ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Electronics, Computer Science and Systems - DEIS

PhD in Bioengineering

Data Mining in Clinical Practice for the Quantification of Motor Impairment in Parkinson's Disease

Thesis submitted for the degree of PhD in Bioengineering - XXIV cycle
Settore Concorsuale: 09/G2
Settore Scientifico-Disciplinare: ING-INF/06

Candidate: Eng. **Luca Palmerini**
Supervisor: Prof. **Guido Avanzolini**
Co-Supervisor: Prof. **Lorenzo Chiari**
Reviewers: Prof. **Paolo Barbini**
Dr. **Paolo Bonato**
PhD Coordinator: Prof. **Angelo Cappello**

Thesis defense: April 2012

A mamma e papà

KeyWords

Data mining
Parkinson's disease (PD)
Posture analysis
Feature selection
Accelerometer
Timed Up and Go
Early detection of PD
Deep brain stimulation
Clinical subtypes of PD
Movement analysis

List of Publications

International Peer-Reviewed Journals

1. L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, L. Chiari, Feature Selection for Accelerometer-Based Posture Analysis in Parkinson's Disease. IEEE Transactions on Information Technology in Biomedicine, Vol. 15, no. 3, pp. 481-90, May 2011.

2. S. Mellone, L. Palmerini, A. Cappello, L. Chiari, Hilbert-Huang Based Tremor Removal to Assess Postural Properties from Accelerometers. IEEE Transactions on Biomedical Engineering, Vol. 58, no. 6, pp.1752-61, June 2011.

Book Chapters

1. L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, L. Chiari, A Clinical Application of Feature Selection: Quantitative Evaluation of the Locomotor Function. Lecture Notes on Communications in Computer and Information Science (CCIS), Vol 272, pp. 151-157, Ed. Springer-Verlag. In Press.

Conference Proceedings published on International Journals

1. L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, L. Chiari, Quantitative evaluation of the instrumented Timed Up and Go in Parkinson's disease. Proceedings of the XI SIAMOC congress, Ferrara, Italy. Gait & Posture, Vol. 33, Suppl. 1, p. S17, April 2011. Award "Stage for young researchers".

2. L. Rocchi, L. Palmerini, S. Mellone, F. Valzania, L. Chiari: Dual tasking during quiet stance allows an accurate classification of subjects with Parkinson's disease and age-matched controls. Proceedings of the 3rd International Congress on Gait and Mental Function, Washington D.C., USA. Parkinsonism & Related Disorders, Vol. 16, Suppl. 1, pp. S74, February 2010.

3. L. Palmerini, S. Mellone, L. Rocchi, F. Valzania, L. Chiari: Estimate of tremor-free postural properties from accelerometers. Proceedings of X Congress of the Italian Society of Movement Analysis in Clinic (SIAMOC), Alghero, Italy. Gait & Posture, Vol.30, Suppl. 1, pp. S55-S56, October 2009.

International Conference Proceedings

1. L. Chiari , I. Liepelt , L. Palmerini , J. Streffer , D. Berg , W. Maetzler Can accelerometer-based evaluation of postural function identify individuals with Enlarged Substantia Nigra Hyperechogenicity? Accepted for ISPGR 2012.

2. L .Palmerini, S. Mellone, L. Rocchi, L. Chiari, Dimensionality Reduction for the Quantitative Evaluation of a Smartphone-based Timed Up and Go Test. Proceedings of the international conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Boston, USA, September 2011.

3. L. Rocchi, L. Palmerini, G.Ganesan, L. Chiari, T. Herman, J. Hausdorff, Characterization of Parkinson's disease subtypes by accelerometer-based postural analysis: a clustering approach. Proceedings of the 6th International Posture Symposium, Smolenice, Slovakia, September 2011

4. L. Palmerini, L. Rocchi, S. Mellone, L. Chiari, Selection of Outcome Measures for the Timed Up & Go Test In Parkinson's Disease. Proceedings of the International Conference on Ambulatory Monitoring of Physical Activity and Movement, ICAMPAM, Glasgow, UK, May 2011.

5. L. Palmerini, L. Rocchi, S. Mellone, L. Chiari, Feature Selection for the Instrumented Timed Up And Go in Parkinson's Disease. Proceedings of the 1st conference on Knowledge Discovery and Information Retrieval (KDIR), Valencia, Spain, pp. 95-99, October 2010.

6. L. Rocchi, L. Chiari, P. Carlson-Kuhta, L. Palmerini, F. B. Horak: Effect of Deep Brain Stimulation on step initiation in subjects with Parkinson's disease. Proceedings of the 1st International Conference on Applied Bionics and Biomechanics, ICABB, Venice, Italy, October 2010.

7. M. Pirini, L. Rocchi, L. Palmerini, L. Chiari: The clinical outcomes of different targets in deep brain stimulation for Parkinson's disease: an interpretation based on a computational model. Proceedings of the 19th ISPGR Conference, Bologna, Italy, pp. 313-314, June 2009.

National Conferences Proceedings

1. L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, L. Chiari: Feature Selection for the Instrumented Timed Up and Go in Parkinson's Disease. Proceedings of the 2nd National Congress of Bioengineering, Torino, Italy, pp. 263-264, July 2010.

2. S. Mellone, L. Palmerini, A. Cappello, L. Chiari: A Method Based on Hilbert-Huang Transformation for Nonlinear Filtering of Tremor from Accelerometers. Proceedings of the 2nd National Congress of Bioengineering, Torino, Italy, pp. 249-250, July 2010.

3. L. Chiari, L. Palmerini, S.Mellone, L. Rocchi: Il rischio di caduta può essere misurato? Proceedings of 54th National Congress of the Italian Society of Gerontology and Geriatrics (SIGG), Florence, Italy. Giornale di Gerontologia, Vol. LVII, No. 6, pp. 307-308, December 2009.

Thesis Abstract

Recent advances in biomedical signal acquisition systems for motion analysis have led to low-cost and ubiquitous wearable sensors which can be used to record movement data in different settings. This implies the potential availability of large amounts of quantitative data. It is then crucial to identify and to extract the information of clinical relevance from the large amount of available data. In a perspective of evidence based medicine, this quantitative and objective information can be an important aid for clinical decision making. Data mining is the process of discovering such information in databases through data processing, selection of informative data, identification of relevant patterns, and interpretation of the results.

The databases considered in this thesis store motion data from wearable sensors (specifically accelerometers) and clinical information (clinical data, scores, tests). The main goal of this thesis is to develop data mining tools which can provide quantitative information to the clinician in the field of movement disorders. This thesis will focus on motor impairment in Parkinson's disease (PD). Even if systems for motion analysis of the main symptoms of PD (tremor, slowness of movements, involuntary movements) are available, techniques that can manage all the available information and extract quantitative knowledge are lacking.

Different databases were considered for this thesis. Each database is characterized by the data related to a specific motor task performed by different groups of subjects (see Table 1). The choice of the groups of subjects is related to a specific aim. Different data mining techniques were used to achieve different aims.

The data mining techniques that were used in this thesis can be divided into supervised (i.e. there are two groups of subjects to discriminate and there is information a priori on the group of each subject) and unsupervised (i.e. no a priori information, the technique is used to find groups of subjects characterized by homogeneous patterns). Feature selection is a technique which was used for every database: it is generally used to find relevant information and to discard useless or redundant data. It can improve the performance, the clarity of the outcome, and the generalizability of the results.

Regarding the first database in Table 1, the aim is to identify subjects who are at a high

			DATA MINING TECHNIQUES					
			Supervised			UnSupervised		
			Feature Selection	Classification	Regression	Feature Selection	Clustering	
DATABASES	High risk	Quiet Standing	X	X				Identify High Risk Subjects for PD
	Early-Mild PD	Quiet Standing	X	X				Characterize differences between PD and healthy
		iTUG	X	X				
	Moderate PD	Quiet Standing				X	X	Identify clinical subtypes of PD
	Advanced PD	UPDRS	X		X			Estimate severity of symptoms
	Population	Motor Task						AIMS

Table 1: Overview of the databases, the techniques and the aims of this thesis.

risk of developing PD by means of a simple postural test. The data is related to healthy adults who are considered as control (CTRL) subjects and healthy age-matched subjects who are at a higher risk of developing PD. The higher risk is based on a priori knowledge derived from medical imaging data. The initial aim can be translated into discriminating between the two classes of subjects (classification). As a result, significant discrimination of the two populations was obtained with a simple and low-cost test such as posture analysis with wearable sensors.

In the second database, data from postural and locomotor analysis of CTRL and early-mild PD subjects was available. Posture analysis was performed on quiet standing trials; locomotor analysis was performed by means of the Timed Up and Go (TUG) clinical test sensorized with an accelerometer (instrumented Timed Up and Go, iTUG). The related aim is to characterize the differences between the motor performance of PD subjects in an early-mild stage of the disease (when symptoms are not severe) and CTRL subjects. Various classification techniques were used to discriminate and describe the motor patterns of the two groups. Classification of postural and locomotor patterns was achieved with high accuracy (only 5% and 10% of the subjects were misclassified in the two motor tasks, respectively).

In the third database, data from posture analysis of PD subjects in a moderate stage of the disease is available. The aim is to identify subtypes of PD subjects with similar motor patterns (clustering). Clusters of subjects with homogeneous patterns were found which also

relate with specific clinical characteristics.

In the fourth, the population that was considered is of PD subjects in an advanced-stage of the disease who had undergone a surgical procedure to reduce their symptoms (deep brain stimulation, DBS). In this study there was a clinical need for an automatic method to estimate the severity of the symptoms in the home setting, without the evaluation of the clinical expert. The available data was recorded during standard tasks of the clinical motor evaluation of PD (Unified Parkinson's Disease Rating Scale, UPDRS). The data was fitted to the scores given by the clinical experts in order to obtain the same outcome (regression). Preliminary results provide evidence of the potential of using wearable sensors and data mining techniques to estimate longitudinally the severity of symptoms in the home setting.

The obtained results have implications both in clinical practice and in clinical research. Regarding the former, data mining tools were developed, together with a low-cost protocol for data acquisition, to allow the evaluation and monitoring (in clinical and home environments) of the motor function of PD subjects. Regarding the latter, motor patterns of PD are studied and characterized in different stages of the disease. This increases the quantitative knowledge about motor impairment in PD and about the evolution of the disease. The information extracted from the acceleration data will be used in the future by clinical experts for correlation studies with functional magnetic resonance images; this, in order to find a relation between a specific brain damage and motor impairment. Moreover it is suggested for the first time that data mining combined with instrumented posture analysis may be able to disclose preclinical signs of a high risk of developing PD.

Although the work of this thesis is focused on a specific clinical application, the proposed approach can be easily extended to other applications with the aim to extract quantitative information from datasets derived from wearable sensors (or other measurement systems for movement analysis).

The thesis is structured as follows:

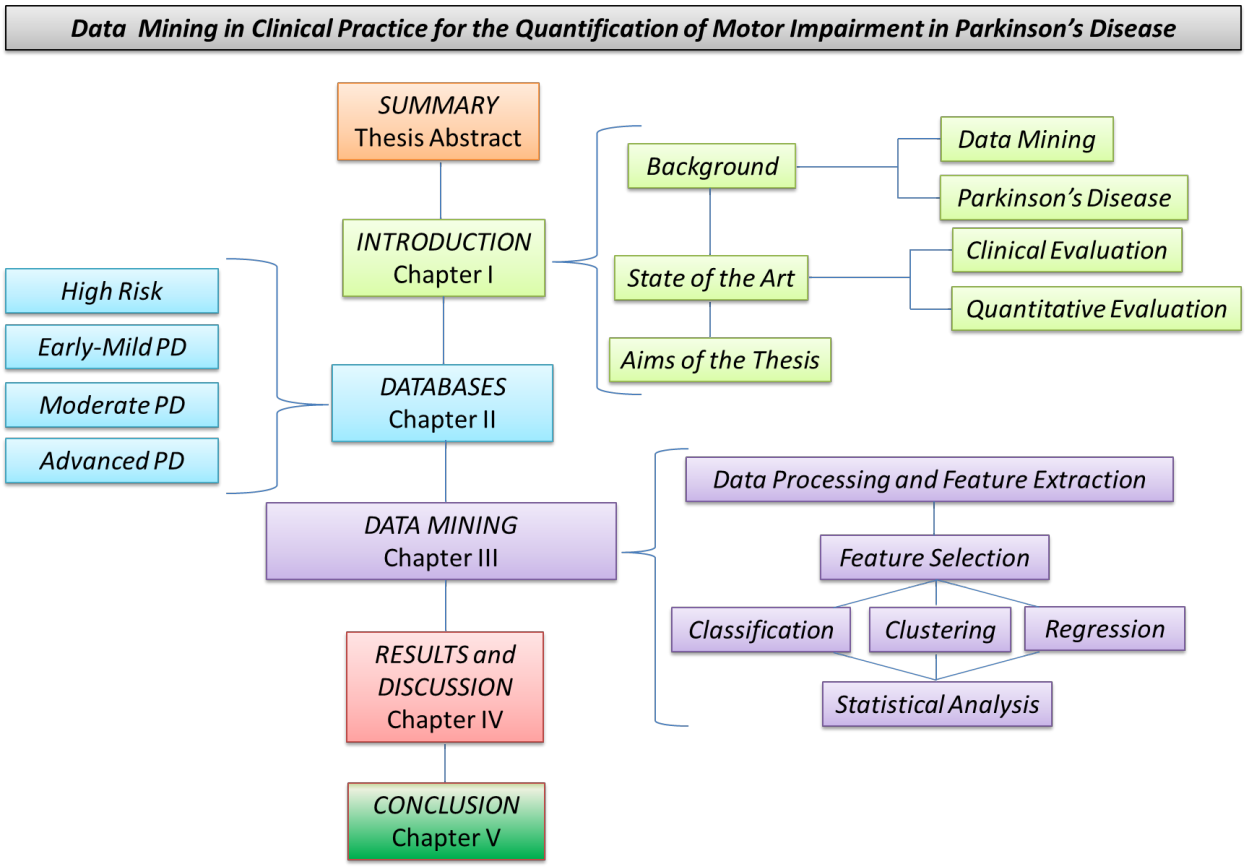

Figure 1: Thesis Outline.

Contents

Chapter 1

Introduction

Recent advances in biomedical signal acquisition systems for movement analysis have led to low-cost and ubiquitous wearable sensors which can be used to record motion data in different settings. This implies the potential availability of large amounts of quantitative data. It is then crucial to identify and to extract the information of clinical relevance from the large amount of available data. In a perspective of evidence based medicine, this quantitative and objective information can be an important aid for clinical decision making. Data mining is the process of discovering such information in databases through data processing, selection of informative data, identification of relevant patterns, and interpretation of the results. The databases considered in this thesis store motion data from wearable sensors (specifically accelerometers) and clinical information (clinical data, scores, tests). The main goal of this thesis is to develop data mining tools which can provide quantitative information to the clinician in the field of movement disorders. This thesis will focus on motor impairment in Parkinson's disease (PD). Even if systems for motion analysis of the main motor symptoms of PD (tremor, slowness of movements, involuntary movements) are available, techniques that can manage the available information and extract quantitative knowledge are lacking.

1.1 Data Mining

Originally data mining was considered as the step of data analysis in the process of knowledge discovery from databases (KDD); now the two concepts (data mining and KDD) are usually used as synonyms. Data mining can be defined as "the nontrivial extraction of implicit, previously unknown and potentially useful information from data" [1]. The process of data mining includes the steps of data processing, feature extraction, feature selection, development of an algorithm, interpretation and evaluation, as shown in Fig. 1.1.

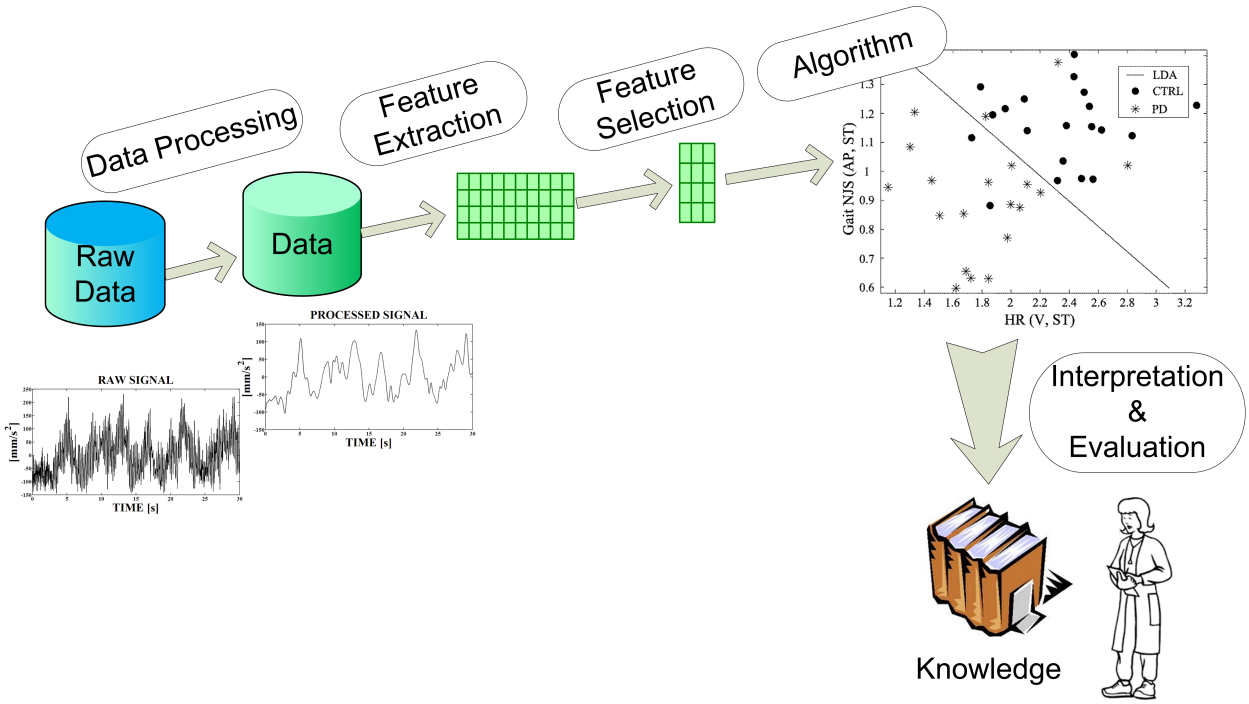

Figure 1.1: The data mining process.

Data mining is a very broad field at the intersection of statistics, machine learning, and pattern recognition.The data mining techniques that were used in this thesis can be divided, following the machine learning definitions, into supervised (e.g. there are two groups of subjects to discriminate and there is information a priori on the class label of each subject) and unsupervised (e.g. no a priori information, the technique is used to find groups of subjects characterized by homogeneous patterns). Supervised techniques can be divided into classification and regression: classification is the process of finding a function (algorithm) that discriminates data classes (e.g. healthy vs pathological). Regression is used to find a function that models numerical data values rather then discrete class labels. Unsupervised techniques are used to extract understandable patterns and associations in the data. Among them, clustering can be used to find homogeneous patterns in the data and to generate class labels. A data mining technique that can be both unsupervised and supervised is feature selection which is used to reduce to extract from the data the informative patterns discarding useless or redundant information.This technique can improve the performance, the clarity of the outcome and the generalizability of the results.

1.2 Parkinson's Disease

Parkinson's disease (PD) is a neurodegenerative disorder which affects mostly older people (1 out of 100 over 75 year [2]). It is characterized by the progressive loss of dopaminergic neurons [3]. Its main motor symptoms are:

- tremor;
- muscle rigidity (stiffness);
- bradykinesia (slowness of voluntary movements);
- postural and balance impairment.

These symptoms have a negative impact in quality of life, can severely limit motor abilities, and lead to adverse events such as falls. Pharmacological treatment based on dopaminergic medications can minimize or reduce most of the symptoms but after prolonged periods of treatment it is normal to develope motor complications like:

- dyskinesia: involuntary movements;
- dystonia: involuntary muscle contractions;
- fluctuations of symptoms severity: abrupt transitions from periods when the medication is effective (ON-periods) and periods when the symptoms are high although the subject is under medication (OFF-periods).

As it was shown here motor impairment has many different characteristics in PD: it is then of great clinical interest to detect and monitor motor impairment for PD, in order to have information on the disease progression, the effect of medications and other treatments.

1.3 Clinical Evaluation of Motor Impairment in Parkinson's Disease

In the clinical field, the evaluation of the motor impairment in PD is limited to the use of scales and questionnaires. The unified Parkinson's disease rating scale (UPDRS) [4] is the most widely used clinical scale to evaluate PD symptoms.

1.3.1 Clinical Scales

UPDRS has four sections:

1. mentation, behavior and mood;
2. activities of daily living;
3. motor examination;
4. complications of therapy

Sections 1,2, and 4 are based on the "history" of the subject (e.g. the clinician asks to the subject how severe a symptom has been in the previous week). Section 3 is also referred as the motor section of UPDRS or motor UPDRS[1]: it is based on the examination by the clinician. This rating scale evaluates the severity of PD symptoms in a 5-points scoring system (0 for no symptom and 4 for a marked severity of the symptom). Unfortunately the UPDRS, like any other semi-objective rating scale, has many limitations:

- intra and inter-observer inconsistencies;
- evaluation can depend on the experience of the clinician;
- time consuming;
- it is affected by recall bias: the scores based on the history of the patient rely on the patient's memory and perception of his own symptoms;
- poor resolution (only five possible values);
- need for the clinical expert: UPDRS evaluation cannot be used in the home setting.

