirelman, S.S. Shiratzky, N. Giladi, et al. The transition between turning and sitting in patients with Parkinson's disease: A wearable device detects an unexpected sequence of events. *Gait & Posture*, 67:224–229, 2019. doi: 10.1016/J.GAITPOST.2018.10.018.

[233] A. Atrsaei, A. Paraschiv-Ionescu, H. Krief, Y. Henchoz, B. Santos-Eggimann, et al. Instrumented 5-Time Sit-To-Stand Test: Parameters Predicting Serious Falls beyond the Duration of the Test. *Gerontology*, 68:587—600, 2021. doi: 10.1159/000518389.

[234] S.F. Castiglia, A. Tatarelli, D. Trabassi, R. De Icco, V. Grillo, et al. Ability of a set of trunk inertial indexes of gait to identify gait instability and recurrent fallers in parkinson's disease. *Sensors*, 21(10):1–17, 2021. doi: 10.3390/s21103449.

[235] T. Mitchell, D. Conradsson, and C. Paquette. Gait and trunk kinematics during prolonged turning in Parkinson's disease with freezing of gait. *Parkinsonism & Related Disorders*, 64:188–193, 2019. doi: 10.1016/j.parkreldis.2019.04.011.

[236] A. Zhan, S. Mohan, C. Tarolli, R.B. Schneider, J.L. Adams, et al. Using smartphones and machine learning to quantify Parkinson's disease severity:

the mobile Parkinson disease score. *JAMA Neurology*, 75(7):876–880, 2018. doi: 10.1001/jamaneurol.2018.0809.

[237] A. Landolfi, C. Ricciardi, L. Donisi, G. Cesarelli, J. Troisi, et al. Machine Learning Approaches in Parkinson's Disease. *Current Medicinal Chemistry*, 28(32):6548–6568, 2021. doi: 10.2174/0929867328999210111211420.

[238] J.F. van der Heeden, J. Marinus, P. Martinez-Martin, C. Rodriguez-Blazquez, V.J. Geraedts, et al. Postural instability and gait are associated with severity and prognosis of Parkinson's disease. *Neurology*, 86(24):2243–50, 2016. doi: 10.1212/WNL.0000000000002768.

[239] D. Safarpour, M.L. Dale, V.V. Shah, L. Talman, P. Carlson-Kuhta, et al. Surrogates for rigidity and PIGD MDS-UPDRS subscores using wearable sensors. *Gait & Posture*, 91:186–191, 2021. doi: 10.1016/j.gaitpost.2021.10.029.

[240] R. Zia Ur Rehman, L. Rochester, A.J. Yarnall, and S. Del Din. Predicting the Progression of Parkinson's Disease MDS-UPDRS-III Motor Severity Score from Gait Data using Deep Learning. In *43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 249–252, 2021. doi: 10.1109/EMBC46164.2021.9630769.

[241] M. Giuberti, G. Ferrari, L. Contin, V. Cimolin, C. Azzaro, et al. Automatic UPDRS Evaluation in the Sit-to-Stand Task of Parkinsonians: Kinematic Analysis and Comparative Outlook on the Leg Agility Task. *IEEE Journal on Biomedical and Health Informatics*, 19(3):803–814, 2015.

[242] P. Kassavetis, T.A. Saifee, G. Roussos, L. Drougkas, M. Kojovic, et al. Developing a Tool for Remote Digital Assessment of Parkinson's Disease. *Movement Disorders Clinical Practice*, 3(1):59–64, 2015. doi: 10.1002/mdc3.12239.

[243] R.B. Postuma, D. Berg, M. Stern, W. Poewe, C.W. Olanow, et al. Mds clinical diagnostic criteria for parkinson's disease. *Movement Disorders*, 30(12): 1591–1601, 2015. doi: https://doi.org/10.1002/mds.26424.

[244] .C.L Tomlinson, R. Stowe, S. Patel, C. Rick, R. Gray, et al. Systematic review of levodopa dose equivalency reporting in parkinson's disease. *Movement Disorders*, 25(15):2649–2653, 2010. doi: https://doi.org/10.1002/mds.23429.

[245] D. Comotti, M. Galizzi, and A. Vitali. nememsi: One step forward in wireless attitude and heading reference systems. In *2014 International Symposium on Inertial Sensors and Systems (ISISS)*, pages 1–4. IEEE, 2014. doi: 10.1109/ISISS.2014.6782521.

