

POLITECNICO DI TORINO

DIPARTIMENTO DI INGEGNERIA MECCANICA E
AEROSPAZIALE (DIMEAS)



MASTER'S DEGREE IN BIOMEDICAL ENGINEERING

**Analysis of movement patterns in
individuals with Parkinson's disease
experiencing motor fluctuations**

Supervisor

Prof. Danilo DEMARCHI

Candidate

Matteo SERAFINO

Torino, December 2019

Academic Year 2018-2019

Abstract

Parkinson's disease (PD) is a gradual neurodegenerative disorder defined by the loss of dopaminergic neurons and the presence of Lewy bodies in the basal ganglia area. The main manifestations of PD are motor symptoms, such as bradykinesia (i.e., slowness of movement), tremor, rigidity and balance instability, nonetheless non-motor symptoms, as sleep disorders and dementia, are common too.

Levodopa is the gold standard medication; its effectiveness is remarkable at the beginning, but it decreases with the progress of the therapy. Besides, motor complications appear as side-effects of the medication causing dyskinesia (i.e., involuntary movement) and motor fluctuations.

To reduce the medication drawbacks and optimize the doses of levodopa, periodical assessments at the hospital are necessary. However, a longitudinal and continuous monitoring of individuals with late-stage PD would increase the efficacy of the drug titration.

The use of wearable technologies can help the fulfillment of objective and longitudinal monitoring of motor symptoms and complications in unconstrained environments.

This work aims to investigate the feasibility of a system able to track the severity of bradykinesia and motor fluctuations in naturalistic settings based on wearable sensors and Machine Learning (ML) techniques.

The dataset used for the study is part of the Blue Sky project and includes 25 participants with late-stage PD. The data are gathered during two visits, one in the laboratory and the other in a simulated apartment, using accelerometer sensors placed on the wrists and ankles. The tasks belong to standardized and activity of daily living (ADL) and the clinical scores follow the Unified Parkinson's Disease Rating Scale (UPDRS).

The data processing consists in filtering, resting period removal, signal segmentation, feature extraction, data cleaning, and training of the Random Forest (RF) algorithm. The signals are filtered to retain the frequency components related to bradykinesia. After the removal of the rest periods, the continuous signals are segmented into 5s windows to simplify the analysis.

Once determined the feature set, the predictors are extracted from the windows and the redundant predictors are discarded applying a proposed method based on feature correlation and ReliefF. Four movement patterns groups are identified to reduce the complexity of the bradykinesia estimation.

Data cleaning to discard outliers and improve class separation is implemented before the training of the RF regressor. The same pipeline is applied to the apartment data adding a movement pattern classifier before the model estimate.

The cross-validation (CV) results on the laboratory data are obtained using k-fold and leave-one-subject-out (LOSO). The performance is measured in terms of root mean square error (RMSE) for regression tasks; whereas accuracy, specificity, and sensitivity are used for classification assignments.

The validation results are encouraging, the overall RMSE is 0.5 in regression range between 0 and 3 with the LOSO CV, and the test results are promising for the longitudinal monitoring of bradykinesia and motor fluctuations in-home setting during ADL with a RMSE of 0.89.

In conclusion, this work demonstrates the feasibility of this approach applied in natural settings despite some limitations, such as the low number of subjects and a reduced amount of labels. This result can be a starting point for future improvements in PD severity monitoring.

Dedication

I dedicate this work to my father Mauro and my mother Anna Maria for allowing me to take this journey and for supporting me during this period away from home. Thanks to my girlfriend Giulia, who managed to stay close to me even though there was a whole ocean between us.

I can not thank my best friends Lorenzo and Pietro for sharing important moments throughout our lives.

Finally, a special thanks to all my family and my true friends.

Dedico questo lavoro a mio papà Mauro e a mia mamma Anna Maria per avermi dato la possibilità di poter intraprendere questo percorso e per avermi sostenuto durante questo periodo lontano da casa.

Grazie alla mia fidanzata Giulia, che é riuscita a starmi vicino nonostante ci fosse un intero oceano tra di noi.

Non posso che non ringraziare i miei migliori amici Lorenzo e Pietro per aver condiviso momenti importanti durante tutta la nostra vita.

Infine, un ringraziamento speciale a tutta la mia famiglia e ai miei veri amici.

Acknowledgements

I would like to thank prof. Danilo Demarchi and prof. Paolo Bonato for giving me the extraordinary opportunity to complete my studies in the city of Boston, at the Motion Analysis Lab in the Spaulding Rehabilitation Hospital.

I would also like to thank Federico Parisi and Stefano Sapienza for following me during the entire period of work.

Finally, thanks to all the boys of the MAL for the fantastic moments spent together.

Acronyms

ADASYN Adaptive Synthetic

ADL Activity of Daily Living

AI Artificial Intelligence

AUROC Area Under Receiver Operating Curve

BBB Blood Brain Barrier

CFS Correlation Feature Selection

CNN Convolutional Neural Network

CV Cross-validation

DBS Deep Brain Stimulation

DC Direct Current

DL Deep Learning

DTFT Discrete Time Fourier Transform

EM Expectation Maximization

FN False Negative

FP False Positive

GABA Gamma-Aminobutyric Acid

GPe Globus Pallidus external segment

GPi Globus Pallidus internal segment

HAR Human Activity Recognition

HP Highpass

IIR Infinite Impulse Response

IMU Inertial Measurement Unit

IQM Inter-Quartile Range

IQR Inter-Quartile Mean

LOO Leave-one-out

LOSO Leave-one-subject-out

LP Lowpass

MAD Mean Absolute Deviation

MAE Mean Absolute Error

MDS-UPDRS Movement Disorder Society Unified Parkinson's Disease Rating Scale

MEAD Median Absolute Deviation

ML Machine Learning

mRMR minimum Redundancy Maximum Relevancy

OOB Out-of-bag

PCA Principal Component Analysis

PD Parkinson's Disease

PSD Power Spectral Density

RF Random forest

RMS Root Mean Square

RMSE Root Mean Square Error

SADL Scripted Activity of Daily Living

SMA Signal Magnitude Area

SMOTE Synthetic Minority Over-Sampling Technique

SNE Stochastic Neighbor Embedding

SNc Substantia Nigra Compacta

SNr Substantia Nigra Reticulata

STN Subthalamic Nucleus

SVM Support Vector Machine

TN True Negative

TP True Positive

t-SNE t-distribution Stochastic Neighbor Embedding

UPDRS Unified Parkinson's Disease Rating Scale

Contents

Abstract	ii
Dedication	iii
Acknowledgements	iv
Acronyms	vii
List of Figures	xi
List of Tables	xii
1 Introduction	1
1.1 Parkinson's Disease	3
1.1.1 Pathophysiology	3
1.1.2 Symptoms	4
1.1.3 Diagnosis	7
1.1.4 Etiology	8
1.1.5 Treatments	9
1.1.6 Clinical assessment	9
1.2 Motor Fluctuations	10
1.3 Wearable sensors	12
1.3.1 Accelerometer sensor	12
1.4 Machine Learning algorithms	13
1.4.1 Supervised Learning	14
1.4.2 Unsupervised Learning	16
1.5 Conclusions	17
2 State of the Art	18
3 Material and Methods	22
3.1 Protocol and Data collection	23
3.1.1 Sensors	23
3.2 Signal pre-processing	26
3.2.1 Filtering stage	27
3.2.2 Signal segmentation	28
3.3 Movement detection	29
3.4 Windowing	30
3.5 Feature extraction	31
3.5.1 Time domain features	32

3.5.2	Frequency domain features	33
3.5.3	Segment velocity features	34
3.5.4	Additional features	35
3.6	Laboratory dataset analysis	37
3.6.1	t-distribution Stochastic Neighbor Embedding	37
3.6.2	Data visualization	38
3.7	Laboratory dataset cleaning	41
3.7.1	Redundant feature removal	43
3.7.2	Data cleaning	44
3.8	Data Balancing	46
3.8.1	Synthetic Minority Over-sampling Technique	46
3.8.2	Adaptive Synthetic sampling	46
3.9	Learning Algorithm	48
3.9.1	Random Forest	49
3.9.2	Cross-validation approaches	50
3.9.3	Performance measures	50
3.10	Apartment data analysis	53
3.11	Conclusions	55
4	Results	56
4.1	Filtered signals	57
4.2	Movement detection validation	57
4.3	Feature selection validation	59
4.4	Data cleaning results	60
4.4.1	Selected features	60
4.4.2	Data cleaning	62
4.5	Data Balancing results	63
4.6	Learning algorithm parameters	65
4.7	Laboratory bradykinesia prediction	66
4.7.1	Bradykinesia prediction after cleaning	66
4.7.2	Bradykinesia prediction	66
4.8	Apartment bradykinesia prediction	71
4.8.1	Movement classifier	72
4.8.2	Longitudinal bradykinesia prediction	73
4.9	Conclusions	75
5	Discussion	78
6	Conclusions	81
	Bibliography	87

List of Figures

1.1	Basal Ganglia region	4
1.2	Connections of basal ganglia	5
1.3	Lewy bodies in histopathological analysis	5
1.4	MDS-UPDRS assessment protocol	10
1.5	Motor fluctuation cycle	11
1.6	Accelerometer sensor model	12
1.7	Capacitive Accelerometer Design	14
1.8	Supervised learning approach	15
1.9	Unsupervised learning approach	16
3.1	Laboratory Protocol	24
3.2	Apartment Protocol	25
3.3	OPAL device	26
3.4	Protocol sensor positions	26
3.5	Filter frequency responses	28
3.6	Laboratory task distribution	29
3.7	t-SNE projection of upper and lower limbs data	39
3.8	t-SNE projection of bradykinesia severity	39
3.9	t-SNE projection of ADL tasks	40
3.10	t-SNE projection of the task clusters	41
3.11	Gross movement t-SNE projection with tasks	41
3.12	Gross movement t-SNE projection with bradykinesia severity	42
3.13	Laboratory dataset partitioning	42
3.14	Data cleaning scheme	45
3.15	Synthetic Minority Over-sampling Technique	47
3.16	K-fold cross-validation	51
3.17	Confusion matrix	52
3.18	Apartment pipeline	54
4.1	Raw and filtered signals	57
4.2	Raw and filtered periodograms	58
4.3	Movement detection validation	58
4.4	Movement detection example	59
4.5	Feature selection validation	60
4.6	ReliefF feature importance	61
4.7	Correlation matrix	61
4.8	Correlation matrix after redundant feature removal	62
4.9	Feature importance after redundant feature removal	62
4.10	Gross movement t-SNE projection before cleaning	63