In Fig. 1.2 and Fig. 1.3 the lack of objective and quantitative information can be seen in the highlighted words.

There is another clinical scale that is used to divide the evolution of the pathology in stages: it is the Hoehn and Yahr (H&Y) scale [5]. The original scale included stages 1 through 5 to indicate the increasing level of disability. To improve the low resolution of the scale, a modified H&Y scale that includes 0.5 increments was presented and has been adopted widely in the clinical field [6]. H&Y (both original and modified) scale suffers from the same problems of UPDRS. Strengths of the H&Y and UPDRS scales include their wide

[1] In this thesis the term motor UPDRS will be used.

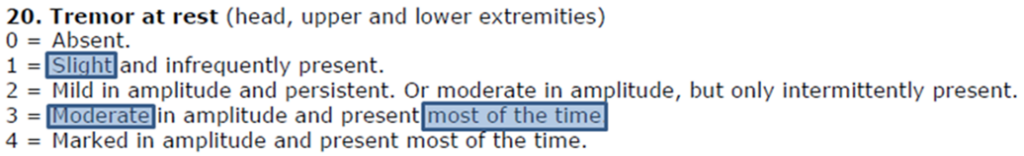

Figure 1.2: UPDRS item of section 3 assessing the severity of tremor at rest.

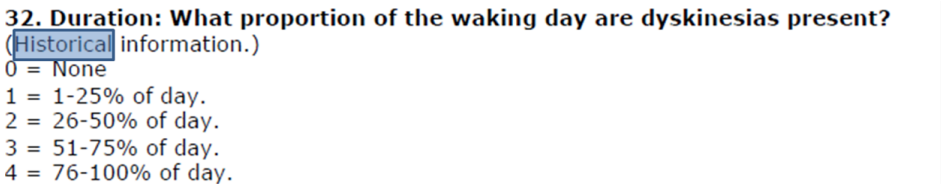

Figure 1.3: UPDRS item of section 4 assessing the duration of dyskinesias.

utilization and acceptance; on the other hand they have several disadvantages, as previously discussed.

The ideal assessment method should provide objective, quantitative measurements that could be easily translated into simple and useful information. In a perspective of *evidence based medicine*, data from measurement systems, processed by data mining techniques, can provide quantitative measures that can integrate and make more objective the available clinical information. This is the *leitmotif* of the work done in this thesis.

The different measurement systems available in literature, and the choice that was made for this thesis are presented in section 1.4; the data mining techniques that were considered in this thesis are shown in Chapter 3.

1.3.2 Clinical Subtypes

Although it is widely accepted by the neurologic community that the classical signs of PD are tremor, bradykinesia, rigidity, and postural instability, in individual PD subjects there is considerable variability in the expression and predominance of each sign. It is then of interest to detect individuals with common characteristics to define the clinical subtypes of Parkinson's disease. Different subtypes could in fact progress and respond to interventions in different ways. In the clinical literature there is no universally accepted or standardized measure for clinical classification in PD subtypes [7]; the most widely used clinical subtypes are the Tremor Dominant subtype (TD) and the Postural Instability and Gait subtype (PIGD) [8]. The division in these two groups is based on the UPDRS scores described in the previous section. Specifically two subscores (summary scores of different UPDRS items)

are defined:

- tremor score: it is a subscore calculated according to [8] as the average of the following UPDRS items: tremor by history (item 16, "activities of daily living" section), tremor at rest by examination (item 20, motor UPDRS), and action tremor by examination (item 21, motor UPDRS);

- postural instability and gait difficulty (PIGD) score: it is a subscore calculated according to [8] as the average of the following 5 items: falling by history, freezing by history, walking difficulty by history (items 13, 14, 15, "activities of daily living" section), gait by examination (item 29, motor UPDRS) and postural instability by examination (item 30, motor UPDRS).

The clinical division in the two subtypes is done in a rather oversimplified way: if the ratio between the tremor score and the PIGD score is greater than or equal to 1.5 then the subject is TD; if the ratio is less or equal than 1.0, then the subject is PIGD. The main disadvantage of the use of this clinical division in subtypes is that it is based on a ratio between items which come from a clinical scale with several limitations (see previous section). Moreover arbitrary (not based on population studies) thresholds are used and there can be subjects who are not classified in none of the two classes (these subjects are referred in this thesis as belonging to the mixed subtype, MX). The study on the "Moderate PD" database (section 2.3) deals with a quantitative evaluation of motor symptoms in order to make the clinical evaluation of PD subtypes more robust.

1.4 Quantitative Evaluation of Motor Impairment in Parkinson's Disease

1.4.1 Measurement systems

Regarding the assessment of the postural and locomotor function in PD subjects, force plates have been used extensively in the past decade; in particular force plates have been used to investigate:

- postural function in static and dynamic conditions [9, 10, 11, 12, 13, 14, 15, 16];

- gait [17] (in some studies in combination with optical systems [18, 19]);

- dyskinesia [20]

Force plates are accurate but they are expensive, require a large space, are difficult to use outside of a laboratory environment, and can be used only for a limited number of different motor task (mostly gait and posture). Developments in microelectronics have led to the mass-market availability of miniaturized wearable sensors that can provide quantitative, objective and pervasive measures of the motor function. These sensors are low-cost, easy to use and can then be easily integrated in existing clinical protocols or used in a home-setting. These are the reasons why wearable sensors (specifically accelerometers) are the measurement systems that were chosen in this thesis. These are also the reasons why wearable sensors (mostly accelerometers and gyroscopes) have been recently used in several studies for the assessment of different motor symptoms and tasks in PD:

- postural function in static condition[21, 22];
- anticipatory postural adjustments [23];
- gait [24, 25, 26, 27, 28, 29];
- turning [25, 26];
- sit to stand [30];
- stand to sit [30];
- activity recognition [31];
- tremor [31, 32, 33, 34, 35, 36, 37];
- bradykinesia and/or hipokinesia [31, 33, 38];
- akinesia (difficulty in initiating voluntary movements)[39];
- ataxia (loss of coodination) [32];
- dyskinesia [33, 40, 41];
- motor fluctuations: ON-OFF periods (see section 1.2) [42].

In these studies different number of sensors were used; in this thesis a minimal set-up was considered, when possible, for the evaluation of the motor impairment of PD; the minimum set up consists of a single accelerometer worn on the lower back. For sake of completeness other measurement systems have been used, less frequently, for the quantification of motor impairment in PD: among them optical systems [43] and force-sensitive insoles [44, 45] for

gait analysis, force/torque sensors [46] and voice recordings [47] for estimating the total score of the UPDRS.

1.4.2 Data Mining on Data from Wearable Sensors in Parkinson's Disease

As it was shown wearable sensors can provide a high amount of data to analyze for different motor tasks: data mining techniques can help to extract meaningful patterns from these available databases. Examples of these applications are the works from Keijsers et al [41, 42] where neural networks are used to predict the UPDRS scores of dyskinesia and to detect ON-OFF states; the work from Patel et al [33] where support vector machines are used to predict UPDRS scores of bradykinesia, tremor, and dyskinesia. These studies provide evidence of the feasibility and of using wearable sensors to automatically assess some of the motor symptoms in PD. On the other hand data mining techniques have never been used for the postural and locomotor (gait and transitions) assessment of PD subjects. These are key components in clinical evaluation of PD because loss of balance and gait impairments are disabling symptoms that can lead to falls and injuries. These are the reasons why the focus of this thesis will be on the use of data mining for the quantitative assessment of the postural and locomotor function in PD.

1.5 Aims of this thesis

The approach of this thesis is to provide objective and quantitative information that can augment and integrate the clinical information available in clinical problems where the traditional assessment shows its limitations. In this thesis data mining techniques are used to:

- identify high-risk subjects for developing PD (UPDRS scale is not feasible for this) by the analysis of the postural function;
- characterize subclinical differences in the postural and locomotor function in early stages of the disease (when symptoms and UPDRS scores are not severe);
- identify clinical subtypes of the disease by the analysis of the postural function (there are limited results on findings based on UPDRS);

- obtain an automated method to predict UPDRS scores of bradykinesia, dyskinesia and tremor in an advanced stage of the disease in the home setting (UPDRS traditional evaluation would require the presence of the clinical expert).

Chapter 2

Databases

In this chapter the characteristics of the databases that were analyzed are presented. To help the reader Table 2.1 is presented here with the summary properties of each database that will be then detailed in the following sections.

Database	Motor Task	Population	Conditions	Stance Position	Measurement System
High Risk	QS	15 CTRL; 21 HRISK	S-EO, S-EOF	Semi-tandem	Tri-axial accelerometer on the lower back
Early-Mild PD	QS	20 CTRL; 20 PD	EO, EC, EODT, EOF, ECF	Standardized averaged preferred position	Tri-axial accelerometer on the lower back
	iTUG		ST, DT	/	
Moderate PD	QS	35 PD	FT-EO, FT-EC, S-EO, S-EC	Feet Together, Semi-tandem	Tri-axial accelerometer on the lower back
Advanced PD	UPDRS	5 PD	/	/	9 tri-axial accelerometers

Table 2.1: Summary properties of the considered databases. QS=Quiet Standing; iTUG=instrumented Timed Up and Go;UPDRS= Unified Parkinson's Disease Rating Scale; CTRL=Control; HRISK=High Risk; PD=Parkinson's Disease; EO=Eyes Open; EC=Eyes Closed; EODT=Eyes Open Dual Task; EOF=Eyes Open on Foam; ECF=Eyes Closed on Foam; ST=Single Task; DT= Dual Task; S=Semi-tandem; FT=Feet Together.

2.1 High Risk of Developing Parkinson's Disease

The study was done in collaboration with the Center for Neurodegenerative Diseases, University of Tuebingen, Germany.

2.1.1 Background and Rationale

Neuromodulatory or even neuroprotective therapy may be available in the future for Parkinson's disease. The best effect of such therapies will be achieved when administered in the earliest as possible disease stage. However the clinical diagnosis is difficult early in the disease when the symptoms and signs may be subtle. Early biomarkers of the disease are therefore needed; these markers, however, are not yet available to a sufficient extent and quality [48]. Enlarged Substantia Nigra Hyperechogenicity (ESNH) assessed by transcranial sonography (TCS) is present in about 90% of subjects with Parkinson's Disease (PD) independent of age and disease stage [49], and may be the best risk marker for PD known to date [50]. It was hypothesized that instrumented analysis of postural function may identify high risk (HRISK) subjects for PD characterized by ESNH.

2.1.2 Population

The subjects were 21 healthy (62 ± 5 years old) subjects characterized by ESNH and therefore at a higher risk of developing PD (HRISK) and 15 age-matched healthy control subjects (CTRL) (64 ± 7 years old).

2.1.3 Protocol

The subjects were asked to stand still for 30 seconds (quiet standing, QS) for posture analysis. The subject performed a single trial in the semi-tandem foot stance with eyes open(S-EO) and a single trial in the semi-tandem foot stance with eyes open on a foam rubber support (S-EOF). The order was randomized. The semi-tandem foot stance is shown in Fig. 2.1. Feet were allowed to be externally rotated for comfortable standing, arms were flagged in self-chosen comfortable position.

2.1.4 Measurement system

A single device that includes a tri-axial accelerometer and a gyroscope (DynaPort Hybrid, McRoberts) with a sample frequency of 100 Hz was worn on the lower back by the subjects. The sensor was fixed with an elastic belt at the level of the third and fourth lumbar spine segment close to the center of mass. Only the signals from the tri-axial accelerometer were considered for the following analysis.

Figure 2.1: The semi-tandem foot stance.

2.2 Early-Mild Parkinson's Disease

The study was done in collaboration with the Department of Neuroscience, University of Modena and Reggio Emilia, Modena, Italy.

2.2.1 Background and Rationale

It is the most complete database of this thesis: acceleration data regarding both the postural and locomotor function of early-mild PD are available together with a complete clinical information. Moreover the data mining techniques (processing and features selection) initially developed for this database have then been used for the databases in sections 2.1 and 2.3. The aims of this study were to characterize i) the postural behavior and ii) the locomotor function of PD subjects in an early-mild stage. In this stage symptoms are not severe and quantitative data may help to detect subtle impairments which may be not evident from a clinical evaluation. For the postural evaluation quiet standing trials were considered in different conditions; for the locomotor evaluation (gait and transitions) a modified version of a well-known clinical test (the Timed Up and Go) was considered and will be described later.

2.2.2 Population

The subjects were 20 early-mild PD OFF medication (62 $\pm$ 7 years old, 12 males) and 20 healthy age-matched control subjects (CTRL, 64 $\pm$ 6 years old, 6 males). The OFF

condition was obtained by a medication washout (a levodopa washout of at least 18 hours and a dopamine agonist washout of at least 36 hours). The mean disease duration of the subjects was 62 ± 49 months (range: 9-170). The Hoehn & Yahr (H&Y) score of the subjects was 2.4 ± 0.25 (range: 1.5-2.5). The unified PD rating scale (UPDRS) was assessed for each subject by qualified medical staff at the department of Neuroscience, University of Modena and Reggio Emilia the same day of the experimental sessions. The average value of motor UPDRS[1] in PD subjects was 26.6 ± 7.1 out of 108 (range: 13-41). The UPDRS complete evaluation was available and the clinical subscores reported in section 1.3.2 were computed.

2.2.3 Protocol - Quiet Standing

Subjects performed quiet standing (QS) trials for posture analysis (Fig. 2.2). Subjects were asked to stand upright, barefoot, with arms crossed on the chest, looking at a visual marker (a black circle, 5 cm in diameter) placed on a wall 2.5 m in front of them. Foot placement was

Figure 2.2: Subject performing a quiet standing trial.

kept consistent over trials using the standardized averaged preferred position traced on the floor [51] (see Fig. 2.3. The subjects were tested in five different QS conditions. Descriptions of conditions, acronyms, and perturbed subsystems are reported in Table 2.2. The dual task administered in the eyes open dual task (EODT) condition consisted of a concurrent cognitive task: counting audibly backwards from 100 by 3's. In each condition, a different aspect of postural control was perturbed in order to detect a possible deterioration of a particular postural control mechanism in PD subjects. The measurement session was organized in three sequential blocks. Each block was made up of 5 consecutive trials, corresponding to

[1] see section 1.3.1.

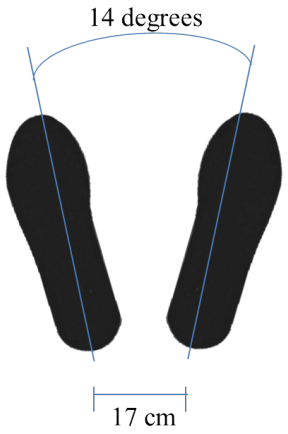

Figure 2.3: The standardized averaged preferred position.

the different conditions presented in the following order: EO, EC, EODT, EOF, and ECF. Each trial lasted 30 seconds.

Acronym	Description	Perturbed Postural Sub-System
EO	Eyes Open on a rigid surface	None
EC	Eyes Closed on a rigid surface	Visual
EODT	Eyes Open with Dual Task on a rigid surface	Attention
EOF	Eyes Open on a Foam-rubber support	Somato-sensory
ECF	Eyes Closed on a Foam-rubber support	Somato-sensory + Visual

Table 2.2: Quiet Standing Conditions.

2.2.4 Protocol - Instrumented Timed Up and Go

The Timed Up and Go test (TUG) is a well-known clinical test [52] to assess balance, mobility and fall risk in Parkinson's disease [53] and in elderly subjects [54]. It consists in rising from a chair, walking 3 meters at preferred speed, turning around, returning and sitting. Because of its simplicity and the ease in which it can be performed in the clinic, the TUG has become a widely used test. The traditional clinical outcome of this test is its total duration, which is usually measured by a stop-watch. Instrumented TUG (iTUG) [25] is a modified version of the TUG with a longer walking part (7 meters) where the subject wears inertial sensors while performing the test. In this study a single accelerometer was worn on the lower back by the subjects. The walking part is longer to permit a reliable estimation of the steps from the acceleration data [25]. A schematic representation of iTUG is shown in Fig. 2.4.

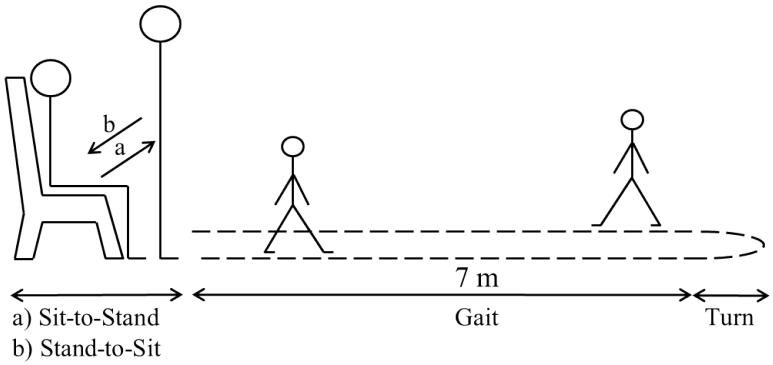

Figure 2.4: the instrumented Timed Up and Go.

The subjects performed three repetitions of iTUG trials (single task, ST) and three repetitions of iTUG trials with a concurrent cognitive task (dual task, DT), which consisted in counting audibly backwards from 100 by 3s. The subjects were instructed to stand up from a chair without armrests, walk at preferred speed on level ground, covering a distance of 7 m from the chair to the turning point, walk back and sit; turns were performed round a traffic cone (see Fig.2.5). The subjects performed the test wearing comfortable shoes.

Figure 2.5: A subject performing the instrumented Timed Up and Go.

2.2.5 Protocol - Total

Each subject, on a single day of evaluation, performed the iTUG protocol (as explained in section 2.2.4) followed by the QS protocol (as explained in section 2.2.3). The two protocols were considered separately in the following analysis in order to obtain both the evaluation of PD postural function (with QS) and of PD locomotor function (with iTUG).

2.2.6 Measurement system

Acceleration, both for the QS and the iTUG, was recorded using a single tri-axial accelerometer: McRoberts Dynaport Micromod (see Fig. 2.6). This accelerometer has a sample rate of 100 Hz, a range of ± 2 g, and a resolution of 1 mg; it was worn by the subjects on the lower back by means of an elastic waist belt, at the level of the fifth lumbar vertebra.

Figure 2.6: The triaxial accelerometer McRoberts Dynaport Micromod.

2.3 Moderate Parkinson's Disease

The study was done in collaboration with the Laboratory for Gait and Neurodynamics, Tel Aviv Sourasky Medical Center, Israel.

2.3.1 Background and Rationale

As it was discussed in section 1.3.2 there is growing evidence that PD subjects belonging to different clinical subtypes can progress and respond to interventions in different ways [8, 7]. It is therefore important to be able to accurately identify and monitor subjects based on their clinical subtype and to follow their motor symptoms over time. However there is no universally accepted clinical subtype classification and the one that is most widely used (TD-PIGD, see section 1.3.2) has many limitations. In addition to this, falls could happen

with different probability in different subtypes of Parkinson's disease: Rudzinska et al [55] showed that, as expected, PIGD subjects are more likely to falls.

The hypothesis at the base of this study was that quantitative assessment of the postural function may help to gain insight in the subtypes of PD. This study had two aims:

1. to test whether data analysis, based on a simple instrumented postural test, may give information on subtypes comparable with the clinical classification (PIGD, TD) based on the standard UPDRS;

2. to evaluate whether this tool may provide information that cannot be obtained by the clinical evaluation alone.

2.3.2 Population

Thirty-five PD subjects were tested OFF medication (with a proper washout of the medication). The subjects had an age of 66 ± 8 years, a Hoehn & Yahr score= 2.2 ± 0.4 (range: 2-3) and a disease duration of 5.4 ± 3.8 years (range: 1-16). With the method presented in section 1.3.2, 18 subjects were clinically classified into the postural instability and gait difficulty (PIGD) subtype, 14 in the Tremor Dominant (TD) subtype, and 3 were in the mixed subtype (MX) .

2.3.3 Protocol

The subjects were asked to stand still for 60 seconds (quiet standing) for posture analysis. For comparison purposes (in order to have trials of the same duration of the other studies that investigate quiet standing) the 60 seconds-trials were considered as two consecutive trials of 30 seconds each. The trials were made in 4 different conditions: 1) feet together with eyes open (FT-EO); 2) feet together with eyes closed (FT-EC); 3) feet in semi-tandem stance with eyes open (S-EO); 4) feet in semi-tandem with eyes closed (S-EC). For the feet positions see Fig. 2.7. It is worth noting that the feet-together position is different from the standardized averaged preferred position in section 2.2.

2.3.4 Measurement system

A single device that includes a tri-axial accelerometer and a gyroscope (DynaPort Hybrid, McRoberts) with a sample frequency of 100 Hz was worn on the lower back by the subjects. Only the signals from the tri-axial accelerometer were considered for the following analysis.

Figure 2.7: The feet-together (on the left) and semi-tandem (on the right) foot stance.