[246] B.R. Greene, D. McGrath, R. O'Neill, Ka.J. O'Donovan, A. Burns, et al. An adaptive gyroscope-based algorithm for temporal gait analysis. *Medical and Biological Engineering and Computing*, 48(12):1251–1260, 2010. doi: 10.1007/s11517-010-0692-0.

[247] D. Gouwanda and A.A. Gopalai. A robust real-time gait event detection using wireless gyroscope and its application on normal and altered gaits. *Medical Engineering and Physics*, 37(2):219–225, 2015. doi: 10.1016/j.medengphy.2014.12.004.

[248] T.H. Nguyen, T.P. Pham, C.Q. Ngo, and T.T. Nguyen. A SVM Algorithm for Investigation of Tri-Accelerometer Based Falling Data. *American Journal of Signal Processing*, 6(2):56–65, 2016. doi: 10.5923/j.ajsp.20160602.03.

[249] C. Pérez-López, A. Samà, D. Rodríguez-Martín, A. Català, J. Cabestany, et al. Assessing Motor Fluctuations in Parkinson's Disease Patients Based on a Single Inertial Sensor. *Sensors*, 16(12):2132, 2016. doi: 10.3390/s16122132.

[250] C. Ni, S. Guan, and Y. Li. Human activity recognition using a improved model based on multi-head cnn-lstm. In *7th International Conference on Information Science and Control Engineering (ICISCE)*, pages 688–693, 2020. doi: 10.1109/ICISCE50968.2020.00147.

[251] T. Zebin, P.J. Scully, N. Peek, A.J. Casson, and K.B. Ozanyan. Design and Implementation of a Convolutional Neural Network on an Edge Computing Smartphone for Human Activity Recognition. *IEEE Access*, 7:133509–133520, 2019. doi: 10.1109/ACCESS.2019.2941836.

[252] F.M. Rueda, R. Grzeszick, G.A. Fink, S. Feldhorst, and M. Ten Hompel. Convolutional neural networks for human activity recognition using body-worn sensors. *Informatics*, 5(2):1–17, 2018. doi: 10.3390/informatics5020026.

[253] T.M. Steffen, T.A. Hacker, and L. Mollinger. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up Go Test, and gait speeds. *Physical Therapy*, 82(2):128–137, 2012.

[254] M.D. Tundo, E. Lemaire, and N. Baddour. Correcting Smartphone orientation for accelerometer-based analysis. pages 58–62, 2013. doi: 10.1109/MeMeA.2013.6549706.

[255] M.D. Latt, H.B. Menz, V.S. Fung, and S.R. Lord. Acceleration patterns of the head and pelvis during gait in older people with Parkinson's disease: A comparison of fallers and nonfallers. *Journals of Gerontology*, 64(6):700–706, 2009. doi: 10.1093/gerona/glp009.

[256] M. Miller, S.J. Ozinga, A.B. Rosenfeldt, and J.L. Alberts. Quantifying turning behavior and gait in Parkinson's disease using mobile technology. *IBRO Reports*, 5:10–16, 2018. doi: 10.1016/j.ibror.2018.06.002.

[257] B. Huijben, K.S. van Schooten, J.H. van Dieen, and M. Pijnappels. The effect of walking speed on quality of gait in older adults. *Gait & Posture*, 65:112–116, 2018. doi: 10.1016/j.gaitpost.2018.07.004.

[258] A. Galán-Mercant and A.I. Cuesta-Vargas. Differences in trunk accelerometry between frail and non-frail elderly persons in functional tasks. *BMC Research Notes*, 7(1):1–9, 2014. doi: 10.1186/1756-0500-7-100.

[259] M. Ponti, P. Bet, C.L. Oliveira, and P.C. Castro. Better than counting seconds: Identifying fallers among healthy elderly using fusion of accelerometer features and dual-task Timed Up and Go. *PLoS ONE*, 12(4):1–21, 2017. doi: 10.1371/journal.pone.0175559.

[260] A. Weiss, T. Herman, N. Giladi, and J.M. Hausdorff. Objective assessment of fall risk in Parkinson's disease using a body-fixed sensor worn for 3 days. *PLoS One*, 9(5):e96675, 2014. doi: 10.1371/journal.pone.0096675.