4.11 k-means clustering outcome	63
4.12 Gross movement t-SNE projection after cleaning	64
4.13 Data balancing results	64
4.14 t-SNE projection of the gross movement cluster after balancing	65
4.15 OOB error curves	66
4.16 Data cleaning results of laboratory data	67
4.17 Laboratory lower limbs confusion matrices using k-fold	68
4.18 Laboratory lower limbs confusion matrices using LOSO	68
4.19 Laboratory upper limbs during walking results using k-fold	69
4.20 Laboratory upper limbs during walking results using LOSO	69
4.21 Laboratory upper fine movement results using k-fold	70
4.22 Laboratory upper fine movement results using LOSO	70
4.23 Laboratory upper gross movement results using k-fold	71
4.24 Laboratory upper gross movement results using LOSO	72
4.25 Selected features of the movement recognition classifier	73
4.26 OOB error curve of the movement recognition classifier	73
4.27 Confusion matrices of fine vs. gross movement classifier	74
4.28 Apartment estimation error over the subjects	75
4.29 Longitudinal bradykinesia prediction result with medication intake	76
4.30 Longitudinal bradykinesia prediction result without medication intake	77

List of Tables

1.1	Parkinson's disease symptoms	6
1.2	Parkinson's disease diagnostic criteria	8
3.1	Laboratory Tasks	24
3.2	Apartment Tasks	25
3.3	OPAL Features	27
3.4	Filter design specifications	28
4.1	Random Forest parameters	65
4.2	Laboratory performance lower limbs	67
4.3	Laboratory performance upper limbs during walking	70
4.4	Laboratory performance fine movements	71
4.5	Laboratory performance gross movements	72
4.6	Laboratory performance classifier for gross and fine movements recognition	74
4.7	Apartment overall estimation error	74

Chapter 1

Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative disorders in the world. Estimates suggest 10 million people are affected by PD worldwide [1]; the amount is going to increase due to the aging progression in developed countries. This thesis is part of a bigger project, called Blue sky, a result of the collaboration of *Motion Analysis Lab, Harvard Medical School, Spaulding Rehabilitation Hospital, Pfizer Inc.* and, *IBM T.J. Watson Research Center*.

The Blue Sky project aims to design a telemonitoring system, to estimate the PD severity using wearable sensors artificial intelligence (AI) methods. Exploiting the predictions, the clinician could monitor the subject constantly, without needing an assessment at the hospital. Besides, the gathered information could lead the physician to a deeper comprehension of the subject state, enabling a personalized and optimized drug treatment strategy.

According to the goals proposed by the project, the PD assessment would be easier, available everywhere, without the presence of a clinician, and constant. In addition, longitudinal monitoring is more suitable to assess the severity of the disease and adjust the drug treatment.

The main advantage of the implementation of the telemonitoring system would be addressed to the patients, providing them a quantitative measure of the severity deleting the subjectivity and bias typical in common clinical assessment, but also the entire society because this system will reduce the cost of the periodical hospital assessments.

The project involves two data collections, the first one in the laboratory, while the second in a simulated apartment. The data has been gathered by inertial measurement units (IMUs) placed on the wrists and the ankles of the subjects, during activities of daily living (ADL) and standardized tasks. A team of clinicians evaluated the performance of the tasks according to the Unified Parkinson's disease Rating Scale (UPDRS) in both the settings.

This work focuses on the PD cardinal symptom, the bradykinesia; the definition of bradykinesia is the progressive slowness of movement during voluntary repetitive activities [2].

The main aim of the thesis is to design a machine learning (ML) predictive model, based on the data collected in the laboratory, to estimate the bradykinesia severity of the subjects in the apartment setting.

During my period spent at the *Motion Analysis Lab, Harvard Medical School, Spaulding Rehabilitation Hospital* in Boston (MA, United States), I dealt with the

signal pre-processing, the feature engineering and feature extraction, the supervised and unsupervised ML approaches to process the data to build a robust predictive model and, the testing phase on the data gathered in the simulated apartment.

In the following sections of this chapter, some details about Parkinson's disease and motor fluctuation will be explained and a general overview of wearable sensors and machine learning algorithms will be provided.

The next chapters are set up as follow:

- Chapter 2: state of the art overview, taking into account the major works in literature published in the last years;
- Chapter 3: material and methods applied to develop the analysis of the data and the predictive model;
- Chapter 4: reporting the results;
- Chapter 5: discussion and comments on the work;
- Chapter 6: conclusion and future advancements on the themes covered by the thesis.

1.1 Parkinson's Disease

PD is a gradual neurodegenerative disorder. It is the second most frequent after Alzheimer's disease among neurodegenerative disorders. It affects, especially, people from 60 years of age, but the onset of the disorder could be earlier. Its predominance is 1% in the population over 60 years [3] and around the 4% over 80 years [4].

To quantify with absolute values, there are 7-10 million people affected by PD worldwide [1]; the overall trend estimate for the future years, according to the constant aging of the global population is around 14 million by 2040 [5].

This trend explains the significant interest in the field of PD research, also because there is no definitive treatment for the disease and the real causes of this disorder are not discovered yet. The majority of PD cases are labeled as idiopathic and only a few cases have genetic reasons.

1.1.1 Pathophysiology

PD occurs as degeneration of the extra-pyramidal structures in the nervous system. These structures are responsible for the voluntary movements. The extra-pyramidal and the pyramidal systems work concurrently, allowing the execution of the voluntary movements.

The decline includes a particular region, located in the basal ganglia area, the substantia nigra (figure 1.1). This anatomical zone is located between the midbrain and the diencephalon, and has this name because there is a high concentrations of melanin pigment.

Basal ganglia handle the likelihood of movement taking place, comparing cortical processing with the overall condition of the nervous system [6].

Basal ganglia have links toward areas in the brain stem and these connections constitute the nigro-striatum motor pathway (figure 1.2).

Motor pathway is divided into two paths: the direct pathway and the indirect pathway; the substantia nigra plays a fundamental role for the correct working in both of them. The inputs to the striatum come from only the cortex area.

The direct pathway is managed by the neurotransmitters Gamma-aminobutyric acid (GABA) and Glutamate; the information streams from the striatum to the globus pallidus internal segment (GPi) and the substantia nigra reticulata (SNr). After that, these regions communicate with the thalamus and, this area sends instructions to the cortex closing the loop.

On the other hand, the indirect pathway proceeds from the striatum to the globus pallidus external segment (GPe); the GPe sends the stream of information to the subthalamic nucleus (STN) and, afterwards to the area of the GPi and SNr. From this region, the path is the same as the direct pathway.

The direct pathway disinhibits the thalamus; instead, the indirect pathway enforces the regular inhibition of the thalamus. The disinhibition of the thalamus causes the excitation of the cortex, meanwhile, the inhibition strengthening induces an inhibition of the cortex.

The decision to use the direct or the indirect pathway is controlled by the substantia nigra pars compacta (SNc). This area sends information to the striatum by the neurotransmitter dopamine; the direct pathway is triggered by dopamine, on the other hand, the indirect path is repressed by dopamine. If the SNc is active the

likelihood of movement will be higher; rather if the it is quiet the motion is inhibited [6].

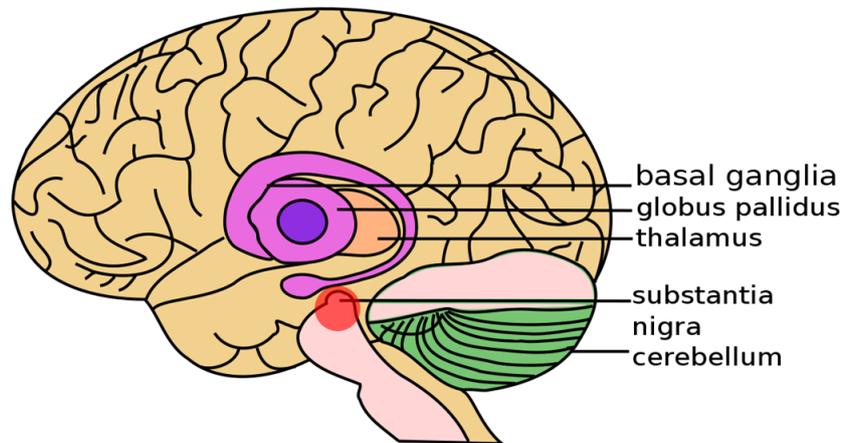


Figure 1.1: Lateral view representation of the brain. The violet area is the basal ganglia region, while the orange one is the thalamus. The red circle indicate the position of the substantia nigra. Adapted from [7].

PD affects the SNc, causing motor and non-motor impairments. Motor issues are clearly explained by the deficit in dopamine level, which triggers one of the two pathways.

The first principal neuropathological finding in PD is the loss of neurons in the pars compacta [2]. The death of these cells is related to a deficit of dopamine in this area, since the dopamine controls the movements through the nigro-striatum pathway, the lack of this neurotransmitter causes not only the deterioration of the motion actions but also cognitive problems. It is estimated that a subject affected loses 50 % to 70 % of the entire neurons of the SNc during his life [4].