2.4 Advanced Parkinson's Disease

The study was done under the supervision of Prof. Paolo Bonato and Eng. Shyamal Patel during a 3-months research period at the Motion Analysis Lab, Spaulding Rehabilitation Hospital, Harvard Medical School.

2.4.1 Background and Aim

Deep Brain Stimulation (DBS) of the subthalamic nucleus of the brain is a surgical procedure which has been shown to be effective in reducing symptoms and motor complications in advanced stages of PD [56, 57]. There are various settings that must be chosen for the stimulation: frequency, voltage, sites [58]. Nonetheless the choice of these settings is now based only on the expertise of the clinician; quantitative methods to optimize the DBS settings are lacking. Moreover the effect of DBS on certain symptoms like bradykinesia is visible only in the long term period which requires a proper monitoring. There is also lack of quantitative method to optimize the medication dosage after DBS.

The aim of this study was to be able to monitor the severity of the symptoms after DBS in the home setting without the need of the clinical expert. In other words, the aim was to provide an automatic method which can provide the same evaluation of the severity of symptoms that an expert clinician would provide. Estimating the symptoms longitudinally could provide important information that could help the clinician to find the best configuration of DBS settings and medication dosage. The symptoms to estimate in this study were: tremor, dyskinesia, and bradykinesia.

2.4.2 Population

Five PD subjects in an advanced stage of the disease, who had undergone DBS surgery, were in the study . Each of them was monitored for 3 visits in the clinical settings and 3 visits at home.

2.4.3 Protocol

Subjects were instructed to perform a series of standard motor tasks that the patient performs during the motor evaluation of UPDRS. The following tasks were then considered for the analysis:

- the finger-to-nose task (reaching and touching a target);
- the finger tapping;
- the repeated hand movements (opening and closing both hands)
- the leg agility task (heel tapping);
- the quiet sitting;
- the alternating hand movements (repeated pronation/supination movements of hands).

Each motor task was performed for about 30 seconds or for a fixed number of repetitions. In the clinical visits this protocol (the series of UPDRS motor tasks) was performed before and after changes in the DBS settings by the clinician; in the home visits the protocol was performed seven times.

2.4.4 Measurement system

Each subject's movements were recorded using 9 tri-axial accelerometers (Freescale MMA7260Q). The accelerometers were placed bilaterally at the midpoint of the forearm and the upper arm, on the shank approximately 10 cm above the ankle, on the thigh approximately 10 cm above the knee and on the upper back (see Fig. 2.8).

Video recordings of each session were gathered for later review and UPDRS scoring by an expert clinician. The UPDRS scores of tremor, dyskinesia, and bradykinesia were assigned by the clinician for each limb(i.e. right and left arm, right and left leg).

In particular tremor and dyskinesia scores were assigned for each motor task; bradykinesia was assigned for the alternating hand movement task, and for the leg agility task.

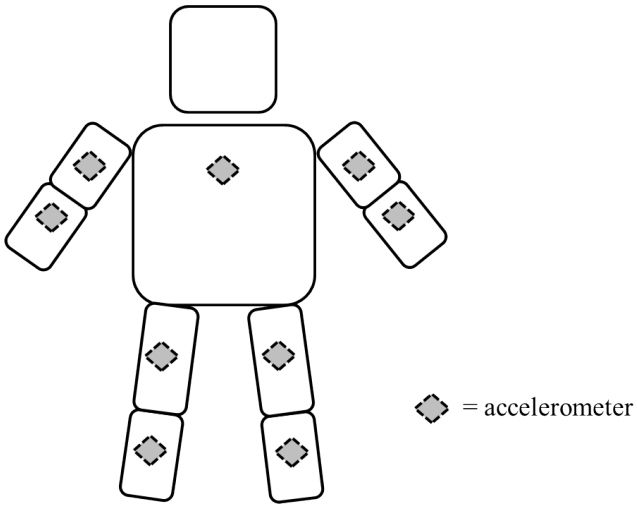

Figure 2.8: Sensor setup.

Chapter 3

Data Mining

In this chapter the data mining techniques applied to the databases are presented. To help the reader, in Table 3.1 the specific techniques that were used on each database are reported, together with the reference to the section of the chapter.

MATLAB R2009b was used for data processing, computation of acceleration-derived measures (feature extraction), feature selection, classification, clustering, and regression. Statistical analysis was performed with MATLAB R2009b and with the statistical software NCSS 2007. Weka software was used for the RReliefF method (section 3.2.3).

Database	Data Processing & Feature Extraction	# of Features	Feature Selection	Algorithm
High Risk	QS (3.1.1)	70	Wrapper with nested CV (3.2.1)	Classification (3.3)
Early-Mild PD	QS (3.1.1)	175	Wrapper with nested CV (3.2.1)	Classification (3.3)
	iTUG (3.1.2)	44		
Moderate PD	QS (3.1.1)	140	Unsupervised Correlation-Based (3.2.2)	Clustering (3.4)
Advanced PD	UPDRS motor task (3.1.3)	19	RReliefF (3.2.3)	Regression (3.5)

Table 3.1: Summary of the data mining techniques that were used in this thesis.

3.1 Data Processing and Feature Extraction

3.1.1 Quiet Standing

In this section the techniques of data processing which were applied to databases where quiet standing was analyzed (sections 2.1, 2.2, 2.3) are presented. The application of this

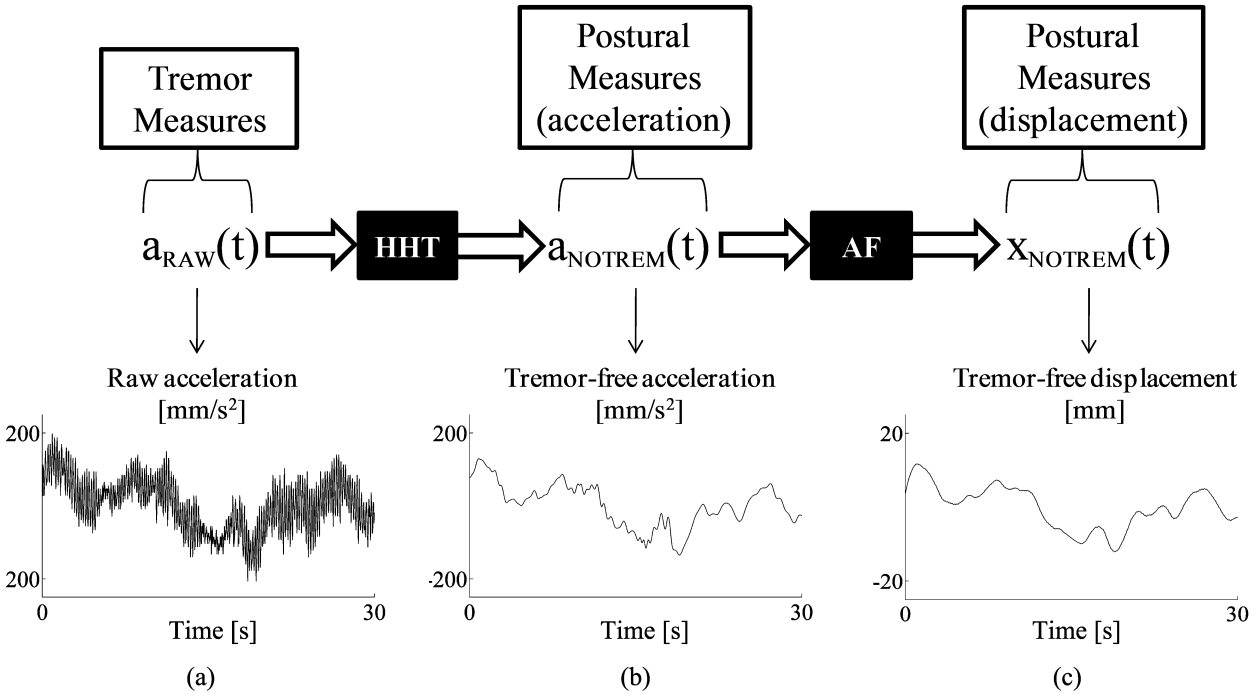

Figure 3.1: Schematic representation of the different processing stages of the acceleration signals. (a) Raw acceleration is filtered by HHT, obtaining (b) tremor-free acceleration; then the anthropometric filter (AF) is applied to get (c) tremor-free displacement.

data processing is then specific to posture analysis for Parkinson's disease by means of quiet standing trials and a single tri-axial accelerometer worn on the lower back. The acceleration signals along the two orthogonal axes of the accelerometer were considered for the analysis: the first aligned with the back/forward direction of sway and coincident with the biomechanical antero-posterior (AP) axis of the body; the second in the left/right direction and coincident with the biomechanical medio-lateral (ML) axis of the body. Three different versions of each signal were considered: raw acceleration, tremor-free acceleration, and tremor-free displacement (approximate center of mass, CoM, displacement). From each of these versions specific measures were computed (Fig. 3.1). The preprocessing of the signals was accomplished following the steps listed as follows (see Fig. 3.1).

1. Ad hoc measures were computed from the raw signals to evaluate the presence and amplitude of tremor (see Fig.3.1(a) and Table 3.2).

2. A filtering procedure based on Hilbert–Huang transform (HHT) [59] was applied to the raw signals (see Fig. 3.1(a)) in order to consider only properties related to postural control (i.e. tremor free). HHT can deal with non-stationary processes (such as tremor)

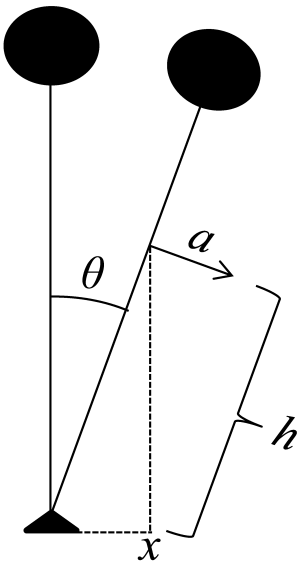

Figure 3.2: Inverted pendulum model of posture.

and nonlinear systems (such as the postural control systems) and can be used for signal artifact reduction[59]; therefore, this technique was used to isolate and later suppress the effect of tremor. Since tremor frequency in PD is usually in the band 4 – 7 Hz [60], the 0 – 3.5 Hz interval was the bandwidth isolated by the HHT procedure and then considered for computation of postural acceleration measures.

3. In order to get measures related to displacement of CoM on the horizontal plane, HHT-filtered acceleration signals underwent a low-pass filtering, with a cutoff frequency of 0.5 Hz and a static gain of $-1/g$, where g is the gravitational acceleration. The transfer function was obtained from a simple biomechanical model, in the sagittal plane, based on inverted pendulum modeling of the human body during QS (see Fig. 3.2)[61]. This assumption and the small angles of sway in quiet standing trials lead to the following equation, which relates acceleration to sway angle and CoM height

$$a(t) = h\ddot{\theta}(t) - g\sin\theta(t) \approx h\ddot{\theta}(t) - g\theta(t) \tag{3.1}$$

where a is the accelerometer output in the AP direction, θ is the sway angle with respect to vertical, and h is the height of the inertial sensor (which is assumed to be the same height as the CoM). The corresponding transfer function is:

$$H(s) = \frac{\theta(s)}{a(s)} = -\frac{1}{g - hs^2} \tag{3.2}$$

Equation 3.2 in the frequency domain can be written as:

$$H(j\omega) = -\frac{1}{g + h\omega^2} = -\frac{1}{g}\frac{1}{1 + \left(\frac{\omega}{\omega_n}\right)^2},\ \omega_n = \sqrt{\frac{g}{h}} \tag{3.3}$$

With the reasonable assumption of h being, on average, approximately 1 meter, displacement of the CoM projection on the ground ($x = h\sin(\theta)$) can be approximated as $x = \sin(\theta) \approx \theta$. Thus equation 3.3 represents the relation between the acceleration and the AP projection of CoM in the frequency domain (i.e. the frequency response of the filter obtained from the biomechanical model). Typical postural displacement measures, traditionally obtained from CoP[62], were computed from x (Fig. 3.1 and Table 3.2). The same processing was performed on the ML signals, to obtain a displacement signal in this direction. It may be considered that the aim of the procedure was not a precise estimation of the CoM, but rather the achievement of a signal reasonably approximating the characteristics of a displacement, in order to compute corresponding displacement-related measures (such as area and main direction of oscillation during QS, as described in detail in Table 3.2).

Table 3.2 summarizes the measures computed from the accelerometers in each test condition (e.g. eyes open, eyes closed). In particular, tremor measures describe the characteristics of the acceleration signal in the frequency domain, assuming that PD tremor is localized between 4 and 7 Hz [60]. Then, several postural measures were computed from tremor-free accelerations to characterize postural sway in the time and frequency domains; the mathematical definitions of these measures are parallel to those traditionally computed from the CoP in posturographic studies and used in clinical practice [63], with the exception of some measures which strictly rely on the acceleration.

Regarding postural measures computed from acceleration, five of them quantify the properties of the acceleration in the frequency domain (F50, F95, CF, FD, and Entropy, see Table 3.2) and two in the time domain (Jerk Score, Normalized Jerk Score, see Table 3.2).

In the frequency domain considering $G(f)$ as the continuous power spectrum of the acceleration signal, $G[m] = G(M\delta f)$ as the discrete power spectrum (that is the one available), δf as the frequency increment, and m as the number that detects the desired frequency ($f = m\delta f$), the k^{th} order spectral moment can be defined as:

$$\mu_k = \sum_{m\Delta f > 0.15hz} (m\Delta f)^k G[m] \tag{3.4}$$

The 0-order spectral moment μ_0 is the area under the power spectrum. As in [62] the frequencies higher than 0.15 Hz were considered, considering lower frequencies not useful to

characterize the behaviour of the postural control system.

Centroidal frequency (CF) was then defined as:

$$CF = \sqrt{\frac{\mu_2}{\mu_0}} \tag{3.5}$$

Frequency dispersion(FD) was defined as:

$$FD = \sqrt{1 - \frac{\mu_1^2}{\mu_0 \mu_2}} \tag{3.6}$$

Entropy was defined as:

$$Entropy = -\sum_{m\Delta f > 0.15hz} \frac{G[m]}{\mu_0} \log_2 \left(\frac{G[m]}{\mu_0} \right) \tag{3.7}$$

Among the time-domain measures, Jerk Score (JS) is computed as a function of the time derivative of the acceleration:

$$JS = \frac{1}{2} \int \left(\frac{da}{dt} \right)^2 dt \tag{3.8}$$

For the Normalized Jerk Score (NJS), the JS is normalized by dividing it by SP^2. The computation of NJS requires information both from acceleration, JS, and displacement, SP. It was considered as an acceleration-based measure in Table 3.2 because it mainly describes properties of the acceleration signal.

Measures derived from displacement are computed in the time domain and describe the amount and direction of sway. Among these, mean velocity (MV) is obtained through the derivative of the displacement. Both tremor and postural measures were computed for the AP and ML directions, with the exception of five postural measures which describe planar (bi-dimensional) characteristics of the displacement (SA, CEA, mSCEA, MSCEA, and |90-Mdir|, see Table 3.2).

In the "Early-Mild PD-QS" database (section 2.2) and in the "Moderate PD" database (section 2.3) for each subject, the mean values of the measures across the trials (three and two respectively) in each QS condition were considered in the following analysis selection procedure. In study 2.1 instead, for each subject a single trial was available for both QS conditions; therefore the values of the measures computed on that single trial were considered in the following feature selection procedure.

Tremor Measures	Domain	Description	Directions
Power HF	Frequency	Fraction of power of the signal for high frequencies (between 4 and 7 Hz) [%]	AP, ML
Peak HF	Frequency	Frequency of the maximum of the PSD for high frequencies (> 4 Hz) [Hz]	AP, ML
RHL	Frequency	Power ratio of the high (3.5 – 15 Hz) to low (0.15 – 3.5 Hz) frequency components (unitless)	AP, ML
Postural Measures (acceleration)	**Domain**	**Description**	
F50	Frequency	50% power frequency: frequency containing 50% of the total power [Hz]	AP, ML
F95	Frequency	95% power frequency: frequency containing 95% of the total power [Hz]	AP, ML
CF	Frequency	Centroidal Frequency: the frequency at which spectral mass is concentrated [Hz]	AP, ML
FD	Frequency	Frequency Dispersion: a unitless measure of the variability of the PSD frequency content (zero for pure sinusoid, increases with spectral bandwidth to one) (unitless)	AP, ML
Entropy	Frequency	Power spectrum entropy of acceleration (unitless)	AP, ML
JS	Time	Jerk Score: a function of the time derivative of the acceleration; fast and large variations in the signal lead to high values of this measure [mm^2/s^5]	AP, ML
NJS	Time	Normalized Jerk Score: JS is normalized by dividing it by SP^2 [$1/s^5$]	AP, ML
Postural Measures (displacement)	**Domain**	**Description**	
MD	Time	Mean Distance from center of CoM trajectory [mm]	AP, ML
RMS	Time	Root Mean Square distance from center of CoM trajectory [mm]	AP, ML
Range	Time	Range of CoM displacement (difference between maximum and minimum value) [mm]	AP, ML
SP	Time	Sway Path: total CoM trajectory length [mm]	AP, ML
MV	Time	Mean Velocity of the CoM, computed as the median value of the absolute value of the time series obtained through the derivative of the displacement [mm/s]	AP, ML
SA	Time	Sway Area: area included in CoM displacement per unit of time [mm^2/s]	planar
CEA	Time	Confidence Ellipse Area: area of 95% confidence ellipse [mm^2]	planar
mSCEA	Time	Minor Semiaxis of CEA [mm]	planar
MSCEA	Time	Major Semiaxis of CEA [mm]	planar
|90-Mdir|	Time	Absolute value of the angular deviation from AP sway of the Max Variance Direction [deg]	planar

Table 3.2: Tremor and postural measures: acronyms and brief descriptions. The last column reports the directions along which they were computed.

3.1.2 Instrumented Timed Up and Go

This technique of signal processing was applied to the data related to the iTUG test (section 2.2.4). In Fig. 3.3 the antero-posterior acceleration signals of a CTRL and a PD subject are reported as an example of the acceleration signals that were recorded during the trials.

Several temporal (including total duration of the test), coordination, smoothness, and variability measures were extracted from the acceleration signals in different directions: antero-poterior (AP), medio-lateral (ML), and vertical (V).

The iTUG was divided into four components (Fig. 2.4): Sit-to-Stand (Si-St), gait, turning, and Stand-to-Sit (St-Si). AP acceleration was used to identify postural transitions [30] and to identify heel strikes [64]; turning was marked manually during the trial by means of a remote control. Unlike [30], the time interval following the acceleration peak was also included as a part of the Si-St and the time interval preceding the acceleration peak as a part of the St-Si. The turning section was excluded from the analysis because a low consistency of the acceleration signals in that component among subjects and/or repetitions was found; this lack of consistency can be ascribed to different turning strategies requiring a more thorough analysis. For similar reasons of low consistency, measures of the Stand-to-Sit were not included.

All the measures related to the gait section were computed considering the time intervals when the subject walks on a straight trajectory, from the Si-St to the beginning of the turn and from the end of the turn to the St-Si.

The complete set of parameters is described and reported in Table 3.3. Step duration (Tstep) was computed identifying heel strikes as described in [64]. Cadence (number of steps/min) was not considered since the same information is already present in Tstep: in fact cadence can be obtained by dividing 60 (seconds) by Tstep. Phase and Phase Coordination Index (PCI) were computed according to [44].

In analogy with [27] a Normalized Jerk Score (NJS) was defined. The acceleration was band-pass filtered between 0.15 Hz and 5 Hz (zero lag fourth order Butterworth filer) to limit the effect of very slow and impulsive variations on the derivative and the integral. NJS was normalized with respect to the total duration of the movement (Sit-to-Stand or step); it was not normalized with respect to the distance because the length of the movement trajectory was not computed; the measurement is therefore not dimensionless but its measurement unit is m.

Harmonic Ratio (HR) was computed as described in [65]; acceleration was low-pass filtered at 20 Hz (zero lag fourth order Butterworth filer). Acceleration was filtered at 20 Hz also before the computation of the acceleration Root Mean Square (RMS) during Si-St.