[261] C.J.C. Lamoth, P.J. Beek, and O.G. Meijer. Pelvis-thorax coordination in the transverse plane during gait. *Gait and Posture*, 16(2):101–114, 2002. doi: 10.1016/S0966-6362(01)00146-1.

[262] S.M. Rispens, M. Pijnappels, K.S. van Schooten, P.J. Beek, A. Daffertshofer, et al. Consistency of gait characteristics as determined from acceleration data collected at different trunk locations. *Gait and Posture*, 40(1):187–192, 2014. doi: 10.1016/j.gaitpost.2014.03.182.

[263] R. Moe-Nilssen and J.L. Helbostad. Estimation of gait cycle characteristics by trunk accelerometry. *Journal of Biomechanics*, 37(1):121–126, 2004. doi: 10.1016/S0021-9290(03)00233-1.

[264] E.P. Doheny, D. McGrath, B.R. Greene, L. Walsh, D. McKeown, et al. Displacement of centre of mass during quiet standing assessed using accelerometry in older fallers and non-fallers. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 3300–3303, 2012. doi: 10.1109/EMBC.2012.6346670.

[265] A.P. Stylianou, M.A. McVey, K.E. Lyons, R. Pahwa, and C.W. Luchies. Postural sway in patients with mild to moderate parkinson's disease. *International Journal of Neuroscience*, 121(11):614–621, 2011. doi: 10.3109/00207454.2011.602807.

[266] A.J. Solomon, J.V. Jacobs, K.V. Lomond, and S.M. Henry. Detection of postural sway abnormalities by wireless inertial sensors in minimally disabled patients with multiple sclerosis: A case-control study. *Journal of NeuroEngineering and Rehabilitation*, 12(1):74, 2015. doi: 10.1186/s12984-015-0066-9.

[267] M. Mancini, A. Salarian, P. Carlson-Kuhta, C. Zampieri, L. King, et al. ISway: A sensitive, valid and reliable measure of postural control. *Journal of NeuroEngineering and Rehabilitation*, 9(1):59, 2012. doi: 10.1186/1743-0003-9-59.

[268] C.H. Lee and T.L. Sun. Evaluation of postural stability based on a force plate and inertial sensor during static balance measurements. *J Physiol Anthropol*, 37(1):27. doi: 10.1186/s40101-018-0187-5.

[269] G. Madzarov, D. Gjorgjevikj, and I. Chorbev. A Multi-class SVM Classifier Utilizing Binary Decision Tree Support vector machines for pattern recognition. *Informatica*, 33:233–241, 2009.

[270] C. Ahlrichs, A. Samà, M. Lawo, J. Cabestany, D. Rodríguez-Martín, et al. 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. doi: 10.1007/s11517-015-1395-3.

[271] L.S. Vargas-Valencia, A. Elías, A. Frizera Neto, and E. Rocón. Body to sensor calibration procedure for lower limb joint angle estimation applied to imu-based gait analysis. In *XXIV Brazilian Congress on Biomedical Engineering*, pages 777–780, 2014.

[272] B. Sijobert, M. Benoussaad, J. Denys, R. Pissard-Gibollet, C. Geny, et al. Implementation and validation of a stride length estimation algorithm, using a single basic inertial sensor on healthy subjects and patients suffering from parkinson's disease. *Health*, 7:704–714, 2015. doi: 10.4236/health.2015.76084.

[273] D. Roetenberg, H.J. Luinge, C.T.M. Baten, and P.H. Veltink. Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(3):395–405, 2005. doi: 10.1109/TNSRE.2005.847353.

[274] M. Caruso, A.M. Sabatini, D. Laidig, T. Seel, M. Knaflitz, et al. Analysis of the accuracy of ten sensor fusion algorithms for orientation estimation using inertial and magnetic sensing under optimal conditions: a single outfit for every season does not work. *Sensors*, 21(7):2543, 2021. doi: 10.3390/s21072543.

[275] R. Bouça-Machado, Co. Jalles, D. Guerreiro, F. Pona-Ferreira, D. Branco, et al. Gait Kinematic Parameters in Parkinson's Disease: A Systematic Review. *Journal of Parkinson's Disease*, 10(3):843–853, 2020. doi: 10.3233/JPD-201969.