The second neuropathological finding is the presence of protein deposits formed by α -synuclein inside the brain cells called Lewy bodies. The Lewy bodies move the other cellular structures from the correct position and their distribution inside the brain tissue is strictly correlated to the expression and the severity of the clinical symptoms. The Lewy bodies are a clear mark of PD; in fact, the diagnosis of PD could be confirmed only during the autopsy by histopathological analysis [2]. If during the examination Lewy bodies are found in the brain tissue the clinical diagnosis will be proved (figure 1.3). This procedure is considered the standard gauge of judgment for the PD diagnosis.

1.1.2 Symptoms

PD symptoms are both motor and cognitive and they are caused by the lack of dopamine in the SNc and by the loss of neural cells.

The hallmark symptoms are motors because they are the first to appear and they induce a significant reduction of quality of life in people affected by PD.

All the symptoms occur gradually and not altogether; as the disease progresses the severity increases and new symptoms arise.

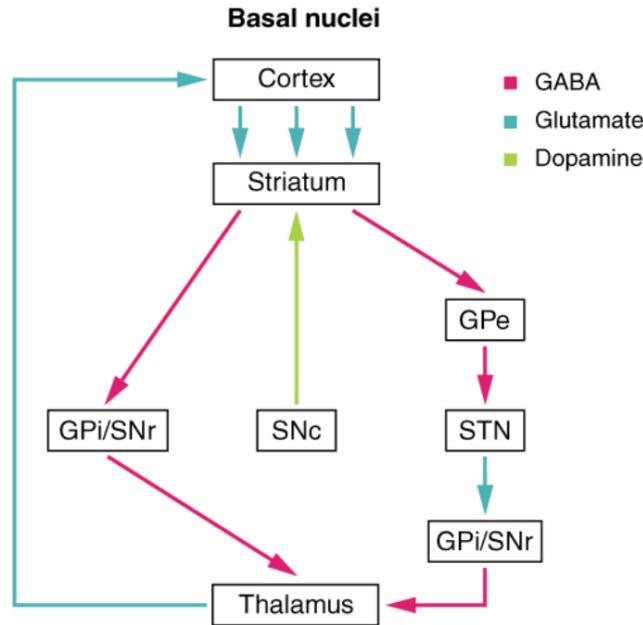


Figure 1.2: Illustration of the basal ganglia, or basal nuclei connections. The cortex controls the striatum using the neurotransmitter glutamate and the striatum has two different paths to tune the cortex actions. The direct pathway involves the striatum, the regions of globus pallidus internal (GPi) and the substantia nigra reticulata (SNr); the indirect pathway includes the globus pallidus external (GPe), the subthalamic nucleus (STN) and the area of GPi and SNr. The substantia nigra compacta (SNc) manages which of the two paths would modulate the cortex. This information is carried by the neurotransmitter dopamine towards the striatum [6]. Retrieved from [6].

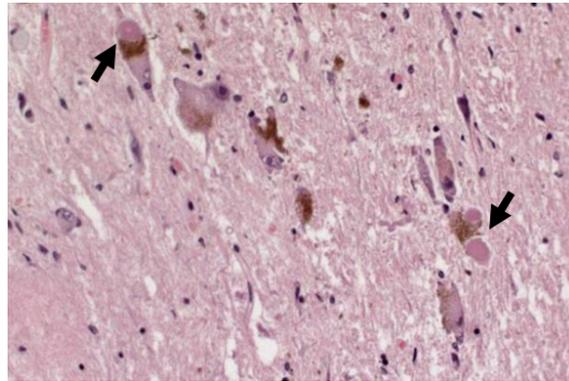


Figure 1.3: Histopathological image of the neuronal tissue with the presence of the Lewy bodies. The protein agglomerates are marked by black arrows to distinguish them from the other tissues; the presence of the Lewy bodies confirms the clinical diagnosis of PD in autopsy.

In table 1.1 the principal symptoms are listed and in the next sections, each symptom will be described more.

The cardinal symptoms of PD are bradykinesia, tremor, and rigidity; nonetheless, they are taken into account for the clinical diagnosis.

Table 1.1: Motor and non-motor symptoms in Parkinson's disease.

Motor Symptoms	Cognitive Symptoms
Bradykinesia	Sleep
Tremor at rest	Cognition
Rigidity	Dementia
Postural Instability	Mood Disturbance
Freezing of Gait	Psychosis
Hypokinesia	Confusion
Akinesia	

Bradykinesia

Bradykinesia is one of the cardinal symptoms of PD, and it is also the most important clinical sign to diagnose the disorder itself.

The clinical definition of bradykinesia is the progressive slowness of movement and decrease of range of motion during repetitive voluntary movements. Indeed, bradykinesia affects the daily activity of the patients hampering the planning and the success of the motions.

The main impairments related to bradykinesia are the decline of automatic movements (i.e., arm swings during walking), difficulty to start movements, general slowness, and abnormal stiffness or diminished facial expression [8]. This manifestation alters the fine motor control [2] as well, for example using kitchen tools, buttoning clothes, or writing.

Bradykinesia is the hallmark sign of basal ganglia degeneration, especially this manifestation is the consequence of the reduced dopaminergic function in the SNc. This symptom may affect only one limb, one-half of the body or even the entire body. According to the features of bradykinesia, the typical tasks for the assessment are repetitive movements, such as finger tapping, finger to nose and alternating hand movement, which consists of a sequence of pronation and supination of the hands. Those tasks allow the clinician to evaluate the severity of the symptom and the overall state of the disorder.

Tremor

Tremor is the most distinctive symptom of PD. It appears during rest with a periodical tremor of the distal part of the limbs. Generally, tremor affects mostly the upper part of the body, but it could manifest in the legs too.

The main characteristic of this symptom is the frequency of the periodical involuntary movements; the values are estimated between 4 Hz to 6 Hz [2]. A secondary aspect is the absence of tremor during voluntary movements and sleep.

Tremor has a fast onset and could change the severity quickly. To estimate the severity of this symptom, the physician should observe the patient during resting periods monitoring the amplitude and the frequency of the manifestation.

Rigidity

Rigidity is one of the hallmark symptom of PD together with bradykinesia and tremor. Rigidity is defined as stiffness of the limbs and joints; this manifestation decreases the range of motion and may cause pain.

Rigidity could not affect only the limbs, but the trunk and the face too. A huge issue of rigidity is the repercussion on the quality of the sleep [8].

Other Motor Symptoms

The other motor symptoms are not less important, but they are less frequent because they occur in the late-stage of the disorder.

Postural instability is the most dangerous symptom because it could cause the subject to fall during walking. The postural instability is due to rigidity and loss of promptness; this means less balancing control while upright posture by the subject. Another cause of falling is the freezing of gait; it could be described as a motor block. The freezing occurs mostly during walk, but it could manifest in the upper part of the body too.

Hypokinesia and akinesia are manifestations quite similar to bradykinesia, nevertheless the former has referred to the inability to perform movements, the latter is the incapacity to start voluntary movement. Those two symptoms are a clear mark of degeneration of SNc; besides, they appear in the late stage of the disease when the loss of dopaminergic neurons is important.

Cognitive Symptoms

Despite these symptoms seem to be on the background compared to motor ones, their incidence of occurrence in PD patients is around 40% [4], and they can not underestimate.

The majority of these symptoms appear in the late stage; for example, sleep disorders are related to a specific motor manifestation, the rigidity. Other cognition dysfunction could be dementia, depression, and mood disturbance, but the real reasons are not well-known [2] by researches and clinicians.

1.1.3 Diagnosis

Up to now, there is no exhaustive analysis for the diagnosis of PD, and it is based essentially on clinical features.

The clinical features are, at first glance, the motor symptoms, because they are the most noticeable signs in people with PD; besides, other considerations are necessary about the response of the medication and the exclusion criteria. The real challenge in PD diagnosis is to recognize the disorder in the early stages, but this could be tough because some symptoms are in common with other diseases [2].

To diagnose PD, the subject must manifest bradykinesia and at least another hallmark between tremor and rigidity; then, the subject must have at least two supportive criteria and the absence of all the exclusion criteria [9]. In table 1.2 the exclusion and the supportive criteria for the diagnosis of PD are shown.

Table 1.2: Diagnostic criteria of the Parkinson's disease; they include the exclusion and the supportive criteria. Adapted from [4].

Exclusion Criteria

History of repeated strokes
Supranuclear gaze palsy
History of repeated head injury
Cerebellar signs
History of definite encephalitis
Early severe autonomic involvement
Oculogyric crises
Early severe dementia with disturbances of memory, language and praxis
Neuroleptic treatment at the onset of symptoms
Babinski's sign
More than one affected relative
Presence of cerebral tumour or communicating hydrocephalus on CT scan
Sustained remission
Negative response to large doses of levodopa (if malabsorption excluded)
Strictly unilateral features after 3 years

Supportive Criteria

Unilateral onset
Rest tremor present
Progressive disorder
Persistent asymmetry affecting side on onset most
Excellent response to levodopa (70 % to 100 %)
Severe levodopa-induced chorea
Levodopa response for 5 years or more
Clinical course of 10 years or more

1.1.4 Etiology

The origin of PD is not yet understood for the most analyzed cases. For this reason, PD is defined as idiopathic disorder. Other causes rely on genetic risk factors, but only for 5% of the PD population, the genetic cause could have been related [9].

The main genetic source is given by the LRRK 2 gene; this gene encodes the protein dardarin, which is correlated with the asymmetric onset and frequent tremor. A second gene mutation regards the SNCA gene; this gene is related to the synthesis of the protein α -synuclein, which generates the protein agglomerates, called Lewy bodies, in the brain cells. The overexpression of this gene leads to an increased production of the protein and a high likelihood to develop Lewy bodies.

Finally, another cause for PD is in mitochondrial genetics. One of the reasons of the brain cells' death is the high oxidative damage sensitivity of the SNc. A mitochondrial dysfunction leads to oxidative phosphorylation defect in PD, hence to the death of the brain cells [4]. A mutation in the PINK 1, that synthesizes the mitochondrial complex may lead to the failure in the antioxidative mechanism.