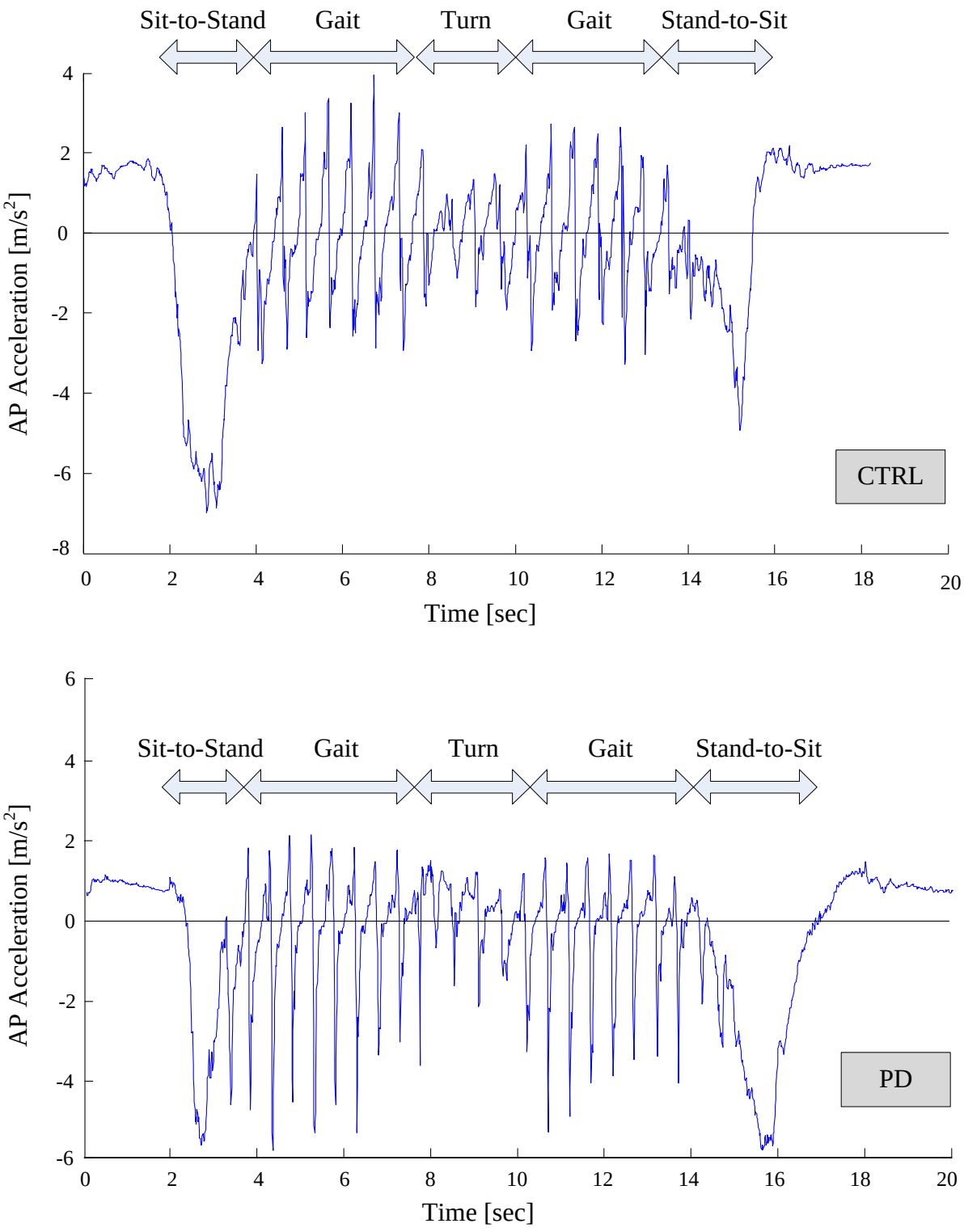

Figure 3.3: The antero-posterior acceleration signals of a CTRL and a PD subject. The different sections of the test are also reported.

ACRONYM	DESCRIPTION	DIRECTION
Tot Dur	Total duration of the Timed Up and Go test defined as the time interval between the beginning of the Sit to Stand transition to the end of the Stand to Sit transition [s]	NA
Si-St Dur	Sit to stand duration [s]	NA
Si-St RMS	Root Mean Square of the acceleration during the Sit to Stand transition [m/s^2]	AP, ML, V
Si-St NJS	Normalized Jerk Score of the acceleration during the Sit to Stand transition [m]	AP, ML, V
Gait Dur	The duration of the walking phase from the first to the last step [s]	NA
Tstep	Mean value of the step duration, computed as the time distance between two consecutive heel strikes, excluding the turn [s]	NA
Tstep STD	Standard deviation of the step duration, computed as the time distance between two consecutive heel strikes, excluding the turn [s]	NA
Tstep CV	Coefficient of Variation of the step duration, computed as the time distance between two consecutive heel strikes, excluding the turn [%]	NA
Phase	Mean value of the phase, in degrees, excluding the turn. The phase measures the step time with respect to the stride time assigning 360° to each stride (gait cycle) [degree]	NA
Phase STD	Standard deviation of the phase, excluding the turn [degree]	NA
Phase CV	Coefficient of variation of the phase excluding the turn [%]	NA
PCI	Phase Coordination Index excluding the turn. PCI measures gait coordination (i.e. the accuracy and consistency of the phase generation) [%]	NA
Gait NJS	Normalized Jerk Score of the acceleration is computed for each step of gait, i.e. between two consecutive heel strikes, and then normalized to the step duration. The mean value across the steps is considered. The turn is excluded from the computation [m]	AP, ML, V
HR	Harmonic ratios measure the rhythm of trunk accelerations during walking and can be viewed as a measure of smoothness of walking (high values are related to high smoothness) [unitless]	AP, ML, V

Table 3.3: iTUG measures: acronyms, brief descriptions, and directions. The directions can be NA = Not Applicable; AP = Antero-Posterior; ML = Medio-Lateral; V = Vertical.

For each measure, both in ST and in DT, the mean value across the three consecutive trials was used for the following feature selection procedure. All the consecutive trials were instead considered for the statistical analysis of reliability, group differences, and differences between trials.

3.1.3 UPDRS Motor Tasks

The data processing developed by Patel et al ([33, 66]) was used for the "Advanced PD" database (section 2.4). The main characteristics of the processing are reported here.

First an high-pass filter was applied to the raw acceleration signals with a cut-off frequency of 1 Hz to attenuate components associated with gross changes in the orientation of body segments[67]. Then an additional filter was applied to isolate the frequency compo-

nents of interest for estimating each symptom. Specifically, the acceleration time series were bandpass filtered with bandwidth 3 – 8 Hz for the analysis of tremor, and low-pass filtered with a cut-off frequency of 3 Hz for the analysis of bradykinesia and dyskinesia. The signals were divided manually, based on the videos of the tests, in the different UPDRS motor tasks (e.g. finger tapping, alternating hand movement). From the portion of the signal of each motor task, 20 windows of 5 seconds were randomly selected; the clinical score related to the entire motor task was assigned to each of these windows for the following estimation.

The data from the x, y and z axis of each accelerometer was combined by taking the square root of the sum of squares of each axis. Measures were computed (as in [66]) from the resulting signals, which represent characteristics of movements such as intensity, modulation, rate, periodicity, complexity, and coordination of movement. For each sensor (i.e. right/left upper arm/forearm, right/left thigh/shank) the following measures were computed:

1. the root mean square value of the de-trended acceleration signal (intensity);
2. the range of the auto-covariance function (modulation: large values of this feature were indicative of intervals of rapid movements interspersed with intervals of slow movements);
3. the dominant frequency component (rate);
4. the ratio between the energy of the dominant frequency component and the total energy (periodicity);
5. the signal entropy (complexity).

For each limb (e.g. right leg) the coordination between its proximal and distal segments (e.g. right thigh and right shank) and between opposite segments (e.g. right and left thigh, right and left shank) is characterized by three measures related to magnitude, delay and similarity. The three measures are:

1. correlation coefficient (magnitude);
2. the time lag corresponding to the peak of the cross-correlation function (delay);
3. the value of the peak of the cross-correlation function (similarity).

The UPDRS clinical scores to estimate were the tremor scores of each limb (arms and legs), the dyskinesia scores of each limb (arms and legs), and, for bradykinesia, the UPDRS scores of alternating hand movement (for arms) and leg agility (for legs).

Considering the right leg as an example, the total number of measures would be: 5 measures for each sensor on the limb (5 for right thigh and 5 for right shank), 3 measures for right/left coordination for both thigh and shank sensors (6 in total), 3 measures for proximal/distal correlation (right thigh/right shank). It is a total of 19 measures. So to estimate the clinical score related to a specific limb, 19 feautures were considered.

3.2 Feature Selection

In this thesis several features (measures) were derived from the acceleration signals in all of the considered databases; generally the number of features was higher than the number of available samples. Feature selection can be used to find the relevant features and discard the irrelevant or redundant ones from the databases. Feature selection can:

- make the algorithms faster (less time to build a classifier if there are fewer features);
- reduce or avoid overfitting (models built on several features tend to be specific to the considered sample and less generalizable);
- improve performance (irrelevant features could degrade the performance of the algorithms);
- gain more insight in interpretation of the results (it is easier to interpret the relations among a low number of features);
- avoid the *curse of dimensionality*: when the number of dimensions of the problem increases, the amount of data (samples) required for an effective result (e.g. classification, clustering) grows exponentially.

In contrast to other dimensionality reduction techniques like those based on projection (e.g. principal component analysis) or compression (e.g. using information theory), feature (subset) selection techniques do not create new variables, but select an optimal subset out of them. Thus, by preserving the original semantics of the variables, they offer the advantage of interpretability by the clinical expert. This is why feature selection was preferred in this thesis to other dimensionality reduction techniques.

As already said, feature selection is the task of choosing a small subset of features that ideally is necessary and sufficient for a target concept. In this thesis the target concept can be the discrimination between two classes (e.g. CTRL vs PD, section 2.2), the division of the data in well-structured clusters (section 2.3) or the regression of clinical scores (section 2.4).

Feature selection can then be applied to both supervised and unsupervised problems: in the supervised problem (classification), where the class labels are known beforehand, the feature selection techniques can be divided into *filter*, *wrapper* and *embedded* methods [68] depending on how the search for the best features is combined with the classification algorithm.

Filter techniques assess the importance of features considering intrinsic properties (e.g. correlation), of the data; a score is calculated, low scoring features are removed and the subset of selected features is then presented as an input to the classifier. These techniques are fast and scalable but they ignore the interaction with the classifier and this could lead to sub-optimal results. They can be univariate and multivariate; multivariate techniques should be generally preferred because they take into account feature dependencies that could degrade or improve the performance of the classifier.

In *wrapper* techniques [69] instead a search procedure in the space of possible feature subsets is defined and the generated subsets are evaluated: a specific classifier is trained and tested and its predictive accuracy is used as the evaluation. Subsets of features that, in combination with a specific classifier, are shown to be optimal (i.e. best accuracy) are selected. The search procedure is "wrapped around" the classifiers. Among the advantages of these techniques the fact that they consider both the interaction with the classifier and the features dependencies. Among the disadvantages, that the dependance from the classifier which may lead (if not taken care of) to a higher risk of overfitting, and the high computational cost, depending on the classifier and on the adopted search strategy (e.g. exhaustive search of all the possible subsets for most problems may be computationally impracticable).

In *embedded* the search for the optimal subset is embedded into the construction of the classifier; only few classifiers can "embed" these techniques (e.g. Decision trees, Naive Bayes, SVM); since the aim of this thesis was to obtain feature selection techniques which could be used with every classifier (in order to see which one performs better), this kind of techniques were not considered.

In this thesis a *wrapper* technique was used in the supervised problems of sections 2.1 and 2.2), in order to obtain the highest possible accuracy in discrimination; the disadvantages of this kind of technique (overfitting and computational cost) were limited by considering a proper procedure for validation and limiting the search strategy to subsets of three features. Then a multivariate *filter* technique was used in the supervised problem of section 2.4; the aim in this section was to estimate (regression) UPDRS scores, each of them with several possible ordered values. So, since the primary output is regression, a technique specifically design for the regression problem was used.

Feature selection in unsupervised problems (clustering) was not studied as much as in supervised problems [68]. In the unsupervised problem of section 2.3 a feature selection based on correlation was used; it can be seen as a modified version of a *filter* method based on correlation which is used in supervised settings.

3.2.1 Wrapper Feature Selection with Nested Cross Validation

In the "High Risk" and "Early-Mild PD" databases (sections 2.1 and 2.2) the number of available features (measures computed from acceleration signals) is very high with respect to the available sample of subjects. This situation of high dimensionality can lead to poor performance and overfitting as previously discussed. This is why an *ad hoc wrapper* feature selection procedure was developed to deal with this situation.

Wrapper techniques are characterized by a search procedure and by the accuracy of a given classifier which is used as the objective function to optimize. The search procedure consisted of an exhaustive search for subsets containing one to three features; the limit of three was chosen to permit a clinical interpretation of the results (it would be difficult to associate too many measures with different aspects of the disease). In addition, this limit keeps the search strategy computationally acceptable. The classifiers that were considered are the ones presented in section 3.3. Since feature selection is part of the design of the classifier, it should be performed only on the training set, in order to avoid the so-called feature selection bias in the final evaluation of the accuracy of the classifier [70]. This bias may occur when the accuracy of a classifier is estimated using all the available data (instead of the training set only). The most common solution to this problem, when the sample size is not large enough to split the data into a training set for feature selection and a testing set for accuracy evaluation, is to use a nested cross-validation procedure [69]. In this case the internal cross-validation for feature selection is repeated for every training set resulting from the external cross-validation, which is used to estimate the accuracy of the classifier. Because of the small sample size of the databases considered, a leave-one-out cross-validation (LOOCV) was implemented both for the internal feature selection ($\mathrm{LOOCV_{int}}$) and for the external evaluation of the classifier ($\mathrm{LOOCV_{ext}}$).

The whole procedure is summarized in Fig. 3.4: considering N as the sample size (number of subjects), $\mathrm{LOOCV_{ext}}$ splits the dataset into N different training and testing sets ($\mathrm{TR_i}$, $\mathrm{TS_i}$ with $1 \leq i \leq 40$); for each $\mathrm{TR_i}$, a different feature selection was performed ($\mathrm{FS_i}$, $1 \leq i \leq N$). The objective function (predictive accuracy) of each feature selection was evaluated by $\mathrm{LOOCV_{int}}$. After each $\mathrm{FS_i}$, a list of optimal subsets of features was generated. There was generally more than one optimal subset with the highest $\mathrm{LOOCV_{int}}$ accuracy. Following the typical nested

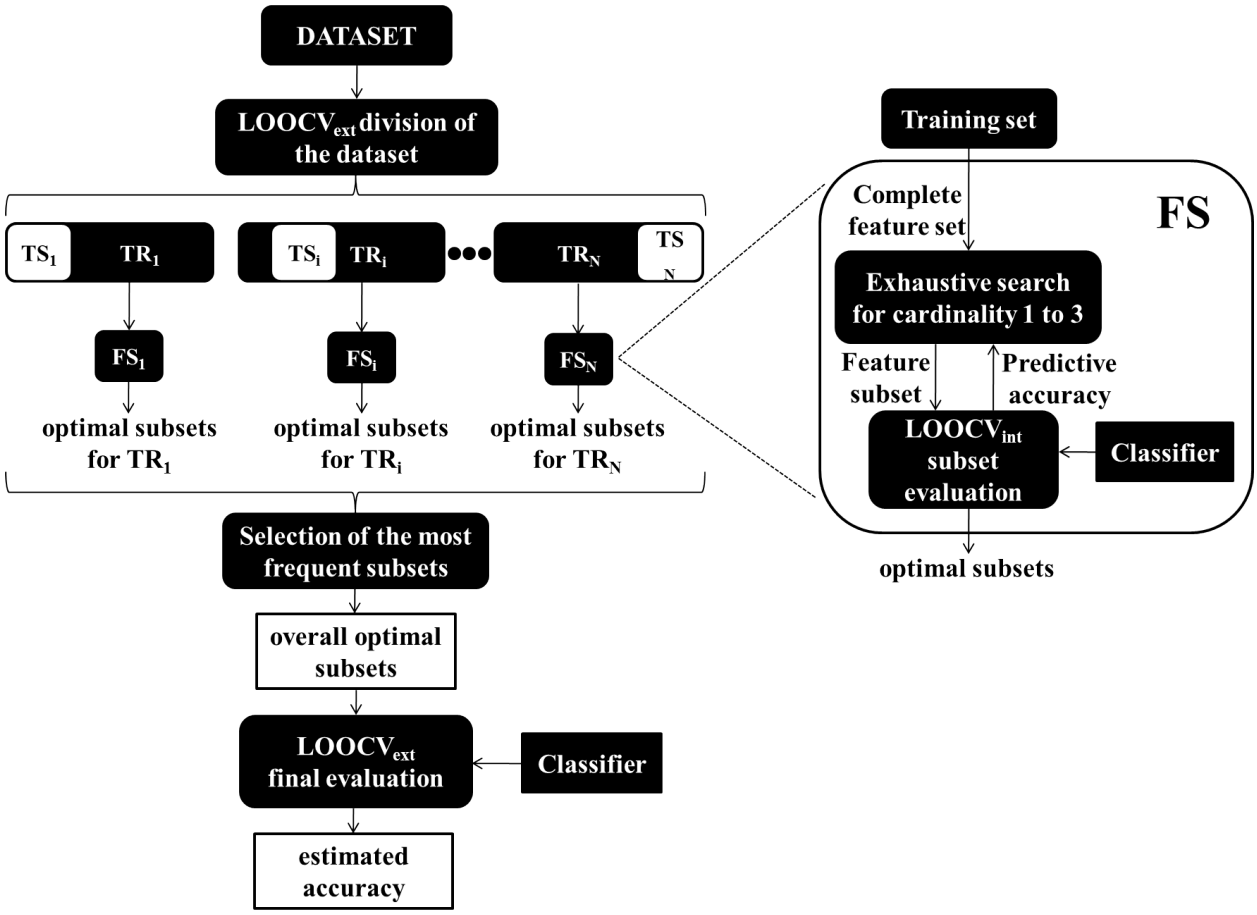

Figure 3.4: The feature selection procedure with the nested cross-validation.

procedure, TS_i should be classified from the classifier built with a single subset chosen by FS_i; however, since more than one optimal subset was found, it was not possible to make a unique choice. Moreover different FS_i's led to different lists of optimal subsets, so we decided to extract those subsets which were selected most frequently as optimal (overall optimal subsets, see Fig. 3.4). The number of times a certain subset was selected as optimal across all of the FS_i's(*selection times*) serves as an index of how robust that subset is to changes in the training set, and therefore to selection bias. *Selection times* can assume values between 1 and N (the total number of FS_i's).

Eventually, the accuracy of the classifier was computed by$LOOCV_{ext}$ for the overall optimal subsets (see Fig. 3.4) and quantified by the misclassification rate (MR), which is the proportion of incorrectly classified subjects. Subsets of different cardinality were considered separately. Confidence intervals (CI) were computed for the MR estimates following the method in [71].

This feature selection procedure was custom-written in Matlab; the whole procedure was parallelized (using the Matlab Parallel Toolbox). This could be done because the inner cross-validations are independent one of another. In this way all of the cores (ore processors) of a computer can be used simultaneously and the computational time can be drastically reduced, permitting the exhaustive search for subsets from one to three features.

3.2.2 Unsupervised Correlation-based Feature Selection

A modified version of the Correlation-based Feature Selection proposed by Hall et al [72] was developed. The central hypothesis of the original method (which is design for supervised settings) is that good feature sets contain features that are highly correlated with the class, yet uncorrelated with each other. To use it in the unsupervised setting of section 2.3 it was modified: the modified algorithm looks for sets of features that are uncorrelated with each other while optimizing the structure of the clusters built with those features (i.e. the clusters obtained from good set of features should have high inter cluster dissimilarity and high intra-cluster similarity). The structure is optimized following a clustering validity index , instead of the accuracy in classification (see section 3.4). In conclusion the aim of this modified version is to find uncorrelated features that discriminate the subjects in homogeneous clusters.

In the "Moderate PD" database a limit of R=0.5 (Pearson's correlation, see section 3.6.4) was empirically set for correlation (i.e. features which correlate more than 0.5 are considered as redundant and therefore are not used in the same subset). Subsets of one to three features were considered for the clustering algorithm for the same reasons of section 3.2.1. Cluster structures of 2 and 3 clusters were considered; more clusters were not considered because of the small sample size (i.e. with four clusters the average cluster size would be of less than nine elements which would not be very generalizable).

3.2.3 RReliefF

The RReliefF [73] algorithm in conjunction with a ranker search method was used in the regression problem of section 2.4 to select the best features and to order them by their importance. It was chosen because it is computationally simple, fast, robust to noisy data [73] and because it is specific for regression problems.

RReliefF is an extension of the ReliefF method which is a filter multivariate method for classification between multiple classes. ReliefF itself is an extension of the basic Relief method which can only be used for two-class problems.

The key idea of the basic Relief algorithm will be described here. The idea is to estimate

the quality of the features according to how well their values distinguish between the instances that are near to each other. For that purpose, given a randomly selected instance R (e.g. a CTRL subject), Relief searches for its two nearest neighbors: one from the same class (e.g. a CTRL subject), called nearest hit H, and the other from a different class (e.g. a PD subject), called nearest miss M. It updates the quality estimation W[F] for all the features F depending on their values for R, M, and H. If istance R and H (e.g. two CTRL subjects) have different values of the feature F then the feature F separates two instances of the same class which is not desirable; therefore the quality estimation W[F] is decreased.

On the other hand if R and M (e.g. CTRL and PD) have different values of F, then the feature F separates two instances belonging to different classes which is desirable; therefore the quality estimation W[F] is increased.

For each instance then W[F] is decreased by the absolute difference between the values of F for R and H (averaged on the number of the instances) and it is increased by the absolute difference between the values of F for R and M(averaged on the number of the instances). The process is repeated for all the instances and the features are ranked based on their quality estimation W[F].

The multiclass and regression extensions are based on the same concept, they use K-nearest neighbors for each class instead of one and a modified increment/decrement of the quality estimation. In the RReliefF algorith (the extension to regression) used in this thesis the number of nearest neighbors K was set to 10 as suggested by [73]. The WEKA [74] implementation of the algorithm was used.