[276] N. Abhayasinghe and I. Murray. Human gait phase recognition based on thigh movement computed using IMUs. In *IEEE 9th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–4, 2014. doi: 10.1109/ISSNIP.2014.6827604.

[277] A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, 2004. doi: 10.1023/B:STCO.0000035301.49549.88.

[278] O.L. Mangasarian and D.R. Musicant. Robust linear and support vector regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9):950–955, 2000. doi: 10.1109/34.877518.

[279] L. Borzì, A. Manoni, A. Zampogna, F. Irrera, A. Suppa, et al. Correlation between wearable inertial sensor data and standardised parkinson's disease axial impairment measures using machine learning. In *IEEE 21st Mediterranean Electrotechnical Conference (MELECON)*, pages 732–736, 2022. doi: 10.1109/MELECON53508.2022.9843018.

[280] P. Bonato, D.M. Sherrill, D.G. Standaert, S.S. Salles, and M. Akay. Data mining techniques to detect motor fluctuations in Parkinson's disease. In *IEEE 26th Annual International Conference of Engineering in Medicine and Biology Society (EMBS)*, pages 4766–4769, 2004. doi: 10.1109/IEMBS.2004.1404319.

[281] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, et al. Monitoring Motor Fluctuations in Patients With Parkinson's Disease Using Wearable Sensors. *IEEE Transactions on Information Technology in Biomedicine*, 13 (6):864–873, 2009. doi: 10.1109/TITB.2009.2033471.

[282] D.A. Heldman, D.E. Filipkowsk, D.E. Riley, C.M. Whitney, B.L. Walter, et al. Automated motion sensor quantification of gait and lower extremity bradykinesia. In *Annual International Conference of Engineering in Medicine and Biology Society (EMBC)*, pages 1956–1959, 2012. doi: 10.1109/EMBC.2012.6346338.

[283] J. Kim, Y. Kwon, Y. Kim, H. Chung, G. Eom, et al. Analysis of lower limb bradykinesia in parkinson's disease patients. *Geriatrics & gerontology international*, 12(2):257–264, 2012. doi: 10.1111/j.1447-0594.2011.00761.x.

[284] R. Moe-Nilssen and J.L. Helbostad. Estimation of gait cycle characteristics by trunk accelerometry. *Journal of Biomechanics*, 37(1):121–126, 2004. doi: 10.1016/s0021-9290(03)00233-1.

[285] D. Rodríguez-Martín, J. Cabestany, C. Pérez-López, M. Pie, J. Calvet, et al. A New Paradigm in Parkinson's Disease Evaluation With Wearable Medical Devices: A Review of STAT-ON. *Front Neurol.*, 2(13):912343, 2022. doi: 10.3389/fneur.2022.912343.

[286] L.A. King, M. Mancini, K. Priest, A. Salarian, F. Rodrigues-De-Paula, et al. Do clinical scales of balance reflect turning abnormalities in people With Parkinson's disease? *Journal of Neurologic Physical Therapy*, 36(1):25–31, 2012. doi: 10.1097/NPT.0b013e31824620d1.

[287] L. Borzì, L. Sigcha, D. Rodríguez-Martín, and G. Olmo. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. *Artificial Intelligence in Medicine*, xx (xx):xx–xx, 2022.

[288] Y. Xia, Z. Yao, Y. Lu, D. Zhang, and N. Cheng. A Machine Learning Approach to Detecting of Freezing of Gait in Parkinson's Disease Patients. *Journal of Medical Imaging and Health Informatics*, 8(4):647–654, 2018. doi: 10.1166/jmihi.2018.2379.

[289] B. Li, Y. Zhang, L. Tang, C. Gao, and D. Gu. Automatic Detection System for Freezing of Gait in Parkinson's Disease Based on the Clustering Algorithm. In *2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pages 1640–9, 2018. doi: 10.1109/IMCEC.2018.8469471.

[290] R. Chee, A. Murphy, M. Danoudis, N. Georgiou-Karistianis, and R. Iansek. Gait freezing in Parkinson's disease and the stride length sequence effect interaction. *Brain*, 132(8):2151–2160, 2009. doi: 10.1093/brain/awp053.