The onset of PD may be associated with environmental factors [3]. According to epidemiological studies, a correlation between the risk of PD and the exposure to pesticides, herbicides, and heavy metals are discovered. These factors lead to dopamine depletion in case of herbicides exposure and increased oxidative stress if heavy metals concentrations are high in the SNc [3].

1.1.5 Treatments

The main goal of the PD treatment is to mitigate the symptoms, especially during the initial stages of the disorder. In this way, the subject is allowed to carry out daily activities without or with reduced impairments. Besides, the treatment must be tolerated by the subject and must trigger the less amount of side-effects.

The gold standard medication is the levodopa (L-dopa), a precursor of the dopamine. This drug is very effective at the beginning of the treatment and, reduces the severity of the symptoms. However, only 5% to 10% of the L-dopa intake can pass through the blood-brain barrier (BBB) and, turned into dopamine, ready to be absorbed by the dopaminergic neurons, by using the DOPA-decarboxylase. The remaining part goes into other districts causing side-effects such as nausea, dyskinesia and joints' rigidity.

The long-term treatment with L-dopa complicates the state of the patient adding new issues, such as dyskinesia, sudden involuntary movements caused by an excessive dose of L-dopa and the motor fluctuations. Motor fluctuations refer to different states of the subject before and after the medication intake. In those states, the subject could experience the total absence or a reduced presence of the symptoms otherwise a significant severity. Besides, the motor fluctuations features are related to the dose of medication and, the disorder duration [4].

1.1.6 Clinical assessment

Up to now, the gold standard to assess PD is the clinical assessment according to the Movement Disorders Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS). The clinical assessment typically takes place at the hospital led by a neurologist, and it is based on the MDS-UPDRS protocol (figure 1.4). Besides, the overall outcome is the ensemble of the subject's feelings and, the clinician's judgments.

The judgment of the neurologist is a score given to each task of the protocol between 0-4; score 0 means absence of the impairment, on the other hand, score 4 suggests an important hampering to perform the assigned task.

The clinical assessment could not be considered a perfect prediction of the severity of the disease, because due to quantitative measures, the outcome is a general idea of the subject's status [5].

The MDS-UPDRS protocol attempts to span all the aspects of the disease's clinical manifestations; in fact, it has four sections divided as follows:

- Part I: assessment of non-motor symptoms during daily living;
- Part II: examination of motor aspects in daily living;

- Part III: motor examination, assessing the hallmarks motor symptoms using standardized tasks. For example, the bradykinesia assessment has performed using finger-to-nose movements and alternating hand movements (pronation-supination of the hands);
- Part IV: assessment of motor complications due to drug treatment.

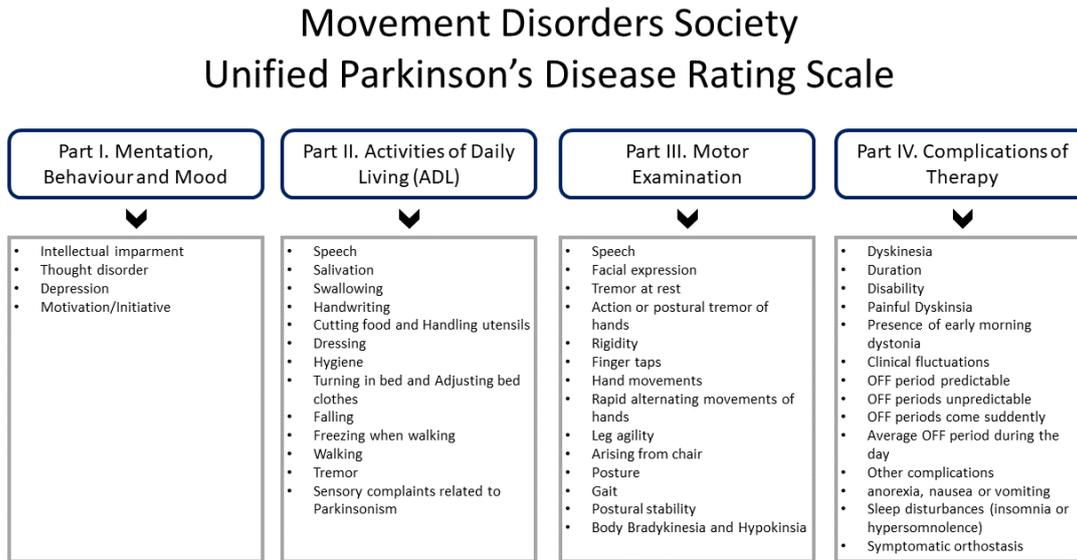


Figure 1.4: The MDS-UPDRS covers all the points for the PD assessment. The score for each item is between 0-4. The part I takes into account the non-motor symptoms in daily living, while the part II assesses the motor complication during activities daily living; the part III is the motor examination, where the major motor symptoms are assessed by standardized tasks and finally the part IV consists in observe the therapy drawbacks and the motor fluctuations. Retrieved from [10].

However, thanks to an important advancement in the technological field new forms of assessment have risen. The most popular technologies to assess quantitatively the PD severity are accelerometer, gyroscopes and magnetometer sensors [5]; these tools allow to get objective measurement, improving the repeatability and the overall accuracy of the examination.

1.2 Motor Fluctuations

Motor fluctuation is not a real symptom of PD, but it is defined as motor complication. It is related to the medication treatment because the gold standard drug is L-dopa, and as time progress the side-effects become significant as the PD symptoms. Studies observed subjects experienced motor complications after 5 years of L-dopa therapy with an incidence of 40% [11].

Motor fluctuation describes the alternation of the patient's status along before, during and after the medication intake [12]. The states of the fluctuation are OFF, wearing ON, ON and, wearing OFF (figure 1.5).

- **OFF state:** it occurs when the subject is no under the effect of the medication; this means the patient experiences the common PD motor symptoms;
- **Transition to ON:** it is the transition phase between the state OFF and ON; it corresponds to the medication intake.
- **ON state:** it is the timing when the drug is working, increasing the dopamine level in the substantia nigra. During this period, the subject has a heavy reduction or the disappearance of the symptoms' severity.
- **Wearing OFF:** it is the transition between the ON and the OFF states. This state is also known as end-of-dose deterioration. Over this phase, the motor symptom severity increases meaning that the level of dopamine is lower and the control for the motor pathway becomes harder.

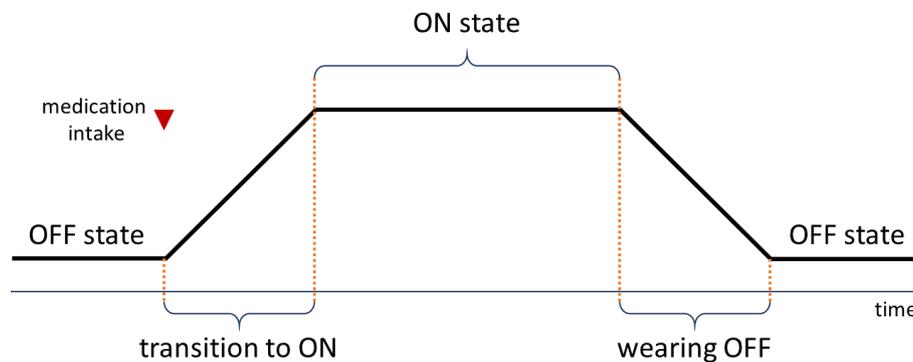


Figure 1.5: Motor fluctuation cycle illustration. The state before the medication intake (red triangle) is defined OFF; after the transition, the level of dopamine in the substantia nigra is high, this state is called ON. At the end of the drug effect, the wearing OFF transition takes place till a new state OFF period. Retrieved from [13].

This fluctuation between OFF and ON states causes movement impairments during the daily living of the patients. These oscillations could have avoided optimizing and personalizing the medication treatment.

Another drawback of L-dopa is the onset of a motor complication during the ON phase. This side-effect is called dyskinesia and its definition is a sudden and involuntary movement that could affect a single limb or the whole body [11].

The real cause of the dyskinesia onset is still unknown, however, a plausible explanation might be the continuous fluctuation of dopamine level due to the several medication intake during the day. The physiological reason relies on the reduced quantity of medication that passes through the BBB; all the remaining dose goes in other areas where it is converted into dopamine; this causes the involuntary choreic movements.

1.3 Wearable sensors

Wearable sensors have been a success since their appearance in the market. Thanks to technological progress into miniaturization of the Micro-electromechanical systems (MEMS), it was possible to condense different sensors on a single device.

Wearable sensor purposes are various, from the health tracking of normal people to the monitoring of individuals with motor impairments.

Typically, wearable sensors consist of accelerometer, gyroscope, and magnetometer and they are capable to communicate with other devices, such as smartphones, tablets, and computers using wireless communication, like Bluetooth.

The reason why they are so popular is the movements are not obstructed by wires and their low weight. Besides, these devices are easy to use because they do not need a particular setup to work properly.

Since in this work, the exploited sensors are accelerometers a deeper discussion about these kinds of sensors in terms of working principle and design will be presented in the next section.