3.3 Classification

The classifiers that were used in combination with the feature selection technique in section 3.2.1 are some of the most commonly used classifiers which indeed could be easily integrated into the specific feature selection procedure that was designed in section 3.2.1. Statistical classifiers which are easy to interpret (especially when considering 3 or less features) were considered: these are the linear and quadratic discriminant analysis [75] (LDA and QDA, respectively). the Mahalanobis classifier [75] (MC), and the logistic regression [76] (LR). For example LDA in two dimensions (two features) is represented by a discriminating line, and in three dimensions by a discriminating plane. Then classifiers belonging to the machine learning area were considered: the K-nearest neighbors (KNN)[77], and support vector machines (SVM)[78]. Also these two algorithms are easily understandable (for the aim of understandability, the linear version of the support vector machines was chosen). A brief

description of each of them is reported in Table 3.4. On the other hand classifiers which could be only interpreted as a black box (e.g. neural networks) or classifiers with several parameters to tune (which would have required a further validation) were not considered in this thesis.

CLASSIFIER	ACRONYM	DESCRIPTION	NOTES
Linear Discriminant Analysis	LDA	LDA assumes normal distribution of the data, with equal covariance matrix for both classes. The separating hyperplane is obtained by seeking the projection that maximizes the distance between the two classes' means and minimizes the interclass variance.	
Quadratic Discriminant Analysis	QDA	Similar to LDA, but it does not assume the equal covariance matrix for both classes.	
Mahalanobis Classifier	MC	It assumes a normal distribution for each prototype of a certain class and the feature vector is assigned to the nearest prototype according to Mahalanobis distance.	
Logistic Regression	LR	It aims to translate the information from continuous independent features to the [0, 1] interval which contains the probability associated with the dichotomous choice outcomes (0, 1).	The dichotomous choice outcomes (0,1) are, for example, CTRL and PD.
K-Nearest Neighbors	KNN	It assigns an unseen point to the dominant class among its k nearest neighbors within the training set. These neighbors are obtained using a distance metrics.	Euclidean distance and k=3 were chosen.
Support Vector Machines	SVM	To identify classes, it uses a discriminant hyperplane which maximizes the margins, i.e. the distance from the nearest training points.	The version with linear boundaries was used.

Table 3.4: Classifiers: acronyms and brief descriptions.

3.4 Clustering

K-means clustering was used [79]: this algorithm defines the centroid of a cluster as the mean value of the points within the cluster. This is how the algorithm works: first, it randomly selects k samples (subjects) as the initial centroids: each remaining sample is assigned to the cluster of the nearest centroid (as a distance metric, the squared Euclidean Distance was used for this thesis). Then the algorithm for each cluster computes the mean of the samples and considers this as the new centroid. Iteratively all the samples are re-assigned to the nearest centroid and the procedure goes on until the assignment does not change anymore (i.e. the clusters formed in the current round are the same as the ones formed in the previous round). It can be noted that the result of the algorithm depend on the initial random choice of the centroids; this is why it is better to run K-means clustering several times in order to obtain a stable result. For the work in this thesis it was run 10 times.

Moreover the K-means algorithm, which is based on distance computation, is sensitive

to the different scales of the features ([7]). This is why the values of the acceleration-derived measures were normalized (the mean was subtracted and the result was divided by the standard deviation) before being used in the clustering algorithm.

The clustering results were assessed by the Silhouette value [80] which is a measure of the statistical validity of the structure of the clusters. Silhouette values range from -1 to 1 and measure how well each sample (subject) has been clustered by comparing its dissimilarity within its cluster to its dissimilarity with the samples in the other clusters. According to the Silhouette value, clustering structures are to be preferred when the dissimilarity intra-cluster is low (high homogeneity of subjects belonging to the same cluster) and the dissimilarity inter-clusters is high (subjects belonging to different clusters have different characteristics). An average silhouette value was computed (across all samples); Kaufman and Rousseeuw[81] proposed to interpret values higher than 0.7 and in the range 0.51 - 0.7 as markers of a strong and a reasonable structure captured in the data, respectively.

3.5 Regression

To estimate the values of clinical scores for the "Advanced PD" database in section 2.4, a regression technique was used on the first five features[1] selected by the RReliefF method (section 3.2.3): the regression technique that was chosen is the algorithm of Random Forest (RF) by Breiman et al [82]. It is an ensemble of weakly correlated decision trees that generate an output as an aggregate result of predictions by individual trees. RF introduces an additional level of randomness to bagging by training individual trees using a randomly selected subset of features. RF has been shown to outperform several other techniques while being robust against overfitting [82]. Random forests were used to provide continuous estimates of the clinical scores between 0 and 4 for each of the 20 windows of the corresponding motor task.

To measure the goodness of the estimates, the root mean square error (RMSE) was used: it measures the degree of difference between the estimates and the values of the scores given by the clinician. It was used to provide an index of the performance of random forests in estimating the values of each clinical score.

$$RMSE = \sqrt{\sum_i (estimate(i) - value(i))^2} \tag{3.9}$$

[1]The limit of five was set empirically; future work should assess how the results may change by varying this limit

The estimate of the scores was subject-based: so, in order to compute the RMSE, the available data of the considered subject was randomly divided: 50% for the training and 50% for the test. Then the RF was trained on the training set and RMSE was computed on the test set. This procedure was repeated ten times and the RMSEs were averaged in order to obtain the final RMSE.

3.6 Statistical Analysis

3.6.1 High Risk

Two-sample two-tailed T-Tests were performed to detect differences in the values of acceleration-based measures between the two groups: High risk and PD subjects.

3.6.2 Early-Mild PD - Quiet Standing

Two-sample two-tailed T-Tests were performed to detect differences in the values of acceleration-derived measures between the two groups: CTRL and PD subjects. A correlation analysis was performed between different acceleration-derived measures and between acceleration-derived measaures and clinical information. Pearson's correlation coefficient was considered for correlation analysis, in order to find linear correlations. The clinical information consists of:

- the clinical subscores defined in section 1.3.2;
- the total score of motor UPDRS.

3.6.3 Early-Mild PD - Instrumented Timed Up and Go

In iTUG there is an addition in the statistical analyses with respect to the QS: in QS only the average of the measures across the three consecutive trials was considered; here instead, the three trials were considered to assess the Test-Retest reliability of the acceleration-derived measures and to find group differences (between CTRL and PD).

Test-Retest reliability has been assessed by means of Intraclass Correlation Coefficient [83] (ICC). The ICC (3,1) was chosen for the analysis of reliability. Then a repeated-measures ANOVA was used to identify the differences between PD and CTRL group with the three consecutive trials as the *within* factor and with group as the *between* factor.

Pearson's correlation analysis was performed between different acceleration-derived measures (considering all the trials), and between clinical information and acceleration-derived measures (considering the mean values).

The clinical information consists of:

- the clinical subscores defined in section 1.3.2;
- the total score of motor UPDRS;
- the disease duration;
- the gait/posture subscore [25];
- the rigidity subscore [25];
- the bradykinesia subscore [25].

3.6.4 Moderate PD

A Pearson's correlation analysis was performed between different acceleration-derived measures in order to find linear correlations.

Chapter 4

Results and Discussion

In this chapter the results of data mining applied to the databases are shown and discussed.

4.1 High Risk

4.1.1 Results

In this study only the best subset (and the corresponding classifier) selected by the feature selection procedure will be shown. The best subset was the one made of:

- frequency dispersion in the medio-lateral direction in the S-EOF condition;
- F50 in the antero-posterior direction in the S-EOF condition.

The corresponding classifier is the QDA(Quadratic Discriminant Analysis) (see section 3.3); this classifier, built on the two selected measures, has a 16.6% of misclassification rate. The subset is robust to changes in the training set: in fact it has a *selection times* of 35 out of 36. Other subsets of two features (with different classifiers) performed with a misclassification over 20% and therefore are not shown. Similarly, considering subsets of three features did not improve the misclassification rate and therefore they were not considered. The discrimination result is visible in Fig. 4.1. The specificity of the classifier was 86.7% and the sensitivity was 81%. While F50 AP values showed no group differences, FD ML values in HRISK were significantly lower than in CTRL subjects (Fig. 4.2).

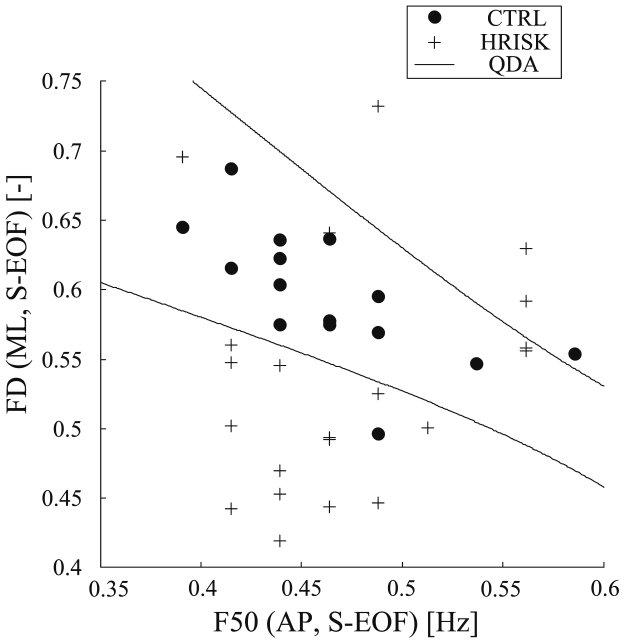

Figure 4.1: Plot of the values of F50(AP, S-EOF) and FD(ML, S-EOF) for each subject of the database. The quadratic discriminant classifier is also displayed.

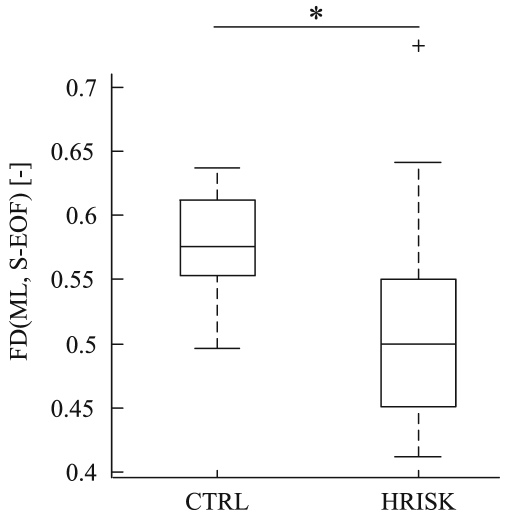

Figure 4.2: T-Test results of FD(ML, S-EOF) values between the CTRL and PD group. HRISK subjects show significantly (*, $p<0.05$) lower values than HRISK subjects.

4.1.2 Discussion

The limitations of this study are the small sample size and the fact that it is not known how many subjects at a high risk of developing PD will eventually become PD subjects. This last limitation could be overcome in the future since the follow-up data of this subject will be available in the next years. Other limitations are the fact that no actual PD subjects were considered (for comparison purposes with both HRISK and CTRL) and that the standardized foot placement (as in the database of section 2.2) was not used in the protocol, thus making the effect of the semi-tandem stance difficult to analyze. However, given all these limitations, this result suggests for the first time that instrumented posture analysis may be able to disclose preclinical signs of a high risk of developing PD.

4.2 Early-Mild PD - Quiet Standing

4.2.1 Selected Features and Performance of the Classifiers

In Table 4.1 a summary of the results of the feature selection procedure is reported, with the subsets of three measures which were selected more frequently as optimal subsets (overall optimal subsets, see section 3.2.1). For each classifier, subsets of three features selected more than 20 times are shown. The number of times a specific subset was selected (*selection times*, see section 3.2.1) out of the 40 different feature selection procedures is reported in Table 4.1, as well as the corresponding $\text{LOOCV}_{\text{ext}}$ final accuracy estimates (misclassification rate, MR). Subsets of three features, with an MR of at most 7.5%, were more accurate than subsets of fewer features; in fact for all classifiers, the best MR for pairs of measures was 10% and the best MR for a single measure was 17.5% (not shown). It can be observed that subsets with the same MR may have different (*selection times*: when this happens subsets with higher (*selection times* should be preferred because they show a higher robustness with respect to feature selection bias, as discussed in section 3.2.1.

All classifiers performed at the same misclassification rate (5%), even on different subsets, except KNN, which has a misclassification rate of 7.5% in all its best subsets. *Selection times* vary among subsets within the same classifier, and between different classifiers. The highest *selection times* value (36) is found for a subset chosen by the feature selection procedure based on MC, consisting of i) Power HF in ML direction during EODT condition; ii) FD in AP direction during EOF; iii) Range in AP direction during ECF.

In Fig. 4.3 boxplots of these three measures for CTRL and PD groups are shown; a statistical difference was detected between the two groups for FD and Power HF. Although

CLASS	OVERALL OPTIMAL SUBSETS Measure (direction, condition)			SELECTION TIMES (out of 40)	MR % [CI]
LDA	Power HF(ML,EODT)	Entropy(AP,EODT)	F50(AP,EOF)	26	5 [1.4-16.5]
QDA	Power HF(ML,EODT)	CF(AP,EODT)	Range(AP,ECF)	25	5 [1.4-16.5]
	Power HF(ML,EODT)	FD(AP,EOF)	Range(AP,ECF)	24	5 [1.4-16.5]
MC	Power HF(ML,EODT)	FD(AP,EOF)	Range(AP,ECF)	36	5 [1.4-16.5]
	Power HF(ML,EODT)	CF(AP,EODT)	Range(AP, ECF)	21	5 [1.4-16.5]
LR	JI(AP,EO)	F95(AP,EODT)	RHL(ML,EODT)	30	5 [1.4-16.5]
KNN	Entropy(AP,EODT)	RHL(ML,EODT)	Entropy(ML,EOF)	27	7.5 [2.6-19.9]
	Power HF(ML,EO)	Power HF(ML,EC)	Power HF(ML,EODT)	22	7.5 [2.6-19.9]
	Power HF(ML,EO)	Power HF(ML,EODT)	mSCEA(EOF)	22	7.5 [2.6-19.9]
SVM	F95(AP, EO)	Power HF(ML,EODT)	|90-Mdir|(EOF)	31	5 [1.4-16.5]

Table 4.1: Classifiers performance. For each classifier the overall optimal subsets are reported with the relative misclassification rate (MR) and the confidence intervals (CI). Only subsets with a *selection times* higher than 20 (out of 40) are reported.

no statistical difference was found for Range, its interaction with the other two measures acts to improve the classifier discriminative ability.

In Fig. 4.4 the values of these 3 measures are shown in a 3D view; a good separation between the two groups is graphically detectable. It can further be noted in Table 4.1 that for 3 classifiers (QDA, MC, SVM) the best (in terms of highest *selection times*) subsets consist of: i) a tremor-related measure, computed from the signal along the ML direction during EODT condition; ii) a postural measure in the frequency domain from the AP signal in EOF or EO or EODT; and iii) a postural measure related to CoM displacement in EOF or ECF.

4.2.2 Relation between Tremor Features and Clinical Measures

A positive correlation (R= 0.75, p=$1.3 \cdot 10^{-4}$) was found between Power HF in ML direction in EODT condition and the tremor subscore (see section 1.3.2). A similar correlation (R= 0.47, p=0.038) was found with RHL in ML direction in EODT condition. These similar correlations can be explained by the fact that the two measures highly correlate with each other, especially in the CTRL group (R=0.96, p=$3 \cdot 10^{-11}$ in CTRL, R=0.71, p=$5 \cdot 10^{-4}$). These two measures can then explain the effect of tremor on the postural function; Power HF seems to be the most adequate between the two because it is chosen more frequently in the best subsets, it correlates more with the clinical tremor score, and because RHL, being a ratio, is more sensitive to outliers (as it can be seen in Fig. 4.5).

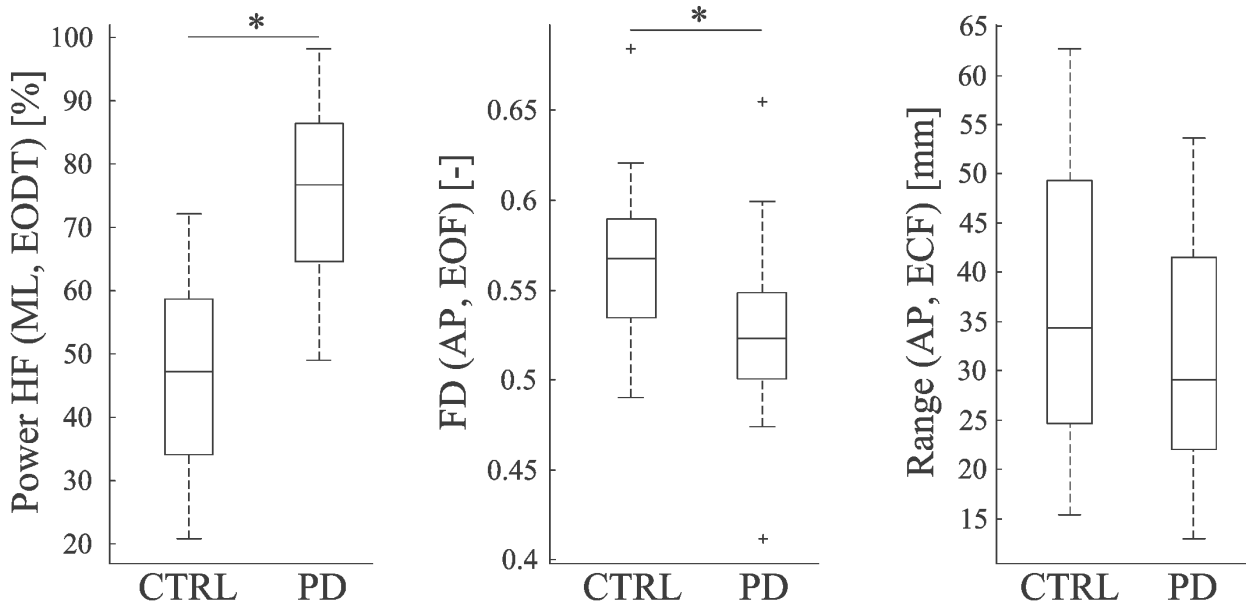

Figure 4.3: Boxplots of the 3 measures selected for the Mahalanobis classifier with the overall highest *selection times*: Power HF in ML direction during EODT condition, FD in AP direction during EOF, and Range AP during ECF. Significant differences between the CTRL and PD groups are indicated with * ($p<0.05$).

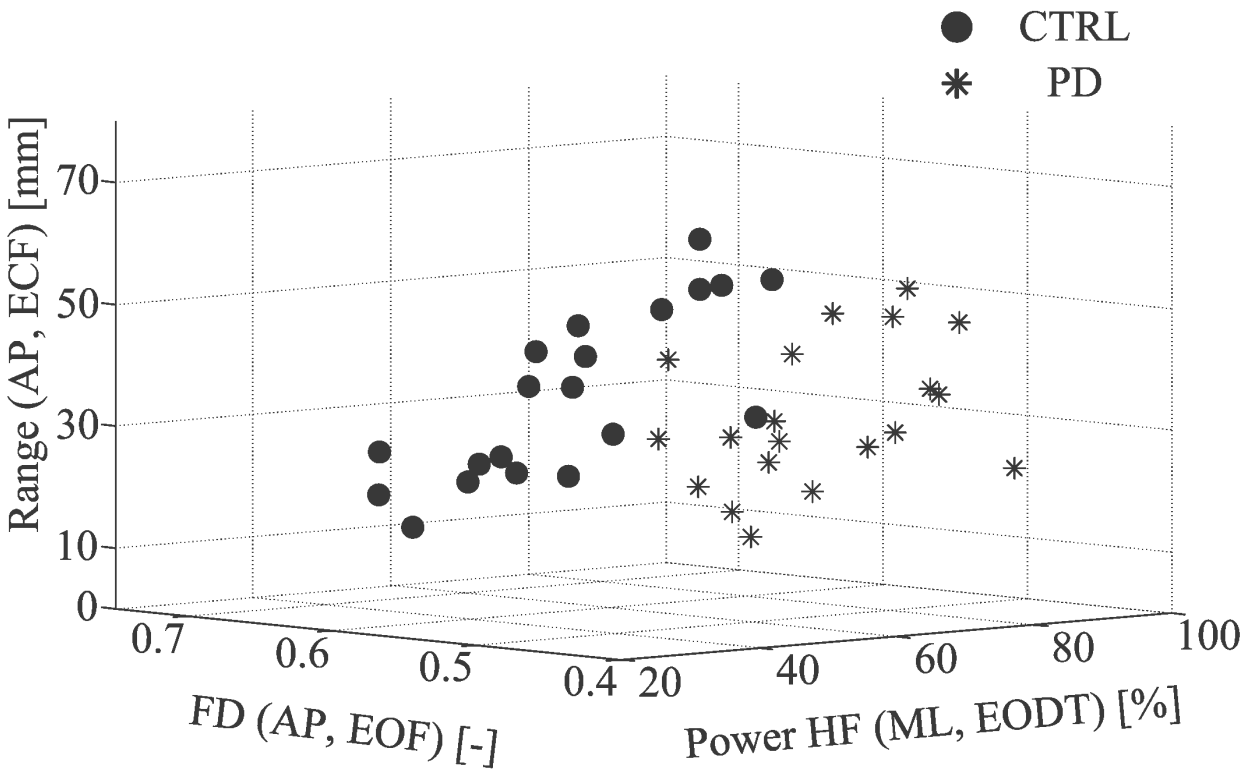

Figure 4.4: Three-dimensional view of the values of the three measures selected for the Mahlanobis classifier with the overall highest *selection times*.