[291] H. Terashi, T. Taguchi, Y. Ueta, O. Yoshihiko, M. Hiroshi, et al. Analysis of non-invasive gait recording under free-living conditions in patients with Parkinson's disease: relationship with global cognitive function and motor abnormalities. *BMC Neurology*, 20(1):161, 2020. doi: 10.1186/s12883-020-01729-w.

[292] Z. Wu, M. Zhong, X. Jiang, B. Shen, J. Zhu, et al. Can Quantitative Gait Analysis Be Used to Guide Treatment of Patients with Different Subtypes of Parkinson's Disease? *Neuropsychiatric Disease and Treatment*, 9(16):2335–2341, 2020. doi: 10.2147/NDT.S266585.

[293] C. Curtze, J.G. Nutt, P. Carlson-Kuhta, M. Mancini, and F.B. Horak. Gait Analysis in Parkinson's Disease: An Overview of the Most Accurate Markers for Diagnosis and Symptoms Monitoring. *Movement Disorders*, 30(10):1361–1370, 2015. doi: 10.1002/mds.26269.

[294] D. de Kam, J. Nonnekes, L.B. Oude Nijhuis, et al. Dopaminergic medication does not improve stepping responses following backward and forward balance perturbations in patients with parkinson's disease. *Journal of Neurology*, 261(12):2330–2337, 2014. doi: doi.org/10.1007/s00415-014-7496-3.

[295] I. Di Giulio, R. St George, E. Kalliolia, A.L. Peters, P. Limousin, et al. Maintaining balance against force perturbations: impaired mechanisms unresponsive to levodopa in Parkinson's disease. *Journal of Neurophysiology*, 116(2):493–502, 2016. doi: 10.1152/jn.00996.2015.

[296] A. Zampogna, I. Mileti, F. Martelli, M. Paoloni, Z. Del Prete, et al. Early balance impairment in Parkinson's Disease: Evidence from Robot-assisted axial rotations. *Clinical Neurophysiology*, 132:2422—2430, 2021. doi: 10.1016/j.clinph.2021.06.023.

[297] M.T. Barbe, M. Amarell, A.H. Snijders, E. Florin, E.L. Quatuor, et al. Gait and upper limb variability in parkinson's disease patients with and without freezing of gait. *Journal of Neurology*, 261(2):330–342, 2014. doi: 10.1007/s00415-013-7199-1.

[298] M. Plotnik, N. Giladi, Y. Balash, C. Peretz, and J.M. Hausdorff. Is freezing of gait in parkinson's disease related to asymmetric motor function? *Annals of Neurology*, 57(5):656–663, 2005. doi: 10.1002/ana.20452.

[299] Y. Nan, N.H. Lovell, S.J. Redmond, K. Wang, K. Delbaere, et al. Deep learning for activity recognition in older people using a pocket-worn smartphone. *Sensors*, 20(24):1–14, 2020. doi: 10.3390/s20247195.

[300] S. Parvaneh, J. Mohler, N. Toosizadeh, G.S. Grewal, and B. Najafi. Postural Transitions during Activities of Daily Living Could Identify Frailty Status: Application of Wearable Technology to Identify Frailty during Unsupervised Conditions. *Gerontology*, 63(5):479–487, 2017. doi: 10.1159/000460292.

[301] D. Rodríguez-Martín, A. Samà, C. Pérez-López, et al. Posture transition analysis with barometers: contribution to accelerometer-based algorithms. *Neural Computing and Applications*, 32:335—349, 2020. doi: 10.1007/s00521-018-3759-8.

[302] A. Atrsaei, F. Dadashi, C. Hansen, et al. Postural transitions detection and characterization in healthy and patient populations using a single waist sensor. *Journal of NeuroEngineering and Rehabilitation*, 17(1):70, 2020. doi: 10.1186/s12984-020-00692-4.

[303] J. Kubicek, K. Fiedorova, D. Vilimek, M. Cerny, M. Penhaker, et al. Recent Trends, Construction, and Applications of Smart Textiles and Clothing for Monitoring of Health Activity: A Comprehensive Multidisciplinary Review. *IEEE Reviews in Biomedical Engineering*, 15:36–60, 2022. doi: 10.1109/RBME.2020.3043623.

[304] J.C. Goldsack, A. Coravos, J.P. Bakker, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *npj Digital Medicine*, 3:55, 2020. doi: 10.1038/s41746-020-0260-4.