1.3.1 Accelerometer sensor

The accelerometer sensor is a device that allows to measure the body's acceleration by a moving proof mass. In general, the system consists of a seismic mass, a spring and a damper (figure 1.6). The sensor exploits the mass inertia during a movement, hence the displacement of the proof mass is measured, respect to the fixed device's structure, and this quantity corresponds to a given acceleration using the second Newton law equation 1.1

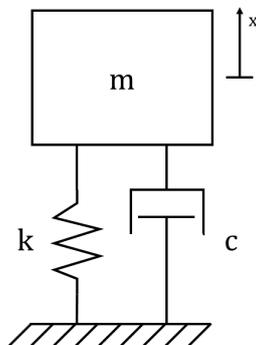


Figure 1.6: The accelerometer sensor is modeled as a mass, spring, damper system. According to this model the mass could move due to external forces; the displacement of the mass between the reference system origin is measured to derive the acceleration values.

$$m \frac{d^2 x}{dt^2} + c \frac{dx}{dt} + kx = mg, \quad (1.1)$$

where x is the displacement, m is the mass, c is the damping coefficient of the damper, k is the elastic constant of the spring and g is the gravitational acceleration. The mechanical displacement is converted into an electrical quantity by a transducer; the common transducers used in accelerometer sensors are:

- **Piezoelectric crystals:** this device has the characteristic of generating a ΔV as a result of a Δx , where V and x are the voltage and the displacement respectively;
- **Piezoresistive:** the displacement is detected by the resistivity changing of the semiconductor device;
- **Capacitive:** the gap between the two plates of a capacitor is inversely proportional to the capacitance value C (equation 1.2)

$$C = \frac{\epsilon_0 \epsilon_r A}{x}, \quad (1.2)$$

where ϵ_0 is the vacuum permittivity, ϵ_r is the relative permittivity of the medium, A is the overlapping surface area of the capacitor's plates and x is the displacement between the two plates.

The focus is on the capacitive accelerometer sensors because the OPAL device has inside this type of technology.

Capacitive accelerometer sensor

The reason why this kind of technology is the most popular is that the measurement accuracy and the stability are significant; besides, the power dissipation is negligible compared with the other fabrication methods and this sensor has a very low sensitivity to noise and temperature fluctuations.

The sensor consists of a proof mass, equipped with fingers, anchored to the fixed structure of the device by two springs. On the two sides of the proof mass, there are other fingers, such as the proof mass, but these are fixed. The overall layout is presented in figure 1.7.

The design of the sensor is due to generate a greater capacitance variation, because with this architecture there are many capacitors in parallel configuration; if the capacitors are in parallel, the system is equivalent to another one with a single capacitor with a capacitance value equal to the sum of all the capacitance values of the previous system.

When the system is moving with a non-zero acceleration, the proof mass is moving too, causing a variation of the displacement among the moving and the fixed fingers. This displacement modifies the capacitance value of the system, and according to the equations 1.1 and 1.2, the acceleration value can be estimated.

Obviously, the description of the device is only for one axis; to build a triaxial accelerometer is necessary to replicate the same structure on the other two axes.

1.4 Machine Learning algorithms

Machine Learning (ML) is part of the artificial intelligence (AI) field, which is the field where computer systems are designed with the ability to learn from experience like humans do.

Two definitions outline the ML field:

"Field of study that gives computers the ability to learn without being explicitly programmed" (A. Samuel, 1959)

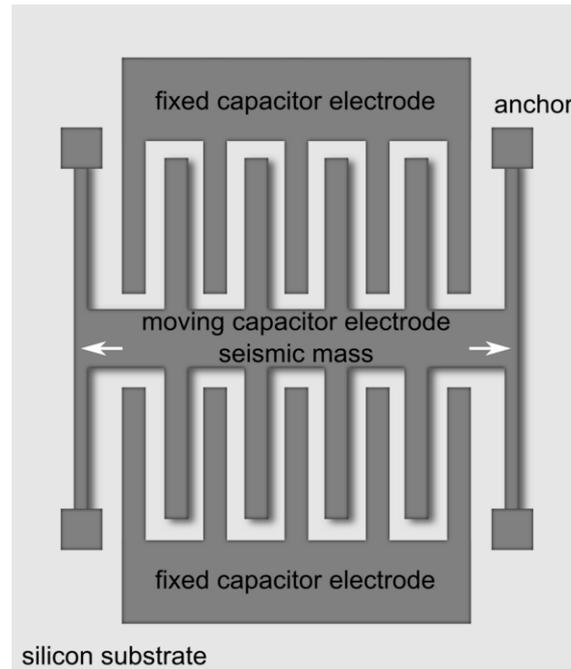


Figure 1.7: Representation of the capacitive accelerometer design. The seismic mass (proof mass), in the middle, is anchored to the fixed silicon substrate; the anchor could be treated as a spring and a damper. On the two side of the seismic mass there are two fixed structures with fingers, these fingers are on the proof mass too. When the seismic mass is moving the distance among the fixed and the moving fingers changes according to the amount of acceleration given to the system.

"A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E ". (T. Mitchell, 1988)

The first definition is the informal one, on the other hand, the second statement describes the basic approach to deal with a problem using ML. To summarize the idea of ML, its goal is to design a mathematical model for a specific task, with the capability to predict the output of new data according to the knowledge previously learned or find structures or patterns inside data. There are different ML approaches according to the task and the kind of available data and the main method of ML are supervised learning and unsupervised learning.

1.4.1 Supervised Learning

The supervised learning is used when the dataset is labeled. A label could be a certain class, a discrete number or a continuous number; regarding the nature of the labels, different approaches could have exploited.

The final purpose of supervised learning is to predict the output according to the experience learned by the data. Usually, the experience of the mathematical model is given by the training set; instead, the new predictions are on new data, which constitutes the test set (figure 1.8).

The training set could be depicted as an $m \times n$ input matrix called X and an $m \times 1$ output array called Y .

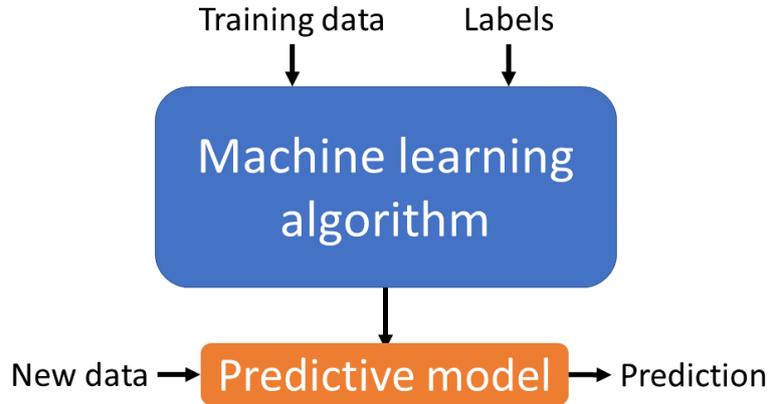


Figure 1.8: Supervised learning approach illustration. The training data and the labels form the experience for the mathematical model. After the learning phase the model is ready to predict the output of new data.

X is the ensemble of all the instances of the sample; each element has represented as a feature array; the feature array locates the element in the feature space. Each row of the matrix X is an instance and each column of the matrix X is a variable. Each element has a label; the ensemble of the labels constitutes the output array Y . The value of Y could be discrete, continuous or categorical.

In mathematical terms, the input matrix and the outputs could have summarized as:

$$D = \left\{ (x^{(i)}, y^{(i)}) \right\}_{i=1}^m, \quad (1.3)$$

where $x^{(i)}$ is the i^{th} sample and $x^{(i)} = x_1^{(i)}, \dots, x_n^{(i)}$.

Finally, the mathematical expression to explain the goal of supervised learning is:

$$Y = f(X) + \epsilon, \quad (1.4)$$

where Y is the output, X the input matrix, f is the model and ϵ is the intrinsic error of the model. According to the nature of the output Y , supervised learning has divided as classification and regression.

Classification

Classification refers to when the output Y has discrete values, such as $0, 1, 2, \dots, n$, or when the outputs are categorical. In this case, the goal of the learning algorithm is to find the proper boundaries among the classes in order to allocate new data points to the correct membership class.

An example of classification task could be to recognize if a mass is tumoral or not, according to some characteristics, like the size and the shape; this particular case belongs to the binary classification tasks because the model has to identify if a new instance belongs to a class or another one.

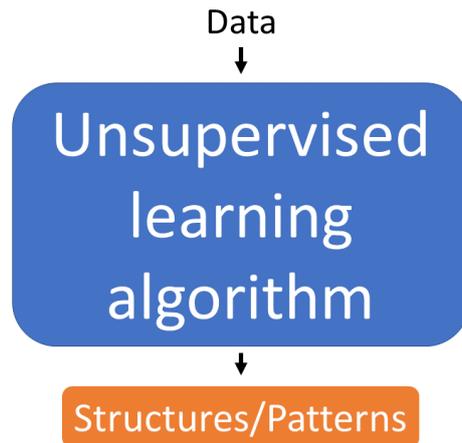


Figure 1.9: Unsupervised learning approach illustration. The input data are processed by the machine learning algorithm; the outcome of the model is how the data have been grouped according to their features.

Another example is to predict the weather, sunny cloudy or rainy, exploiting information like the temperature and the humidity. This task belongs to the multiclass classification because the learning algorithm has to deal with more than two classes.

Regression

Regression is suitable when the outputs Y are continuous values between a defined range.

The task of the learning algorithm is to fit the training data using a linear or non-linear function in order to predict the output values of new data.

An example of a regression problem is to predict houses' prices using some features such as the location and the number of rooms. Clearly, this is a regression because the prices are continuous values.

1.4.2 Unsupervised Learning

Unsupervised learning differs from the supervised by the lack of the output array. Hence, according to equation 1.5, the design of a dataset for unsupervised learning is:

$$D = \left\{ (x^{(i)}) \right\}_{i=1}^m \quad (1.5)$$

The purpose of unsupervised learning is to find out patterns and structures inside the gathered data to extract knowledge. There are two main approaches in unsupervised learning: clustering and dimensionality reduction.

Clustering aims to split into group the input data according to the characteristics of the instances.

On the other hand, dimensionality reduction attempts to represent the elements in the dataset in a lower dimension without losing information.

1.5 Conclusions

The aim of this work is to monitor the bradykinesia severity and the motor fluctuation in people with PD, exploiting wearable sensors and ML techniques, supervised and unsupervised, to design a model able to predict the motor impairment caused by bradykinesia during the daily life.

The supervised learning methods aim to predict the severity of bradykinesia symptom and the motor fluctuation too, exploiting features extracted by the accelerometer signals using a regression learning algorithm.