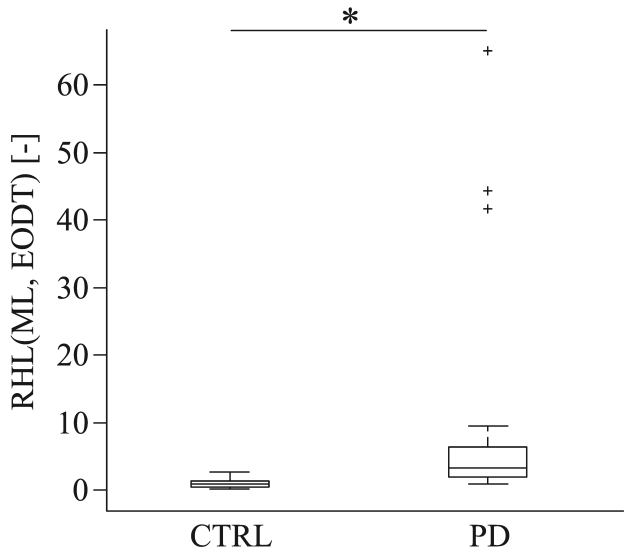

Figure 4.5: The boxplots of the values of RHL(ML, EODT) for CTRL and PD groups are shown; a significant difference ($p<0.05$) was detected between the two groups for FD and Power HF. Outliers in the PD group are clearly visible.

4.2.3 Relation between Postural Features and Clinical Measures

None of the selected features showed a significant correlation with the total score of the motor UPDRS score, with the exception (near the limits of significance) of the angular deviation from AP sway, |90-Mdir|, in EOF condition: this measure in fact has a significant positive correlation ($R= 0.41$, $p= 0.017$), but only after the removal of one outlier.

4.2.4 Discussion on Selected Features

Obtaining good results with only three measures shows that several redundant and irrelevant features were in the dataset. The proposed feature selection hence is useful to:

1. optimize the experimental protocol and reduce the cost of data acquisition;
2. reduce the computational costs;
3. improve clinical understanding of the results;
4. help in data visualization (see Fig. 4.4);
5. improve classification accuracy.

For this last purpose the similar performance of different classifiers suggest that in the present dataset the choice of the classifier is not critical. Therefore in similar circumstances, attention should be focused on the feature selection procedure rather than on building new or complex classifiers. It is noteworthy that the best features (in subsets with the highest *selection times*) selected by half of the classifiers have common characteristics representing the same aspects of postural control. The similar subsets share: a tremor-related measure, a postural measure in the frequency domain, and a displacement-related postural measure. These results confirm that all the different kinds of measures that were proposed (tremor measures, postural measures computed on acceleration and displacement) are essential, and can detect different characteristics of postural behavior of subjects with PD in the early stage.

The best combination of classifiers and selected measures highlighted by this work is the Mahalanobis classifier and the set composed of:

- Power HF (Power of the signal for high frequencies) in ML direction during dual task (EODT);
- FD (Frequency Dispersion) in AP direction with the subjects standing on foam rubber (EOF);
- Range in AP direction during ECF condition.

Power HF positively correlates with the UPDRS tremor score, and its values in PD subjects are significantly higher than its values in CTRL subjects. This measure can identify and objectively quantify tremor characteristics in PD subjects when clinical signs of tremor are not severe (tremor score: 0.74 $\pm$ 0.45 out of 4, range: 0 - 1.375). One possible reason why Power HF was selected in EODT condition is that in stressful conditions (such as dual tasking) PD subjects usually show higher tremor amplitude [60].

FD values are significantly lower in PD than in CTRL subjects, being a sign of a bit more regular and less variable sway. No statistical difference was shown for the remaining measure, Range, which exemplifies a feature that is not able to discriminate by itself but is important for the classification accuracy because of its interaction with the other measures.

From a clinical point of view, PD subjects in the present group did not show pronounced signs of postural impairments. In fact items 28 (posture) and 30 (postural stability) of motor UPDRS had average values of (1.4 $\pm$ 0.9 out of 4) and (0.9 $\pm$ 0.3 out of 4), respectively. Nonetheless quantitative postural measures were able to add discriminative power to tremor-related measures. A positive correlation was found between the motor UPDRS and the angular deviation from AP sway (absolute value) in the EOF condition: the higher the

motor UPDRS, the more the sway deviated from the AP direction, towards ML. A possible explanation is that more pronounced motor impairments in PD subjects are reflected by an increased postural sway in the ML direction. This is in agreement with the results by Rocchi et al [12] who found an increased ML sway for PD subjects, both ON and OFF levodopa, compared to CTRL subjects. Further supporting this interpretation, increased ML sway has also been associated with high fall risk in the elderly [84].

Another important outcome is identifying which quiet standing conditions are capable of disclosing postural differences between PD and CTRL subjects. Interestingly, the condition with eyes closed (EC) does not qualify, since it was not chosen by any classifier as part of the subsets with the highest *selection times*. This study found that three conditions are enough to obtain good accuracy in PD/CTRL classification, even if patients are in the early stage of the disease. From a clinical point of view, it would thus be possible to shorten the postural test protocol accordingly. In addition, it is interesting to note that, for early stage subjects, when clinical signs are not so evident, just a small perturbation of the sensory channels or an attentional overload may create the conditions which differentiate between PD and CTRL subjects.

The eyes open dual task (EODT) condition was shown to be particularly sensitive to the disease, since it was selected in all reported subsets. It is worth noting that, when subsets of three features from the same condition are considered, EODT provides the best performance. We found, in fact, a subset exclusively composed of EODT features which has a misclassification rate of 7.5% and *selection times* of 9 (it is not shown in Table IV since only subsets with *selection times* greater than 20 are reported). This EODT subset, selected by the LR classifier, is made of JI AP, F95 AP, and RHL ML. Thus a good discrimination can be achieved considering the EODT condition alone, even if this subset is less robust (lower *selection times*) than the ones presented in Table IV. This result further confirms the importance of dual tasks in differentiating between the postural behavior of PD and CTRL subjects [14].

4.2.5 Impact on Parkinson's Disease

In summary, quantitative measures computed from accelerometers seem to be able to identify postural and tremor characteristics of PD subjects even when, in an early-mild stage of the disease, these may not be evident from a clinical evaluation. The potential impact of this result will be in both clinical practice, since a low-cost and accessible method for data acquisition is introduced, and in clinical research, since the proposed protocol allows an easy monitoring of the postural function in subjects with PD.

The proposed test protocol is potentially usable in home and clinical environments to analyze the disease's progression and fluctuations, to monitor PD subjects after therapy, or to evaluate the follow-up of a rehabilitative procedure. Since PD symptoms may vary throughout the day and from day to day, the timetable for administering the medication is very important: the proposed protocol could be used to define and control this timetable in order to optimize the effect of the medication on postural function. Moreover, this protocol could also be used to test the effect of physical exercise and physiotherapy on balance impairments in PD subjects. From this perspective the reliability and sensitivity of the selected features with respect to the dynamics of the disease need to be addressed next.

4.2.6 Limitations and Future Developments

The lack of comparison of acceleration-derived measures with a gold standard (force plates for postural evaluation) is a limit of the present study which can be overcome by future experiments.

The relatively small sample size is a limitation, reflected by the large confidence intervals in Table 4.1. This problem could be overcome by widening the presented dataset with PD subjects of comparable disease stages analyzed with the same protocol. It remains also to be seen how specific the selected optimal measures are for PD, since other pathologies might be characterized by similar postural impairments. On the other hand, it is worth mentioning that even if the presented subsets are optimal for classifying early-mild PD, there is no guarantee that they would be optimal for monitoring the disease (i.e. sensitive to postural changes during the progression of the disease) or for detecting changes after a medical treatment. In this context, a follow-up of the study on the same subject will be done to evaluate the performance of selected measures over time.

4.3 Early-Mild PD - Instrumented Timed Up and Go

4.3.1 Properties of the acceleration-derived measures

Mean value and standard deviation of each measure (considering the three consecutive trials) are reported in Table 4.2 for both PD and CTRL groups in ST and DT conditions. Table 4.2 also shows Test-Retest reliability and statistical significances ($p<0.05$) of the ANOVA repeated-measures for the *within* (consecutive trials) and the *between* (group) factors. Regarding the Test-Retest reliability, in 4.2 all ICCs are reported along with their 95% confidence interval; negative values, due to a negative average covariance among items

Parameter	Control subjects				Parkinsonian Patients			
	Single Task		Dual Task		Single Task		Dual Task	
	Mean (STD)	ICC(3,1) [LB-UB]	Mean (STD)	ICC(3,1) [LB-UB]	Mean (STD)	ICC(3,1) [LB-UB]	Mean (STD)	ICC(3,1) [LB-UB]
Tot Dur[a,c]	16,08 (2,42)	0.89 [0.74-0.95]	18 (3,8)	0.89 [0.75-0.95]	17.54 (2.95)	0.81 [0.65-0.91]	20.12 (4.39)	0.88 [0.78-0.95]
Si-St Dur[a]	1,24 (0,23)	0.47 [0.2-0.71]	1,36 (0,38)	0.49 [0.23-0.73]	1.3 (0.46)	0.75 [0.55-0.88]	1.57 (0.59)	0.52 [0.26-0.75]
Si-St RMS AP	2,67 (0,53)	0.74 [0.53-0.87]	2,52 (0,59)	0.67 [0.44-0.84]	2.43 (0.53)	0.63 [0.4-0.82]	2.32 (0.46)	0.46 [0.2-0.71]
Si-St RMS ML	0,68 (0,3)	0.38 [0.11-0.65]	0,62 (0,19)	0.63 [0.39-0.81]	0.59 (0.22)	0.55 [0.29-0.77]	0.54 (0.2)	0.18 [0-0.48]
Si-St RMS V[b,d]	1,77 (0,46)	0.61 [0.36-0.8]	1,65 (0,55)	0.85 [0.72-0.93]	1.51 (0.4)	0.78 [0.6-0.9]	1.3 (0.38)	0.63 [0.4-0.82]
Si-St NJS AP	11,52 (6,71)	0.39 [0.11-0.66]	13,98 (12,22)	0.39 [0.12-0.66]	10.74 (10.42)	0.66 [0.44-0.83]	16.39 (17.96)	0.25 [0-0.55]
Si-St NJS ML	5,9 (7,18)	0.39 [0.12-0.66]	6,43 (6,15)	0.61 [0.37-0.8]	5.02 (6.09)	0.46 [0.2-0.71]	7.13 (6.55)	0.27 [0-0.57]
Si-St NJS V	11,23 (7,32)	0.43 [0.15-0.69]	13,48 (13,28)	0.39 [0.12-0.66]	9.88 (9.78)	0.54 [0.29-0.76]	14.5 (13.86)	0.15 [0-0.46]
Gait Dur	7,79 (1,59)	0.84 [0.59-0.94]	9,16 (2,55)	0.93 [0.82-0.97]	8.72 (2.18)	0.63 [0.37-0.81]	10.39 (3.05)	0.82 [0.66-0.92]
Tstep[a,c]	0,52 (0,05)	0.94 [0.83-0.98]	0,56 (0,06)	0.93 [0.85-0.97]	0.52 (0.04)	0.86 [0.74-0.94]	0.56 (0.06)	0.95 [0.91-0.98]
Tstep STD[b]	0,02 (0,01)	0.37 [0.11-0.64]	0,03 (0,02)	0.75 [0.55-0.88]	0.03 (0.02)	0.1 [0-0.42]	0.04 (0.02)	0.64 [0.4-0.82]
Tstep CV[b]	4,6 (2,09)	0.4 [0.13-0.66]	6,03 (2,57)	0.68 [0.46-0.84]	6.51 (4.3)	0.12 [0-0.44]	7 (3.46)	0.62 [0.37-0.81]
Phase	179,93 (4,46)	0 [0-0]	180,66 (5,71)	0.06 [0-0.39]	178.72 (6.29)	0.15 [0-0.46]	180.61 (7.22)	0 [0-0.31]
Phase STD	5,57 (2,79)	0.25 [0.01-0.54]	6,94 (3,76)	0.68 [0.46-0.84]	7.17 (5.79)	0.33 [0.05-0.61]	7.36 (4.38)	0.48 [0.22-0.72]
Phase CV	3,1 (1,56)	0.24 [0-0.53]	3,83 (2,07)	0.68 [0.45-0.84]	4.06 (3.41)	0.31 [0.03-0.6]	4.08 (2.43)	0.5 [0.24-0.73]
PCI	5,98 (2,96)	0.37 [0.11-0.64]	7,59 (3,79)	0.75 [0.56-0.88]	7.97 (5.24)	0.33 [0.06-0.62]	8.59 (4.38)	0.59 [0.34-0.79]
Gait NJS AP [b,d]	1,15 (0,14)	0.81 [0.65-0.91]	1,21 (0,2)	0.77 [0.59-0.89]	0.92 (0.2)	0.96 [0.92-0.98]	0.93 (0.24)	0.95 [0.91-0.98]
Gait NJS ML[b,d]	1,11 (0,26)	0.93 [0.86-0.97]	1,11 (0,26)	0.88 [0.76-0.94]	0.75 (0.33)	0.95 [0.9-0.98]	0.76 (0.31)	0.96 [0.92-0.98]
Gait NJS V	1,1 (0,24)	0.9 [0.8-0.95]	1,24 (0,38)	0.85 [0.72-0.93]	1.08 (0.25)	0.9 [0.79-0.95]	1.19 (0.31)	0.91 [0.82-0.96]
HR AP[b,d]	1,81 (0,28)	0.72 [0.51-0.86]	1,78 (0,26)	0.55 [0.29-0.77]	1.48 (0.32)	0.92 [0.85-0.97]	1.46 (0.3)	0.71 [0.5-0.86]
HR ML[b,c,d]	2,02 (0,47)	0.71 [0.5-0.86]	1,81 (0,39)	0.63 [0.38-0.82]	1.59 (0.55)	0.83 [0.67-0.92]	1.51 (0.38)	0.62 [0.38-0.81]
HR V[b,d]	2,33 (0,43)	0.69 [0.47-0.85]	2,22 (0,43)	0.82 [0.66-0.92]	1.82 (0.41)	0.79 [0.63-0.9]	1.78 (0.43)	0.85 [0.71-0.93]

ICC = Intraclass Correlation Coefficient; LB = Lower Bound of 95% confidence interval; UB = Upper Bound of 95% confidence interval; ANOVA repeated measures - within factor: Consecutive trials - between factor: Group. [a] Time in ST is significant; [b] Group in ST is significant; [c] Time in DT is significant; [d] Group in DT is significant

Table 4.2: Values of iTUG measures for PD and CTRL trials in single task and dual task conditions. Test-retest reliability is assessed by means of ICC(3,1); differences ($p<0.05$) between groups and between consecutive trials are assessed by Anova Repeated Measures.

which violates reliability model assumptions, are replaced by zeros.

As expected, the total duration of the test (Tot Dur) is very reliable (as in [85]) but is not sensitive to group differences for early stages of the disease (as in [26]). It could be argued that in the current study an acceleration-derived duration of a 7m-TUG was used instead of the traditional duration of a 3m-TUG measured by a stopwatch; however Zampieri et al [26] already showed that in a comparable population the traditional duration did not show any significant group differences.

Considering the sit-to-stand component, its duration shows no group differences (as in [30]) and low reliability; instead the root mean square (RMS) of the acceleration in the vertical direction shows a fair to good reliability and it is significantly lower in PD both in ST and DT. This reflects a greater range of motion of the trunk in the vertical direction for CTRL subjects with respect to PD. This result could be analyzed with future studies, especially to focus on the subcomponents of the sit-to-stand process (as in [86, 30]). RMS in the vertical direction, even if not chosen by the feature selection process, is a clear example of how an acceleration-based measure can detect differences between the two groups when the UPDRS clinical evaluation does not. In fact the "arising from a chair" item of the UPDRS motor examination is zero for all the PD subjects (zero equals to the absence of impairment).

Considering the gait component, most of the normalized jerk scores and harmonic ratios show good to excellent reliability and significant group differences. Interestingly Tstep (cadence) shows no group differences in contrast with [25] where PD subjects of comparable age and disease severity (Hoehn & Yahr and total of motor-UPDRS) and healthy subjects of comparable age were considered; the two main differences between the two data sets are the smaller sample size in [25] (12 PD, 12 CTRL) and the fact that PD subjects in that study had never taken anti-parkinson medications. The difference in sample sizes seems to be the most plausible reason for the contrasting results since the PD subjects of the current study were tested OFF-medication.

Among the measures of gait variability, phase-related measures (Phase STD, Phase CV and PCI) do not show any group differences. Instead the variability of Tstep in single task is significantly higher in PD; this measure on the other hand shows poor reliability (which slightly improves in dual task with respect to single task both for CTRL and PD).

Total duration and Tstep (cadence) show a significant within factor (consecutive trials) in both ST and DT conditions: they tend to decrease from the first to the third trial, showing a possible learning effect of the test. The same consideration can be applied to the duration of sit-to-stand which has a significant within factor in ST condition and a decreasing trend across consecutive trials; the same trend is identifiable in DT condition although it is not

CLASS	OVERALL OPTIMAL SUBSETS Measure (Direction, Condition)		SELECTION TIMES (out of 40)	MR % [CI]
LDA	Gait NJS (AP, ST)	HR (V, ST)	37	12.5 [5.5 26.1]
	Gait NJS (AP, DT)	HR (AP,DT)	37	12.5 [5.5 26.1]
QDA	Gait NJS (AP, ST)	Tstep CV (ST)	37	10 [4-23]
MC	Gait NJS (AP, DT)	HR (V, DT)	39	12.5 [5.5 26.1]
LR	Gait NJS (AP, DT)	HR (AP, DT)	40	12.5 [5.5 26.1]
KNN	Gait NJS (AP, ST)	HR (V, ST)	38	10 [4-23]
SVM	Gait NJS (AP, DT)	HR (AP, DT)	37	12.5 [5.5 26.1]

Table 4.3: Classifiers performance. For each classifier the overall optimal subsets are reported with the relative misclassification rate (MR) and the confidence intervals (CI). For each classifier only subsets which were selected with the highest *selection times* are reported.

significant. HR ML also has a significant within factor in DT condition with an increasing trend (i.e. increasing smoothness of the ML trunk sway while walking) across trials: this could be related to a learning effect of the subtraction task.

4.3.2 Selected Features and Performance of the Classifiers

In Table 4.3, a summary of the results of the feature selection procedure is reported for subsets of two features: for each classifier, only the subsets with the highest *selection times* (see section 3.2.1) are presented; MRs are reported together with confidence intervals. The best MR considering each feature alone is 22.5%(CI: 12.3-37.5) (not shown) for the Gait NJS(AP, ST). Considering pairs of features improves MR, as it can be seen in Table 4.3 where the minimum MR is 10%.

On the other hand considering three features together improves MRs only very slightly. In fact the best MR with three features (not shown) is 7.5% (CI: 2.6-19.9). A lower number of features (2) was preferred to this slight (and possibly not significant, given the large CIs) improvement in MR; therefore subsets of two features were preferred and reported in Table 4.3.

It is worth noting that only measures related to the gait component of the iTUG are selected. All classifiers with two features performed at similar MRs (10-12.5%), even on different subsets. For five classifiers (out of six) the best (in terms of highest *selection times*)

subsets consist of two measures (computed in the AP or V directions and during ST or DT):

- the harmonic ratio during gait;
- the normalized jerk score during gait.

In order to choose a single subset among the ones presented, which are mostly equivalent, the following considerations were done: i) classifiers which are easy to interpret and to visualize should be preferred; ii)subsets with measures only from single task should be preferred (there is no need to add a factor of variability and complexity, the dual task, if this does not bring significant improvements); iii) subsets consisting of reliable features should be preferred (i.e. the subset related to QDA should not be chosen because Tstep CV has a low reliability).

Based on these considerations the chosen subset is the one related with the LDA classifier which consists of HR(V, ST) and Gait NJS(AP, ST). LDA is easy to interpret and to visualize; in a 2 dimensional space the discrimination is visualized through a line that divides the subjects' values in two parts, as in Fig. 4.6. A further advantage of LDA is the possibility of computing the distance of a PD subject from the discrimination line which can be seen as a summary measure of how far the pathological motor pattern is from a normal motor pattern. Supporting this interpretation, this distance (positive values for PD subjects under the line, negative otherwise) was found to be positively correlated in PD with the gait/posture subscore (R=0.46, p=0.04) and the PIGD subscore(R=0.47, p=0.038): the higher the distance, the higher the severity of the clinical subscores (it has to be said that the two subscores share two common UPDRS items [8, 25]). More of the same, a positive trend was found, even if not significant (p=0.07) between the distance of PD subjects from the discriminant line and their disease duration. To give a preliminary evidence of the robustness and generalizability of the classifier to new data, the same LDA line computed on the mean values of the two measures was used (without re-training of the classifier) to discriminate all the three consecutive trials of each subject (Fig. 4.7, left panel). Even if this data is not properly "new" because it comes from the same subjects, it was never used by the classifier to train; the left panel of Fig. 4.7 shows that the discrimination is still accurate considering the three consecutive trials for each subject instead of their means. To gain insight into the two proposed measures, the trials corresponding to the minimum and maximum values of both (they are shown by points a),b),c),d) in the left panel of Fig. 4.7 were considered. It can be noted that the trials associated to the minimum values of Gait NJS (AP, ST) and HR (V, ST) both correspond to PD subjects; on the contrary the trials associated to their maximum values both correspond to healthy subjects. For HR (V, ST) the acceleration signals in the vertical direction corresponding to its minimum and maximum

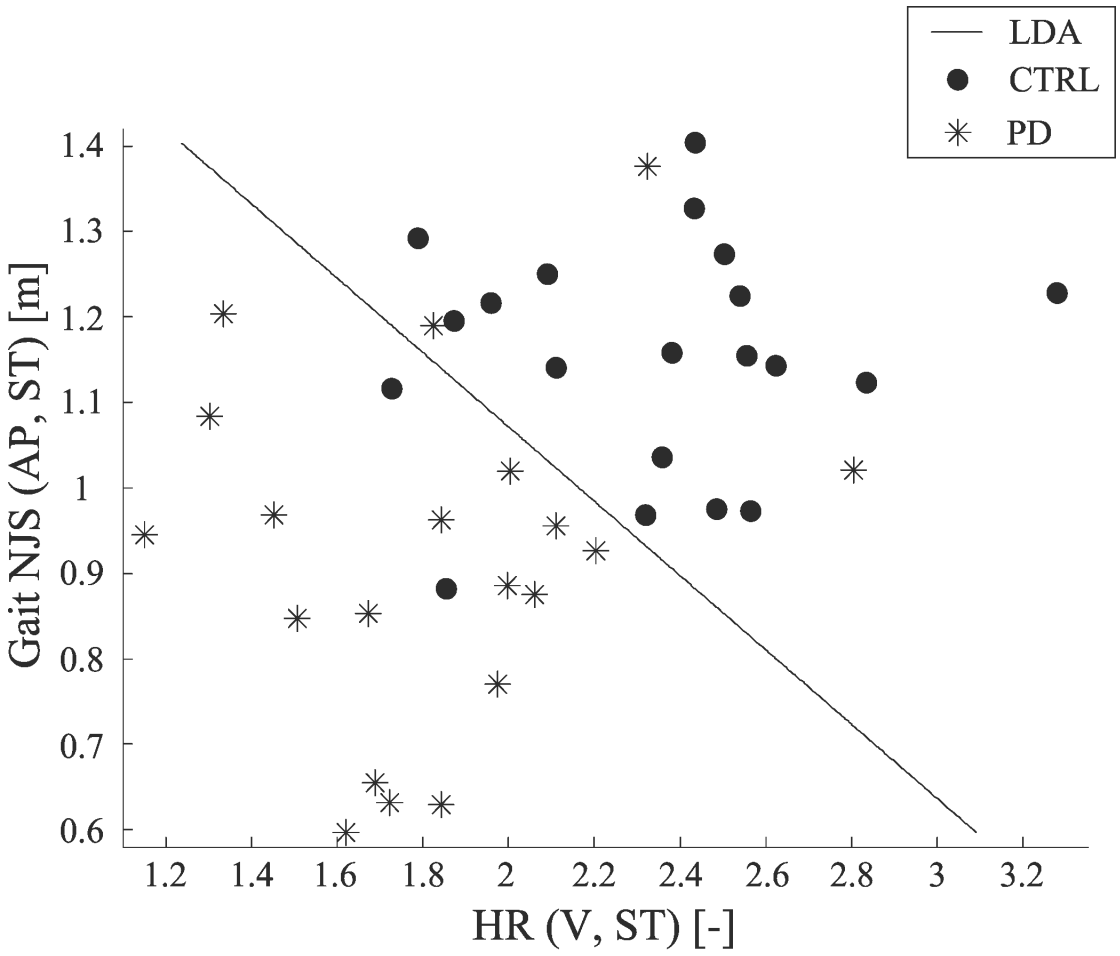

Figure 4.6: Plot of the values of the Gait NJS (AP, ST) and the HR (V, ST) measures considering the mean of the three trials for each subject. The discrimination line resulting from the LDA classifier is also reported.

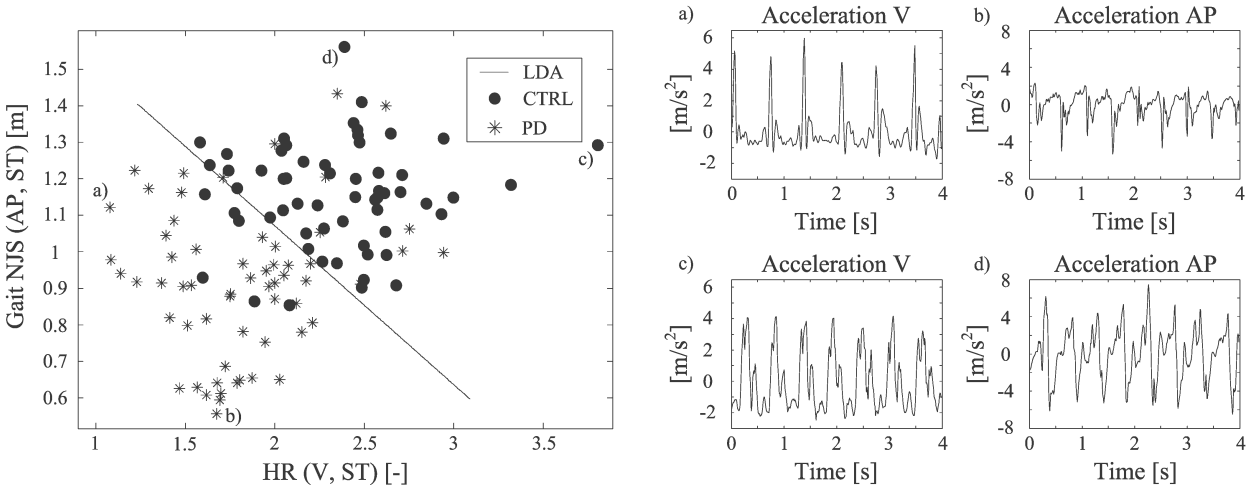

Figure 4.7: Left panel: Plot of the values of the Gait NJS (AP, ST) and the HR (V, ST) measures considering all the three trials for each subject. The discrimination line is the same as the one in Fig. 4.6. Right Panel: Acceleration signals related to the minimum and the maximum values of the measures plotted in the left panel: 4 seconds of gait are shown. The offset has been removed from each signal. a,c) Vertical acceleration associated to the minimum (a) and maximum (c) value of HR (V, ST); ,db) Antero-posterior acceleration associated to the minimum (b) and maximum (d) value of Gait NJS (AP, ST).

value are reported in Fig. 4.7(a) and Fig. 4.7(c), respectively. Similarly, for NJS (AP, ST) the acceleration signals in the AP direction corresponding to its minimum and maximum value are reported in 4.7(b) and 4.7(d), respectively.

Harmonic ratios are a measure of smoothness of walking [87] and have already been found to be lower in PD subjects [29, 28]; they also have been found to be lower in elderly fallers with respect to non-fallers and in individuals with peripheral neuropathy with respect to healthy individuals. HR values are lower in PD (Table 4.2) subjects in accordance with the results of [29, 28]; this is further supported by the fact that in the current study PD subjects with increasing gait and posture impairments tend to have decreasing values of HR(V, ST), as reflected by the negative correlation (R=-0.56, p=0.01) with the gait/posture clinical subscore. HR(V, ST) identifies a lower smoothness in the vertical acceleration signal of the PD subject: while in healthy subjects the foot contact with the floor is characterized by a smooth heel-to-toe pattern, this may be altered in PD, thus leading to a prominent flat-footed gait and a reduced roll-off [88], Baltadjieva2006].

Normalized jerk score during gait is a relatively new measure: it measures the time-normalized dynamic of the acceleration during a step: fast and large variations in the signal lead to high values of this measure. As reported in Table 4.2, Gait NJS both in the AP and ML directions are very reliable, especially for the PD group in both ST and DT; PD subjects show lower NJS in both AP and ML; this reduced dynamic (low and small variations in the acceleration signals) and high repeatability could be related to a general loss of complexity of the motor control system which could lead to a loss of adaptivity in motor strategies. This hyphotesis is in accordance with the findings of [Mitoma 2000]. In this study the ground reaction forces (GRF) during walking have been found altered in PD subjects: the characteristic GRF peaks at the beginning and end of the stance phase became nearly vertical; this have been linked to a breakdown in gait control or ataxia. Finally, NJS (AP, ST) was not found to be associated with the clinical information available and thus seems to bring complementary information.

4.3.3 Dual Task

In this study dual task was considered as a perturbing effect that could increase the differences in the motor performance between CTRL and PD subjects [89]. From the analysis that was performed it seems that dual task does not have the expected added value (that it had in quiet standing): in fact group differences in DT are also present in ST; instead some of the group differences in ST are not present in DT. Moreover the highest accuracy (lowest MR) in discriminating between the two groups was found with two features both related to

ST. It has to be said that investigating the effects of dual task on PD motor function was not the aim of this study; in the future further analysis should focus on variations in measures between ST and DT could provide more insight on the effect of a concurrent cognitive task on PD locomotor patterns.

4.3.4 Impact on Parkinson's Disease

This study quantitatively evaluates the locomotor function of early-mild PD subjects by considering a minimum set up (a single accelerometer) to measure the performance of a simple clinical test (Timed Up and Go). Reliability and discriminative power of acceleration-derived measures computed from the gait and sit-to-stand components of the test are evaluated. A subset of two measures is then chosen by using a feature selection process: these two measures are highly reliable and their combination can well discriminate, with the aid of a simple and easily interpretable classification algorithm, between the locomotor function of CTRL and PD subjects. This, in an early-mild stage of the disease, when the traditional outcome of the test would not be able to discriminate between healthy and pathological performances. The final outcome of the classification algorithm is found to be sensitive to clinical scores which assess symptoms related to gait and posture. These results suggest that the proposed iTUG plausibly characterizes PD locomotor impairment and, hence, may be used for evaluation, follow-up, and possibly remote monitoring. The quantitative assessment of the locomotor function of the current study is a step forward in the direction of obtaining a multifactorial and complete assessment of the motor impairments of individuals with Parkinson's disease.

4.3.5 Limitations and Future Developments

The relatively small sample size is a limitation, reflected by the large confidence intervals of intra-class correlation coefficients in Table 4.2 and of misclassification rates in Table 4.3. Also correlation results could be biased by the small sample size and should be verified in a larger sample. This problem could be overcome by future studies that will implement this set-up and protocol in other PD subjects.

Another limitation is the exclusion of the turning from the analysis: it was not considered because a low consistency of the acceleration signals in that component (among subjects and/or repetitions) was found. It is possible that measures computed from the turning phase could increase the discrimination ability of the classification algorithms. In fact duration of turning was found to be significantly lower in early-mild PD subjects (with a comparable disease severity) with respect to healthy subjects in [25]. To overcome this limitation in future

studies it could be possible to consider a gyroscope sensor together with the accelerometer; this could help to segment the different components of the iTUG accurately.

In this study acceleration-based measures were found to be associated with clinical measures of disease severity. Future studies should consider different stages of the disease and longitudinal monitoring to assess the sensitivity of acceleration-based measures to the severity and progression of the disease.

The proposed method (single accelerometer-based iTUG and feature selection) could be used to quantify the locomotor function of populations characterized by different motor impairments.

Instrumented TUG, given its simplicity and ease of reproducibility, could be used outside the clinical setting; in a pilot study by [90] iTUG is used for remote home monitoring. For this kind of applications a smartphone-based version of the presented iTUG could be used, which uses the smartphone's accelerometer as the measurement device [91]. This would improve the pervasiveness, ease of use, and cost of the protocol.

4.4 Moderate

4.4.1 Clustering Results

The best unsupervised clustering result (see Fig. 4.8) was found with two features in a three-clusters structure: the two features were both the semi-tandem eyes-open condition:

- the percentage of Power in High Frequency components in AP direction (which is related to tremor in AP)
- the Sway Path in the ML direction.

This clustering structure has an average silhouette value of 0.7, which can be interpreted as a reasonable-to-strong structure captured in the data (see section 3.4).

The clustering structure is made of 3 clusters:

A) cluster with high AP tremor;

B) cluster with low AP tremor and low ML sway;

C) cluster with low AP tremor, high ML sway.

Two of the three obtained clusters are very similar to the clinical subtypes:

- cluster A) $\approx$ TD;

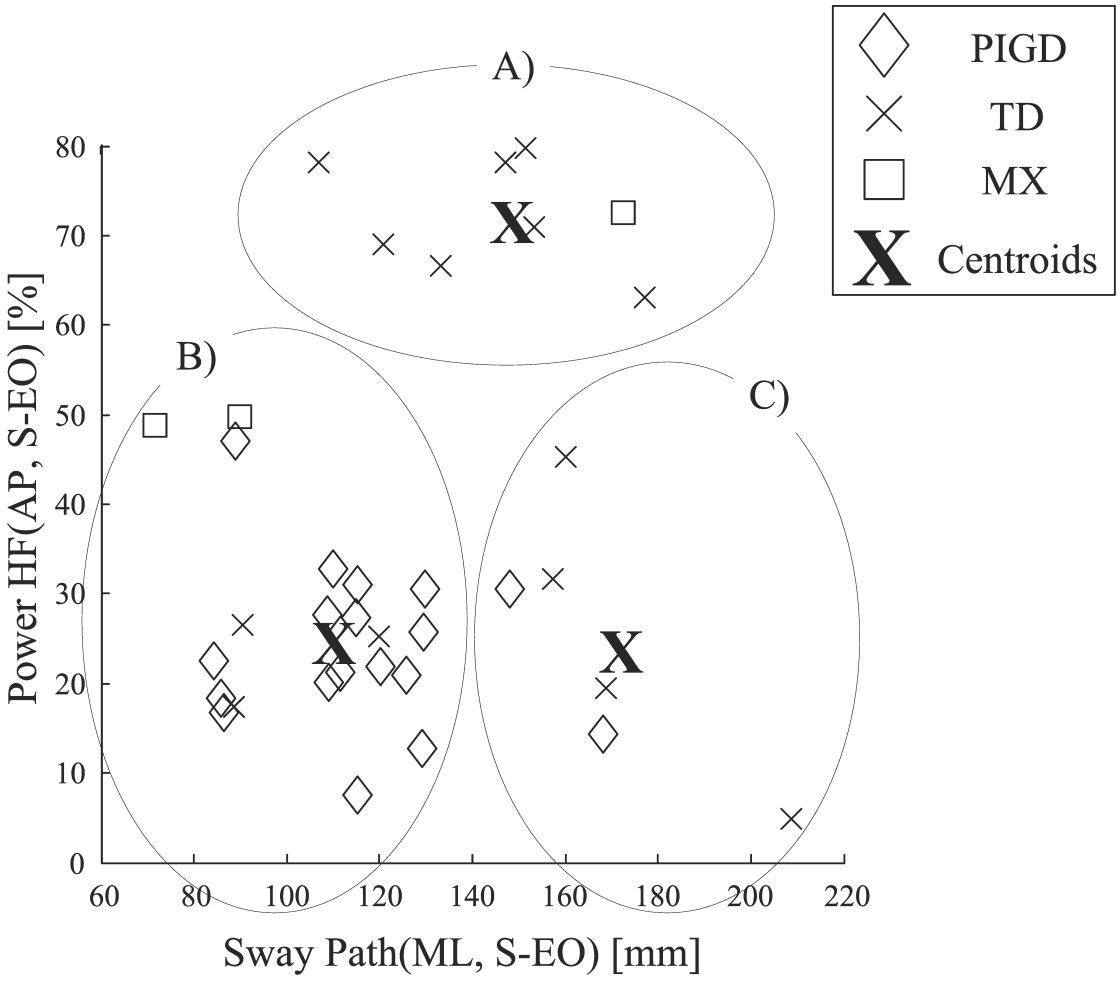

Figure 4.8: Optimal clustering result.

- cluster B) ≈ PIGD;
- cluster C) has no direct relation with clinical classification.

Clinically-classified TD subjects were described only by high tremor related quantitative measure. This variable seems to be specific but not enough sensitive for the TD subjects (specificity:100%, sensitivity 50%), since some are classified in other clusters. This could be considered as an added value to the clinical classification: even if these subjects are considered to be TD, their tremor is not detected during postural tests. The PIGD group is well described by cluster low ML sway and low AP tremor (specificity: 79%, sensitivity: 89%). A possible explanation is that rigidity may lead to low ML sway.

4.4.2 Impact on Parkinson's Disease

The present study suggests that accelerometer-derived measures may augment the clinical evaluation and help to classify, monitor and quantify different PD subtypes and their evolution over time, such as the response to therapeutic interventions.

4.4.3 Limitations and Future Developments

The main limitation are given by the small sample size, the fact that control subjects were not available, and the fact that only the postural function was considered: a more complete protocol, ad example with the data from iTUG, may lead to better results. Other limitations are the fact that only pairs of measures were considered (considering more could lead to better results) and the fact that the standardized foot stance was not used in the protocol. It is also possiblethat considering other feature selection or clustering techniques may lead to better results.

4.5 Advanced PD

4.5.1 Results in Estimating Clinical Scores

Only the preliminary results of ongoing work are presented in this thesis. The results are related to the estimation of the UPDRS clinical scores of the first subject. The error in estimation RMSE (see section 3.5) was averaged between the right and left limbs. Tremor and Dyskinesia were only estimated for the legs because neither tremor nor dyskinesia were present during the testing on the arms of the considered subject. The results are reported in Table 4.4.

	Tremor	Dyskinesia	Bradykinesia (Alternating Hand Movements)	Bradykinesia (Leg Agility)
RMSE	0,4	0,3	0,3	0,7
STD	0,12	0,2	0,1	0,4

Table 4.4: Regression Results: mean and standard deviation (STD) of the RMSE across the ten splits of the database in training and testing.

4.5.2 Discussion

From Table 4.4 it can be seen that a reasonable estimation of the clinical scores related to tremor, dyskinesia, and bradykinesia in alternating hand movements was obtained, especially considering the fact that UPDRS scores assigned by a clinical expert have an inherent inter and intra-examiner variability [92]. On the other hand the estimation of the bradykinesia in the leg agility task does not seem as accurate as the other ones.

These results provide preliminary evidence of the potential of using inertial sensors to facilitate the process of selecting the optimal combination of DBS configuration and medication dosages. If these results were confirmed, the inertial sensors could be used to monitor over long periods of time, and without expert clinical symptoms resulting from changes in DBS configurations and/or medication dosages.

Chapter 5

Conclusion

The motor impairment of Parkinson's disease was evaluated with quantitative methods in different stages of the disease: acquisition of data from wearable sensors was combined with data mining techniques to extract relevant information for clinical purposes. For each considered stage of the disease, a different aim was considered and appropriate techniques were applied. To deal with specific characteristics of the different databases, *ad hoc* techniques of feature selection were developed.

The preliminary result obtained in the identification of healthy subject at a high risk of developing PD suggests for the first time that instrumented posture analysis may be able to disclose preclinical signs of a high risk of developing PD.

Regarding the early stage of the disease, when symptoms are still not severe, quantitative measures computed from accelerometers seem to be able to identify subtle postural, tremor and locomotor characteristics of PD subjects which may not be evident from a clinical evaluation.

For the moderate stage of the disease, clusters of subjects with homogeneous postural patterns were found which are also related with specific clinical characteristics.

For the advanced stage of the disease, the preliminary results provide evidence of the potential of using wearable sensors and data mining techniques to estimate longitudinally the severity of symptoms in the home setting.

The results obtained in this thesis have implications both in clinical practice and in clinical research. Regarding the former, the combination of the data mining tools developed and the low-cost protocols for data acquisition allows the evaluation and monitoring (in clinical and home environments) of the motor function of PD subjects. Regarding the latter, motor patterns of PD were studied and characterized in different stages of the disease. This increases the quantitative knowledge about motor impairment in PD and about the evolution

of the disease. The information extracted from the acceleration data in the database of early subjects is the most complete database considered in this thesis: the follow-up of the same subjects is ongoing and the results will be used in the future by clinical experts for correlation studies with functional magnetic resonance images; this, in order to find a relation between a specific brain damage and motor impairment.

The main limitation of the presented studies is the relatively small sample size of the datasets, reflected by the large confidence intervals in accuracies. This problem could be overcome in the future by widening the presented databases of PD subjects and by considering follow-up studies. This would make the present findings statistically more robust. Future work could also deal with different feature selection procedures to explore in different ways the space of the combinations of the features since it is possible that this would result in improved accuracy of the classification.

Finally, the proposed approach of this thesis, even if it is focused on a specific clinical application, can be easily extended to other applications with the aim to extract quantitative information from datasets derived from wearable sensors (or other movement analysis devices).

List of Figures

List of Tables

Bibliography

[1] W. J. Frawley, G. Piatetsky-shapiro, and C. J. Matheus, "Knowledge Discovery in Databases: An Overview," *AI Magazine*, vol. 13, no. 3, pp. 57–70, 1992.