On the other hand, unsupervised approaches are primarily used to understand how the data have placed in the feature space and what are the relationships among the data.

In the next chapter, an explanation of the related work on the bradykinesia prediction will be presented, listing some of the main works published.

Chapter 2

State of the Art

The need to reduce as much as possible the L-dopa drawbacks has carried out feasibility and reliability studies for automatic assessment systems for individuals with PD, monitoring objectively and continuously to optimize the drug titration [14].

Usually, the clinical examinations are periodical and take place at the hospital; during the visit, the neurologist estimates the patient's status in a limited time frame and counts on the subject's diary, where motor complications manifestations are reported [15].

At a glance, it may perceive the limits of the clinical assessment, which is marked by a subjective evaluation and not long enough.

The chance of a long-term telemonitoring has been achieved with the introduction of the wearable sensors, which allow the collection of data without obstructing the subject's movements [14, 16, 17].

The remotely monitoring could be applied to every PD motor symptom, but a focus on bradykinesia and motor fluctuations tracking will be discussed. The most common wearable sensors used in this field are IMUs, in particular accelerometers and gyroscopes. These sensors can record the acceleration along three orthogonal axes and the angular velocity around the same orthogonal axes, during any kind of motion.

The first published works about the prediction of PD symptoms using IMUs exploited the UPDRS-III movement tasks because these specific movements can highlight the manifestation and the severity of the motor symptom; besides, these motor tasks are very simple to perform and to analyze due to their intrinsic periodicity [18].

Other authors started using ADL to predict and monitor the severity of bradykinesia and motor fluctuations at the hospital [19]; the idea is to transfer this analysis to unconstrained settings to be able to track the symptoms remotely and get a better awareness of the patient's state.

After the first published works, other authors investigated different data collection setups, attempting various motor tasks, sensors in diversified areas of the body, and different environments like laboratory and patient's home.

For what concerns the motor tasks, they may be standardized, proposed by UPDRS and ADL, extracted by the common daily activity.

The standardized tasks for bradykinesia are finger-to-nose, finger tapping, repeated

hand movement, and alternating hand movement. These tasks are exploited to discriminate motor fluctuations status and bradykinesia severity [20, 21, 22].

On the other hand, ADL are less used because are not so clear to detect the PD symptoms. ADL are mainly used to design system for home monitoring, without forcing the subject to perform particular movement during the day.

Depending on the setting of the investigation, the type of sensors changes. Usually, in laboratory assessments, the trend is to use accelerometers, gyroscope or together as well; while, in unconstrained settings and long-term monitoring the choice is the accelerometer sensors because they are the most efficient in terms of power consumptions ensuring a long battery life, despite the small dimensions of the device. The number of sensors is defined by the examination setting as well. In laboratory investigations, the tendency is to use a huge number of sensors such as a body area network on the subjects' body made of 17 sensors [23] or place sensors on the fingers using for examples gloves [24]. Otherwise, the trend in naturalistic environment analysis is to use one sensor for each limb [25] or only one sensor place on the waist [26].

The final results of these works are various according to the setups described before. The most implement ML [18] or Deep Learning (DL) [27] pipelines to obtain predictions, usually continuous using regression algorithms, of the bradykinesia severity using as true labels the UPDRS scores [28] or a longitudinal classification of the patient's motor state [25]. Others prefer to generate severity indices based on the extracted features from the gathered signals [29, 23] proving that these indices are highly correlated to the UPDRS scores.

Patel et al. [18] in 2009 used a uniaxial accelerometer sensor place on the limbs to gather data during standardized tasks for bradykinesia and dyskinesia monitoring. The data collection took place in the laboratory investigating the motor fluctuations as well, gathering the data in OFF and ON states. The motor tasks were scored by clinicians according to UPDRS and the scores were used as true labels for the learning algorithm.

This study introduced the signal analysis into windows, proving the optimal window length is 5s and proposing invariant features respect to the motor tasks.

The results of this investigation are interesting getting 2.2% estimation error for bradykinesia prediction using a support vector machine (SVM) and suggesting the extension of this analysis to ADL.

Cancela et al. [19] in 2010 used wearable accelerometers on the limbs, trunk, and belt, during ADL for bradykinesia monitoring.

The data collection was in the laboratory and the ADL included walking, drinking and opening and closing a door. The final result of the analysis, using different learning algorithms, is between 70% and 86% of accuracy.

This work introduced the concept of the classifier outcome post-processing, adjusting the estimates avoiding that they can not change rapidly because the bradykinesia has a very slow dynamic.

Tzallas et al. [30] in 2014 designed a system, called PERFORM, for motor symptoms monitoring using accelerometers placed in the limbs and an embedded sensor with accelerometer and gyroscope in the waist.

The system consists of the gathering, storing and analyzing the data; the main purpose of this device is to link the clinician and the patient for more precise monitoring. The collection of the data is at the patients' home and the activities are ADL. The bradykinesia severity estimation reached 74.5 % of accuracy using SVM classifier.

Hammerla et al. [25] in 2015 proposed a double data collection, the first one was in the laboratory, and the second was in an unconstrained environment for one week. The investigation exploited DL using as data accelerometer signals gathered from the wrists.

Nevertheless, the goal of this investigation was to monitor the motor fluctuations attempting to predict the state of the subjects into four classes: ON state, OFF state, asleep, dyskinesia.

They used the laboratory data as a training set for the DL algorithm and the test was on the long-term data in the unconstrained setting.

The result of the investigation was a longitudinal track of subjects during the week of data recording and the overall classification performance was 0.6 in terms of F1 score.

Eskofier et al. [27] in 2016 investigated the feasibility of bradykinesia detection using a Convolutional Neural Network (CNN) and accelerometer sensors placed in the two wrists during standardized tasks.

The study compared traditional expert-defined pipelines with DL techniques. The prediction of the symptom was barely its presence or absence; this aspect may be considered as a limit but the main purpose was the comparison of the two approaches.

The result of CNN was encouraging reaching 90 % of accuracy.

Sama et al. [31] in 2017 assessed bradykinesia severity exploiting the only gait and using a unique triaxial accelerometer placed on the waist.

The assessment and the data collection were in the laboratory, however, the promising result suggested to extend this approach to the unconstrained environment.

Using the recorded data, they extracted the fluidity index highly correlated to the UPDRS scores during walking. Despite the system architecture was extremely simple, the detection of bradykinesia reached 90 % of accuracy thanks to the analysis of gait, which is the best ADL to monitor bradykinesia because it is a repetitive and periodical movement.

Daneault et al. [28] in 2017 investigates the minimum number of accelerometer sensors to estimate accurately bradykinesia during standardized tasks and walking. The data collection was partitioned into three visits, the first two in the laboratory and the last at the patient's home. The total number of sensors was 8, two for each limb; however, the best results for alternating hand movement was in terms of RMSE 0.5 using only one sensor for each upper limb.

The RMSE for the walking was 0.6. The results are expressed as error because they design a regressor model to predict the bradykinesia severity using as true labels the UPDRS scores.

Pulliam et al. [32] in 2017 investigated all the PD motor symptoms, tremor, bradykinesia and dyskinesia, and motor fluctuation as well.

The sensors are accelerometers and gyroscopes placed on the wrists and ankles during standardized and ADL tasks, such as dressing, eating, hygiene, and laundry. The data collection was setup in a simulated apartment studying the progression from the state OFF to the state ON.

The rating scale was the UPDRS to estimate the activities during the recording period. For the severity estimation, a regressor model was used. The results of bradykinesia prediction were expressed as area under the receiver operating curve (AUROC) is 0.82.

Predict bradykinesia and motor fluctuation is not an easy task in the laboratory and nonetheless in unconstrained settings. Even more accurate studies are proposed in literature attempting to reduce the limitations of the previous works or designing more reliable systems for the long-term monitoring of the motor symptoms. Over the years, some concepts have been exploited as the windowing introduced by [18] or the assessment only during ADL [19].

In this work, some of these concepts will be used following the previous investigations; however new ideas will be proposed for the analysis of the data.

In the next chapter, the materials and the method of the work will be explained following the followed processing pipeline.

Chapter 3

Material and Methods

This chapter describes the processing pipeline for the gathering and the analysis of the data aimed at the accomplishment set by the Blue Sky project.

The Blue Sky project has the objective to monitor continuously the motor symptom severity and the motor fluctuations in PD subjects in natural environments such as the patient's home, by using data collected by wearable sensors and ML algorithms. The derived prediction will be useful for the clinicians to personalized the medication strategy attempting to decrease the side-effects of the medication and the severity of the PD symptoms and reducing the lack of objectivity during the assessment.

The project begins with a data collection in two settings, laboratory, and apartment. The data gathering took place at the *Spaulding Rehabilitation Hospital*, Boston, MA, United States, involving clinicians and engineers.

The data collection is justified by the fact that amount of wearable sensors data recorded in laboratory or during clinical visits is significant; on the contrary data collected in a natural environment, such as an apartment, are not so typical.

After the data collection, the data have been processed by using the software MATLAB, Mathworks, Natick, MA, United States; in particular, the signal processing and the ML algorithms toolboxes. The analysis aims to build a predictive model for bradykinesia and motor fluctuations to assess the symptom severity in the apartment.

3.1 Protocol and Data collection

The participants involved in the Blue Sky project are 25; the subjects' age range is from 42 to 80 and the age average and standard deviation are 66.1 ± 8.2 .

Each subject in the study has a diagnosis of PD conforming to the UK Parkinson's Disease Society Brain Bank Criteria and must be able to perceive the wearing off states and attest an improvement of the condition after the medication intake.

The exclusion criteria are the presence of other neurological disease and an implanted medical device, for example, the Deep Brain Stimulation (DBS).

The data are gathered using IMUs called OPAL, designed by APDM, Inc. which embed accelerometer, gyroscope, and magnetometer. The sensor positions are on the two ankle and the two wrists of the subject. This layout will be the same both for the laboratory and the apartment data collections.