[2] M. E. Morris, "Movement disorders in people with Parkinson disease: a model for physical therapy.," *Physical therapy*, vol. 80, pp. 578–97, June 2000.

[3] C. D. Marsden, "Parkinson's disease.," *Journal of neurology, neurosurgery, and psychiatry*, vol. 57, pp. 672–81, June 1994.

[4] S. Fahn and R. Elton, "The UPDRS development commitee. Unified Parkinson's disease rating scale.," in *Recent developments in Parkinson's disease.* (S. Fahn, C. Marsdern, D. Calne, and M. Goldstein, eds.), pp. 153–163, Florham Park, New Jersey: Mac Millan Healthcare Information, 1987.

[5] M. M. Hoehn and M. D. Yahr, "Parkinsonism: onset, progression and mortality.," *Neurology*, vol. 17, pp. 427–42, May 1967.

[6] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, M. D. Yahr, and L. Seidl, "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations.," *Movement Disorders*, vol. 19, pp. 1020–8, Sept. 2004.

[7] S. M. Van Rooden, W. J. Heiser, J. N. Kok, D. Verbaan, J. J. van Hilten, and J. Marinus, "The identification of Parkinson's disease subtypes using cluster analysis: a systematic review.," *Movement Disorders*, vol. 25, pp. 969–78, June 2010.

[8] J. Jankovic, M. McDermott, J. Carter, S. Gauthier, C. Goetz, L. Golbe, S. Huber, W. Koller, C. Olanow, and I. Shoulson, "Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group.," *Neurology*, vol. 40, pp. 1529–34, Oct. 1990.

[9] A. Nardone and M. Schieppati, "Balance in Parkinson's disease under static and dynamic conditions.," *Movement Disorders*, vol. 21, pp. 1515–20, Sept. 2006.

[10] J. W. Błaszczyk, R. Orawiec, D. Duda-Kłodowska, and G. Opala, "Assessment of postural instability in patients with Parkinson's disease.," *Experimental brain research*, vol. 183, pp. 107–14, Oct. 2007.

[11] J. W. Błaszczyk and R. Orawiec, "Assessment of postural control in patients with Parkinson's disease: sway ratio analysis.," *Human movement science*, vol. 30, pp. 396–404, Apr. 2011.

[12] L. Rocchi, L. Chiari, and F. B. Horak, "Effects of deep brain stimulation and levodopa on postural sway in Parkinson's disease," *Gait & Posture*, pp. 267–274, 2002.

[13] L. Rocchi, L. Chiari, A. Cappello, and F. B. Horak, "Identification of distinct characteristics of postural sway in Parkinson's disease: a feature selection procedure based on principal component analysis.," *Neuroscience letters*, vol. 394, pp. 140–5, Mar. 2006.

[14] R. Marchese, M. Bove, and G. Abbruzzese, "Effect of cognitive and motor tasks on postural stability in Parkinson's disease: a posturographic study.," *Movement Disorders*, vol. 18, pp. 652–8, June 2003.

[15] M. A. McVey, A. P. Stylianou, C. W. Luchies, K. E. Lyons, R. Pahwa, S. Jernigan, and J. D. Mahnken, "Early biomechanical markers of postural instability in Parkinson's disease.," *Gait & posture*, vol. 30, pp. 538–42, Nov. 2009.

[16] F. B. Horak, D. Dimitrova, and J. G. Nutt, "Direction-specific postural instability in subjects with Parkinson's disease.," *Experimental neurology*, vol. 193, pp. 504–21, June 2005.

[17] A. M. S. Muniz, H. Liu, K. E. Lyons, R. Pahwa, W. Liu, and J. Nadal, "Quantitative evaluation of the effects of subthalamic stimulation on gait in Parkinson's disease patients using principal component analysis.," *The International journal of neuroscience*, vol. 120, pp. 609–16, Sept. 2010.

[18] H. Mitoma, R. Hayashi, N. Yanagisawa, and H. Tsukagoshi, "Characteristics of parkinsonian and ataxic gaits: a study using surface electromyograms, angular displacements and floor reaction forces.," *Journal of the neurological sciences*, vol. 174, pp. 22–39, Mar. 2000.

[19] M. E. Morris, J. McGinley, F. Huxham, J. Collier, and R. Iansek, "Constraints on the kinetic, kinematic and spatiotemporal parameters of gait in Parkinson's disease," *Human Movement Science*, vol. 18, pp. 461–483, June 1999.

[20] K. A. Chung, B. M. Lobb, J. G. Nutt, J. McNames, and F. Horak, "Objective measurement of dyskinesia in Parkinson's disease using a force plate.," *Movement Disorders*, vol. 25, pp. 602–8, Apr. 2010.

[21] L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, and L. Chiari, "Feature Selection for Accelerometer-Based Posture Analysis in Parkinsons disease," *IEEE transactions on information technology in biomedicine*, vol. 15, no. 99, pp. 1–1, 2011.

[22] M. Mancini, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and L. Chiari, "Trunk accelerometry reveals postural instability in untreated Parkinson's disease.," *Parkinsonism & related disorders*, vol. 17, pp. 557–62, Aug. 2011.

[23] M. Mancini, C. Zampieri, P. Carlson-Kuhta, L. Chiari, and F. B. Horak, "Anticipatory postural adjustments prior to step initiation are hypometric in untreated Parkinson's disease: an accelerometer-based approach," *European Journal of Neurology*, vol. 16, pp. 1028–1034, Sept. 2009.

[24] A. Salarian, H. Russmann, F. J. G. Vingerhoets, C. Dehollain, Y. Blanc, P. R. Burkhard, and K. Aminian, "Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring.," *IEEE transactions on bio-medical engineering*, vol. 51, pp. 1434–43, Aug. 2004.

[25] A. Salarian, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian, "iTUG, a sensitive and reliable measure of mobility.," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 18, pp. 303–10, June 2010.

[26] C. Zampieri, A. Salarian, P. Carlson-Kuhta, K. Aminian, J. G. Nutt, and F. B. Horak, "The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson's disease.," *Journal of neurology, neurosurgery, and psychiatry*, vol. 81, pp. 171–6, Feb. 2010.

[27] B. Caby, S. Kieffer, M. de Saint Hubert, G. Cremer, and B. Macq, "Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry.," *Biomedical engineering online*, vol. 10, p. 1, Jan. 2011.

[28] 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.," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 64, pp. 700–6, June 2009.

[29] K. A. Lowry, A. L. Smiley-Oyen, A. J. Carrel, and J. P. Kerr, "Walking stability using harmonic ratios in Parkinson's disease.," *Movement Disorders*, vol. 24, pp. 261–7, Jan. 2009.

[30] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, I. Maidan, N. Giladi, T. Gurevich, and J. M. Hausdorff, "Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson's disease?," *Medical engineering & physics*, vol. 32, pp. 119–25, Mar. 2010.

[31] D. Zwartjes, T. Heida, J. van Vugt, J. Geelen, and P. Veltink, "Ambulatory Monitoring of Activities and Motor Symptoms in Parkinson¿s Disease.," *IEEE transactions on bio-medical engineering*, May 2010.

[32] J. D. Frost, "Triaxial vector accelerometry: a method for quantifying tremor and ataxia.," *IEEE transactions on bio-medical engineering*, vol. 25, pp. 17–27, Jan. 1978.

[33] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors.," *IEEE transactions on information technology in biomedicine*, vol. 13, pp. 864–73, Nov. 2009.

[34] J. I. Hoff, E. A. Wagemans, and B. J. van Hilten, "Ambulatory objective assessment of tremor in Parkinson's disease.," *Clinical neuropharmacology*, vol. 24, no. 5, pp. 280–3, 2001.

[35] A. Salarian, H. Russmann, C. Wider, P. R. Burkhard, F. J. G. Vingerhoets, and K. Aminian, "Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system.," *IEEE transactions on bio-medical engineering*, vol. 54, pp. 313–22, Feb. 2007.

[36] E. J. W. Van Someren, M. D. Pticek, J. D. Speelman, P. R. Schuurman, R. Esselink, and D. F. Swaab, "New actigraph for long-term tremor recording.," *Movement Disorders*, vol. 21, pp. 1136–43, Aug. 2006.

[37] J. Jankovic and J. D. Frost, "Quantitative assessment of parkinsonian and essential tremor: Clinical application of triaxial accelerometry," *Neurology*, vol. 31, pp. 1235–, Oct. 1981.

[38] R. J. Dunnewold, C. E. Jacobi, and J. J. van Hilten, "Quantitative assessment of bradykinesia in patients with Parkinson's disease.," *Journal of neuroscience methods*, vol. 74, pp. 107–12, June 1997.

[39] S. Katayama, "Actigraph analysis of diurnal motor fluctuations during dopamine agonist therapy.," *European neurology*, vol. 46 Suppl 1, pp. 11–7, Jan. 2001.

[40] P. R. Burkhard, H. Shale, J. W. Langston, and J. W. Tetrud, "Quantification of dyskinesia in Parkinson's disease: validation of a novel instrumental method.," *Movement Disorders*, vol. 14, pp. 754–63, Sept. 1999.

[41] N. L. Keijsers, M. W. Horstink, and S. C. Gielen, "Online monitoring of dyskinesia in patients with Parkinson's disease.," *IEEE engineering in medicine and biology magazine*, vol. 22, no. 3, pp. 96–103, 2003.

[42] N. L. W. Keijsers, M. W. I. M. Horstink, and S. C. A. M. Gielen, "Ambulatory motor assessment in Parkinson's disease.," *Movement Disorders*, vol. 21, pp. 34–44, Jan. 2006.

[43] E. L. Johnsen, P. H. Mogensen, N. A. Sunde, and K. Ø stergaard, "Improved asymmetry of gait in Parkinson's disease with DBS: gait and postural instability in Parkinson's disease treated with bilateral deep brain stimulation in the subthalamic nucleus.," *Movement Disorders*, vol. 24, pp. 590–7, Mar. 2009.

[44] M. Plotnik, N. Giladi, and J. M. Hausdorff, "A new measure for quantifying the bilateral coordination of human gait: effects of aging and Parkinson's disease.," *Experimental brain research*, vol. 181, pp. 561–70, Aug. 2007.

[45] M. Plotnik, Y. Dagan, T. Gurevich, N. Giladi, and J. M. Hausdorff, "Effects of cognitive function on gait and dual tasking abilities in patients with Parkinson's disease suffering from motor response fluctuations.," *Experimental brain research*, vol. 208, pp. 169–79, Jan. 2011.

[46] B. R. Brewer, S. Pradhan, G. Carvell, and A. Delitto, "Application of modified regression techniques to a quantitative assessment for the motor signs of Parkinson's disease.," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 17, pp. 568–75, Dec. 2009.

[47] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests.," *IEEE transactions on bio-medical engineering*, vol. 57, pp. 884–93, Apr. 2010.

[48] W. Maetzler, I. Liepelt, and D. Berg, "Progression of Parkinson's disease in the clinical phase: potential markers.," *Lancet neurology*, vol. 8, pp. 1158–71, Dec. 2009.

[49] A. Gaenslen, B. Unmuth, J. Godau, I. Liepelt, A. Di Santo, K. J. Schweitzer, T. Gasser, H.-J. Machulla, M. Reimold, K. Marek, and D. Berg, "The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of Parkinson's disease: a prospective blinded study.," *Lancet neurology*, vol. 7, pp. 417–24, May 2008.

[50] D. Berg, K. Seppi, S. Behnke, I. Liepelt, K. Schweitzer, H. Stockner, F. Wollenweber, A. Gaenslen, P. Mahlknecht, J. Spiegel, J. Godau, H. Huber, K. Srulijes, S. Kiechl, M. Bentele, A. Gasperi, T. Schubert, T. Hiry, M. Probst, V. Schneider, J. Klenk, M. Sawires, J. Willeit, W. Maetzler, K. Fassbender, T. Gasser, and W. Poewe, "Enlarged substantia nigra hyperechogenicity and risk for Parkinson disease: a 37-month 3-center study of 1847 older persons.," *Archives of neurology*, vol. 68, pp. 932–7, July 2011.

[51] W. E. McIlroy and B. E. Maki, "Preferred placement of the feet during quiet stance: development of a standardized foot placement for balance testing.," *Clinical biomechanics (Bristol, Avon)*, vol. 12, pp. 66–70, Jan. 1997.

[52] D. Podsiadlo and S. Richardson, "The timed "Up & Go": a test of basic functional mobility for frail elderly persons.," *Journal of the American Geriatrics Society*, vol. 39, pp. 142–8, Feb. 1991.

[53] Y. Balash, C. Peretz, G. Leibovich, T. Herman, J. M. Hausdorff, and N. Giladi, "Falls in outpatients with Parkinson's disease: frequency, impact and identifying factors.," *Journal of neurology*, vol. 252, pp. 1310–5, Nov. 2005.

[54] T. Herman, N. Giladi, and J. M. Hausdorff, "Properties of the 'timed up and go' test: more than meets the eye.," *Gerontology*, vol. 57, pp. 203–10, Jan. 2011.

[55] M. Rudzińska, M. Marona, S. Bukowczan, K. Banaszkiewicz, E. Mirek, and A. Szczudlik, "Falls in different types of Parkinson's disease.," *Neurologia i neurochirurgia polska*, vol. 41, no. 5, pp. 395–403, 2007.

[56] Y. Katayama, M. Kasai, H. Oshima, C. Fukaya, T. Yamamoto, K. Ogawa, and T. Mizutani, "Subthalamic nucleus stimulation for Parkinson disease: benefits observed in levodopa-intolerant patients.," *Journal of neurosurgery*, vol. 95, pp. 213–21, Aug. 2001.

[57] Deep-Brain Stimulation for Parkinson's Disease Study Group, "Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease.," *The New England journal of medicine*, vol. 345, pp. 956–63, Sept. 2001.

[58] J. Volkmann, J. Herzog, F. Kopper, and G. Deuschl, "Introduction to the programming of deep brain stimulators.," *Movement Disorders*, vol. 17 Suppl 3, pp. S181–7, Jan. 2002.

[59] S. Mellone, L. Palmerini, A. Cappello, and L. Chiari, "Hilbert-Huang-based tremor removal to assess postural properties from accelerometers.," *IEEE transactions on biomedical engineering*, vol. 58, pp. 1752–61, June 2011.

[60] K. E. Lyons and R. Pahwa, *Handbook of Essential Tremor and Other Tremor Disorders.* 2005.

[61] D. A. Winter, A. E. Patla, and J. S. Frank, "Assessment of balance control in humans.," *Medical progress through technology*, vol. 16, pp. 31–51, May 1990.

[62] T. E. Prieto, J. B. Myklebust, R. G. Hoffmann, E. G. Lovett, and B. M. Myklebust, "Measures of postural steadiness: differences between healthy young and elderly adults.," *IEEE transactions on bio-medical engineering*, vol. 43, pp. 956–66, Sept. 1996.

[63] L. Chiari, L. Rocchi, and A. Cappello, "Stabilometric parameters are affected by anthropometry and foot placement.," *Clinical biomechanics (Bristol, Avon)*, vol. 17, no. 9-10, pp. 666–77, 2002.

[64] W. Zijlstra, "Assessment of spatio-temporal parameters during unconstrained walking.," *European journal of applied physiology*, vol. 92, pp. 39–44, June 2004.

[65] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick, "Acceleration patterns of the head and pelvis when walking on level and irregular surfaces.," *Gait & posture*, vol. 18, pp. 35–46, Aug. 2003.

[66] S. Patel, C. Mancinelli, R. Hughes, A. Dalton, L. Shih, and P. Bonato, "Optimizing deep brain stimulation settings using wearable sensing technology," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pp. 6–9, IEEE, Apr. 2009.

[67] J. I. Hoff, A. A. van den Plas, E. A. Wagemans, and J. J. van Hilten, "Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease.," *Movement Disorders*, vol. 16, pp. 58–61, Jan. 2001.

[68] Y. Saeys, I. n. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics.," *Bioinformatics (Oxford, England)*, vol. 23, pp. 2507–17, Oct. 2007.

[69] R. Kohavi and H. John, "Artificial Intelligence Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273–324, 1997.

[70] R. Simon, M. D. Radmacher, K. Dobbin, and L. M. McShane, "Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification.," *Journal of the National Cancer Institute*, vol. 95, pp. 14–8, Jan. 2003.

[71] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, second ed., 2005.

[72] M. A. Hall, *Correlation-based Feature Selection for Machine Learning*. PhD thesis, University of Waikato, New Zealand, 1999.

[73] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, pp. 23–69, Oct. 2003.

[74] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[75] G. A. F. Seber, *Multivariate Observations*, vol. 28. Hoboken, NJ: John Wiley&Sons, 1984.

[76] D. W. Hosmer, S. Lemeshow, and E. D. Cook, *Applied Logistic Regression*. New York, New York, USA: Wiley, second edi ed., 2000.

[77] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.

[78] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[79] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, third ed., 2011.

[80] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.

[81] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, 1990.

[82] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, Oct. 2001.

[83] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability.," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.

[84] B. E. Maki, P. J. Holliday, and A. K. Topper, "A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population.," *Journal of gerontology*, vol. 49, pp. M72–84, Mar. 1994.

[85] S. Morris, M. E. Morris, and R. Iansek, "Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease.," *Physical therapy*, vol. 81, pp. 810–8, Feb. 2001.

[86] E. Nikfekr, K. Kerr, S. Attfield, and D. E. Playford, "Trunk Movement in Parkinson ' s Disease During Rising from Seated Position," *Movement Disorders*, vol. 17, no. 2, pp. 274–282, 2002.

[87] J. S. Brach, D. McGurl, D. Wert, J. M. Vanswearingen, S. Perera, R. Cham, and S. Studenski, "Validation of a measure of smoothness of walking.," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 66, pp. 136–41, Jan. 2011.

[88] J. R. Hughes, S. G. Bowes, A. L. Leeman, C. J. O'Neill, A. A. Deshmukh, P. W. Nicholson, S. M. Dobbs, and R. J. Dobbs, "Parkinsonian abnormality of foot strike: a phenomenon of ageing and/or one responsive to levodopa therapy?," *British journal of clinical pharmacology*, vol. 29, pp. 179–86, Feb. 1990.

[89] C. M. Campbell, J. L. Rowse, M. A. Ciol, and A. Shumway-Cook, "The effect of cognitive demand on Timed Up and Go performance in older adults with and without Parkinson disease," *Journal of Neurologic Physical Therapy*, vol. 27, no. 1, pp. 2–7, 2003.

[90] C. Zampieri, A. Salarian, P. Carlson-Kuhta, J. G. Nutt, and F. B. Horak, "Assessing mobility at home in people with early Parkinson's disease using an instrumented Timed Up and Go test.," *Parkinsonism & related disorders*, vol. 17, pp. 277–80, May 2011.

[91] S. Mellone, C. Tacconi, and L. Chiari, "Validity of a Smartphone-based instrumented Timed Up and Go.," *Gait & posture*, Mar. 2012.

[92] B. Post, M. P. Merkus, R. M. A. de Bie, R. J. de Haan, and J. D. Speelman, "Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?," *Movement Disorders*, vol. 20, pp. 1577–84, Dec. 2005.

Ringraziamenti

Inizio ringraziando i miei genitori ai quali è dedicata questa tesi; è grazie al loro amore e sostegno incondizionato se sono arrivato fino a qui. Voglio ringraziare subito anche la mia stupenda ragazza, Fiorenza. La vita con lei è fantastica; senza di lei non so come avrei fatto durante l'ultimo periodo della tesi; ringrazio Chi c'è lassù per averci fatto incontrare. Passando ora alle persone con cui ho lavorato durante questo dottorato, voglio ringraziare molto il prof. Lorenzo Chiari per aver creduto nel mio potenziale, per avermi insegnato tanto di tutto e per essere sempre stato disponibile; mi ritengo molto fortunato per tutto questo. Tra i miei compagni di dottorato inizio ringraziando Sabato, compagno di avventure, di bevute e di film ai limiti (e oltre) del trash: il merito di questa tesi lo condivido con lui; senza il suo aiuto ed il suo lavoro non avrei fatto metà delle cose che invece sono riuscito a fare. Poi ringrazio in ordine sparso Gonz, Giacomo, Fabio, Roberto, Elena, Valeria, Carlo, Luigi, Toni: con voi mi son divertito molto, son sempre stato bene, ho imparato ed ho condiviso gioie e dolori (burocratici) del dottorato. Ringrazio anche il prof. Paolo Bonato e Shyamal per avermi seguito durante il mio periodo all'estero; a proposito ringrazio tutta la bella gente che ho conosciuto durante la mia esperienza indimenticabile a Boston: Jaime, Sylvain, Luca, Nicolas, Martina, Sultan, Cahoime (si scrive così?), Giacomo, Chiara, Iahn, Yalgin, Amelie, Virginie...e sto sicuramente dimenticando qualcuno. Poi, last but not least, un grazie ai miei amici storici, che mi son sempre stati vicini, che mi fanno spaccare dal ridere, che mi aiutano e che ci sono sempre: quelli universitari (Cam, Manuel, Poz (si sposaaaaa), Giah, Papu, Elena) e quelli delle elementari/superiori/basket/vr (Wally, Alejandro, Manu, Alby, Claudio, Peruz, Miguel, Deghe, Tine, Vice, Barzi, Batti, Gioio, Enry, Ciaffo, Raji).

Luca