The protocol consists of two different visits:

- **Visit 1:** the environment is the laboratory; the subject has to perform tasks belonging to the ADL category (table 3.1). During the visit, a clinician has to evaluate the severity of the PD symptoms at each task according to the UPDRS score; in particular, the score is given according to the worst movement during each task. The visit is long enough to allow the subject to experience OFF and ON states, hence motor fluctuations too.

Due to the length of the entire examination, it is divided into 5 sessions. During the main sessions (1, 3 and, 5) speech tasks, tasks belonging to MDS-UPDRS part III and scripted activity of daily living (SADL) tasks are performed; on the other hand, during the session between the main ones (2 and 4), there are only the MDS-UPDRS part III tasks (figure 3.1).

During the first session, the subject must be in OFF state and at the beginning of the second session, the patient has to take a dose of medication.

The gathered data will be used as the training set for the predictive models. This choice is justified by an accurate measurement in a constrained environment, like the laboratory, under the eye of clinicians.

- **Visit 2:** the setting is a simulated apartment; the subject has to execute SADL and there is no default order of the tasks; hence, he could achieve them randomly. The list of these tasks is in table 3.2

During the visit the patient is alone in the apartment, indeed the entire assessment is videotaped to allow the clinician to evaluate the symptom severity. Besides, every hour the subject has to submit a motor state self-report using a tablet application (figure 3.2).

Finally, the data gathered by the sensors will be used as the test set for the predictive models, to verify if it is possible to assess the bradykinesia severity and to monitor the motor fluctuation in natural environments.

3.1.1 Sensors

The exploited sensors for the study are the OPALs, designed by APDM, Inc. (figure 3.3).

Table 3.1: Tasks included in the protocol of the laboratory data collection. Most of them belong to the scripted activity of daily living (SADL); although there are standard tasks to assess the bradykinesia severity; in this case the alternative hand movement.

Laboratory tasks		
n	Task	Type
1	Button a lab coat	SADL
2	Carry a book (out and back 10m) and place it on a table	SADL
3	Carry a suitcase (out and back 10m) and hold 90s up with forearm at 90°	SADL
4	Eat with a spoon 2x	SADL
5	On the table, folding a piece of paper in half 4x	SADL
6	Open and close a door	SADL
7	Pour a cup of water and take two drinks	SADL
8	Put on and removing jewelry	SADL
9	Use a remote control	SADL
10	Shake 5x, open bottle, drink and close	SADL
11	Tie a shoe	SADL
12	Write <i>elelelele</i> (cursive) 10x	SADL
13	Write a sentence	SADL
14	Zip a zipper	SADL
15	Alternating Hand Movement	UPDRS

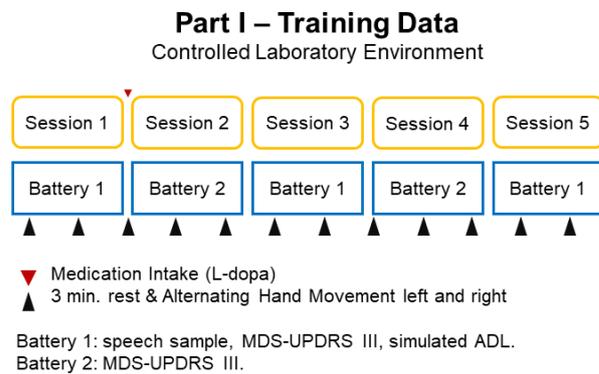


Figure 3.1: Description of the visit 1, including all the 5 sessions. The medication intake is after the first session; every 30 minutes there are 3 minutes of resting and the standardized tasks for bradykinesia are performed. On the right side, a picture of a subject during the laboratory assessment; the sensors exploited for this study are in the two ankles and in the two wrists of the subject.

The sensor is similar to a wristwatch, good to fit around the limbs or waist using strips; the device is an IMU, hence the system embeds a tri-axial accelerometer, a gyroscope, and a magnetometer.

In the Blue Sky project, the OPALs are placed ON the two ankles and the two wrists of the subjects; Besides, two additional OPALs are placed on the back and the sternum of the subject (figure 3.4).

The choice of wearable devices is due to allow to not hamper the movements of the subjects making the analysis closer to the reality, indeed as without the recording devices. Besides, other advantages of the wearable device are the increase spread in

Table 3.2: Tasks included in the protocol of the apartment data collection and all of them belong to the SADL. These tasks are different respect to the laboratory ones and they attempt to mimic a typical day spent at home.

Apartment tasks		
n	Task	Type
1	Walk around every room of the apartment	SADL
2	Put together and taking nuts and bolts	SADL
3	Carry grocery bags, unload them and place the grocery in the fridge	SADL
4	Clean the kitchen using dustpan and broom	SADL
5	Prepare a snack and eating it	SADL
6	Get in and out in the bed	SADL
7	Make the bed	SADL
8	Fold some clothes	SADL
9	Brush teeth	SADL
10	Prepare the meal	SADL
11	Set the table	SADL
12	Load the dishwasher	SADL
13	Clean the table and the kitchen	SADL

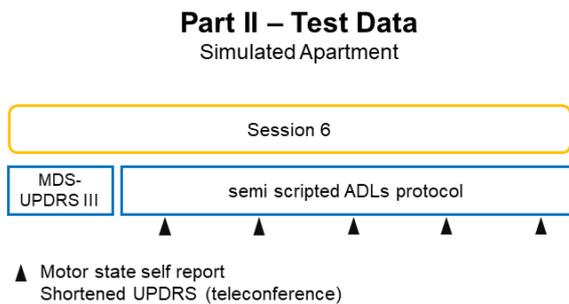


Figure 3.2: Description of the visit 2 in the simulated apartment. The assessment consists on a unique session of 5 hours. The subject has to achieve semi scripted ADL randomly. The monitoring of the subject is by using cameras and the sensors layout is the same of the laboratory examination. On the right side, a picture of a subject while he is closing a shutter and holding a holding a plate.

the market, the low purchase price, the low maintenance cost, and the high measure reliability reached for the monitoring of human activity recognition (HAR) [17].

For the motor fluctuation monitoring, it is decided to exploit only the accelerometer sensors and the signals gathered from the limbs. The main advantage is that the data can be collected continuously and for a long period of time, allowing a longitudinal monitoring of the PD symptoms [14] without frequent recharging times. Therefore, no large batteries are required and the subjects are not forced to charge the sensors daily.

Secondly, the degree of the measurement accuracy is higher compared to the gyroscopes and magnetometers; for example, the magnetometers could record wrong data if near electromagnetic sources.

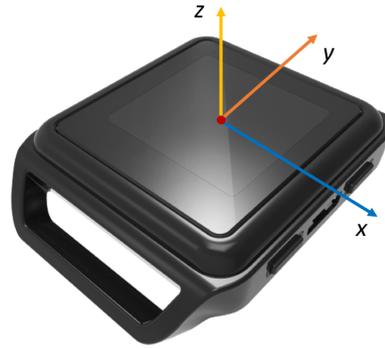


Figure 3.3: Representation of the OPAL sensor. The display helps the user to control the battery state and if the sensor is linked to the network. Inside the OPAL there are an accelerometer sensor, a gyroscope and a magnetometer. Retrieved from [33].

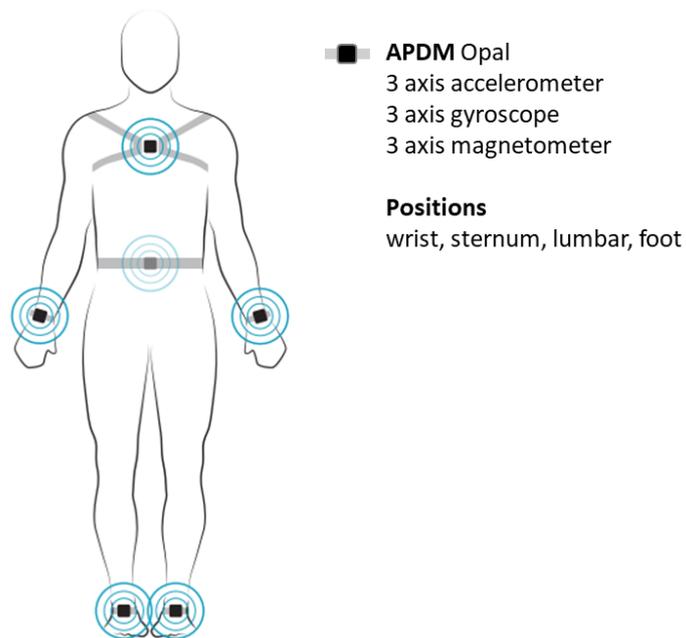


Figure 3.4: Scheme of the sensor positions during the data collection. The OPAL sensors are six and they are placed on the two wrists, the two ankles, the sternum and the lumbar. For the analysis of bradykinesia only the gathered signals from the limbs are used. Retrieved from [33].

In table 3.3 are listed the most important specifications of the OPAL device; for simplicity, only the features of the accelerometer sensor are reported.

3.2 Signal pre-processing

From the continuous raw data acquired by the sensors to the actual data, used for the further analysis, some processing is needed, to handle in a better way the information gathered by the accelerometer signals.

The pre-processing steps in the next sections are the same for both the examinations. They mainly consist of filtering stage and segmentation step. Before that, a

Table 3.3: Technical specifications of the OPAL device. For clarity only the accelerometer specifications have been reported [33]. In the entries with two quantities the first one is referred to the normal accelerometer and the second one to the accelerometer to falls detection.

OPAL specifications	
Specifications	Accelerometers
Axes	3 axes
Range	$\pm 16g, \pm 200g$
Noise	$120\mu g/\sqrt{Hz}, 120mg/\sqrt{Hz}$
Sampling Rate	20Hz to 128Hz
Bandwidth	50Hz
Resolution	14 bits, 17.5 bits
Battery Life	12h (Synchronous Logging) to 16h (Asynchronous Logging)
Internal Storage	8Gb

resampling of the signals is performed, reducing the sampling frequency from 128Hz to 32Hz since the upper-frequency band of accelerometer signals during ADL is below 10Hz [34, 35]. Another reason to justify the resampling is to reduce the amount of information in terms of bytes for the storage of the data.

3.2.1 Filtering stage

The filtering step is one of the critical stages for the detection of bradykinesia, exploiting only accelerometer signals. The frequency band for the analysis depends on the type of movements performed by the subjects.

Patel et al. used a frequency band between 1 Hz to 3 Hz, but the recorded tasks were standardized according to the MDS-UPDRS [18]. Other authors, who exploited ADL tasks for the detection of bradykinesia, used a frequency band between 0.5 Hz to 3 Hz [36].

The differences of band relies on the distinct tasks among the standardized and ADL because the first one must be performed as fast as possible so the frequency content will be a little bit higher respect to daily activities, which are executed with another speed.

The choice of the frequency band for the project is between 0.5 Hz to 3 Hz following the main works in this field. The lower frequency cutoff is set to remove the direct components (DC) inside the signals, like the gravitational acceleration and the gross body orientations. The higher frequency cutoff, instead, is to filter out the components related to the tremor; this symptom has a frequency content from 4 Hz to 6 Hz [2]. Besides, the choice of this frequency band is justified by the nature of the ADL, which is between 1 Hz to 4 Hz [37, 38].

According to the frequency band under investigation, when a movement is affected by bradykinesia the frequency content in the higher frequency should decrease and shift toward the lowest components.

The filtering stage is the same both for the laboratory signals and the apartment data.

To keep low the filter order, instead of designing a bandpass filter, the filter was

split into a highpass (HP), with the cutoff frequency set at 0.5 Hz and a lowpass (LP) with cutoff equal to 3 Hz.

Since the processing is offline, the type of filter is infinite impulse response (IIR) and zero-phase digital filtering has implemented to avoid the phase distortion effect of the filter.

The HP filter and the LP filter are Chebyshev type I and type II respectively. Since the type II is preferred due to the absence of the ripple in the pass band, the type I can better mitigate the low frequency component despite the ripple in the pass band. For this reason the HP filter is a Chebyshev type I.

In table 3.4 are listed the design characteristics of the two filters and in figure 3.5(a) and 3.5(b) are illustrated the magnitude and the phase of the filters.

Table 3.4: The specification of the filters. The specifications are a trade-off between a good attenuation of the not interesting aspects of the signals and a low order to reduce as much as possible the artifact of the filter transitory.

Filter specifications		
	HP filter	LP filter
Type	Chebyshev type I	Chebyshev type II
Order	9	16
Passband frequency	0.4 Hz	2.8 Hz
Stopband frequency	0.6 Hz	3.2 Hz
Passband attenuation	1 dB	1 dB
Stopband attenuation	60 dB	60 dB

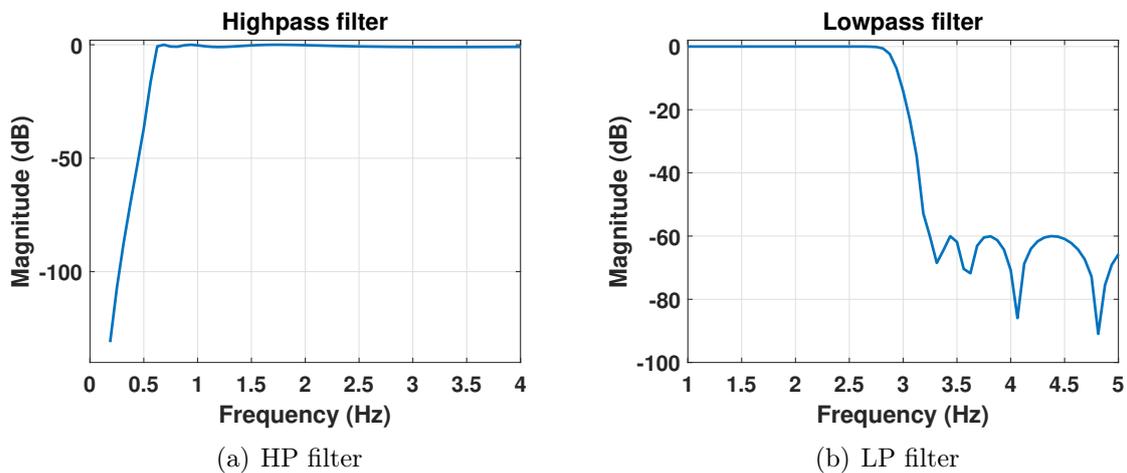


Figure 3.5: On the left side the magnitude of the HP filter is illustrated. On the right side is represented the magnitude of the LP filter. Both the filters comply with the design specifications.

3.2.2 Signal segmentation

The following step, after the filtering, is the segmentation of the filtered continuous signals, gathered in the laboratory, using the temporal markers, recorded to indicate the starting and the ending point of each task.

Afterward the detection of a task, the labeling process is performed; each task has paired with the clinical scores, which evaluate bradykinesia, tremor, and dyskinesia severity. All the signals without a severity score related to bradykinesia are discarded and the bradykinesia severity scores 3 and 4 are merged because the collected sample is too little for the score 4.

Of course, the labeling step has been done only for the gathered data in the laboratory because from this data the training set will be generated.

At the end of these processes, the total number of tasks with a bradykinesia score associated is 2093.

Figure 3.6 illustrates the distribution of the score, for each patient during the laboratory data collection. Besides, the distribution is split into upper limbs and lower limbs sensors. The bradykinesia severity 0 is dominant; also, in the upper limbs the smaller sample is score 1, and in the lower limbs the less depicted class is the 3. Since the samples belonging to class 3 in the lower limbs are only two, these samples have been relabelled as score 2. This choice reflects on the prediction in the apartment data because for the lower limbs the model could predict a severity score between the range 0 to 2.

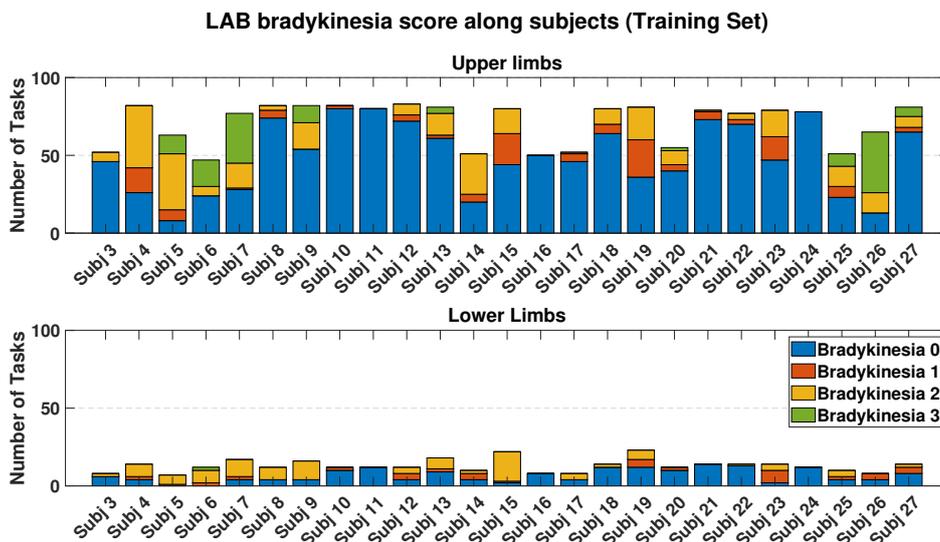


Figure 3.6: Representation of the score distribution for each subject involved in the laboratory data collection. There are two distribution, at the top for the upper limbs and the bottom for the lower limbs. The distribution shows a clear unbalance toward the bradykinesia score 0, low representation of score 1 in the upper limbs, and a low sample size for class 3 in the lower limbs tasks.

3.3 Movement detection

The movement detection is an important step for the prediction of bradykinesia because only during a movement, or an action it is possible to detect the symptom. For the implementation and the analysis are used the laboratory signals; later, the results will be applied to the data gathered in the apartment.

According to the signals gathered during the data collection, a movement is a result of fast acceleration variations due to acceleration and deceleration of the limb to

execute the motion.

Following this physical principle, to detect a movement inside a task, the analysis of the accelerometer signals amplitude and the range is necessary.

To decrease the noise during this analysis, instead of using the filtered data, the envelope of the magnitude is taken into account.

The extraction of the envelope of each task consists of two steps:

1. Compute the magnitude signal from the three channels according to the following formula:

$$\|a\| = \sqrt{a_x^2 + a_y^2 + a_z^2}, \quad (3.1)$$

where a_x , a_y , and a_z are the three channels of the accelerometer sensor.

2. Extract the envelope of the magnitude signal, using an LP filter with a cutoff frequency equal to 0.5 Hz. The type of the filter is Chebyshev II and the order is 5; the choice of the attenuation is 40 dB.

After the extraction of the envelope, a thresholding on the envelope signal is performed to detect movements inside each task. The threshold value is based on a multiple of the noise root mean square (RMS)

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (3.2)$$

of the accelerometer sensor; the information of the noise of the sensor is in table 3.3 and retrieved in the sensor datasheet [33]. To optimize this phase of the process different multiplication factor of the noise level has been attempted; in particular, the attempts are from 15 to 50 in steps of 5. Another optimization around the previous optimal value is executed in steps of 1.

The results are compared with the recorded labels during the data collection, which provide the information if inside a task a movement is present or not.

Since the optimization phase is on the laboratory signals, to detect the movement in the apartment data the best threshold value will be used according to the results on the laboratory data.

In the apartment setting, since the signals are continuous and not divided into tasks, the envelope will be extracted in the same way as described before and the threshold based on the RMS value of the sensor noise is chosen according to the results of the optimization.

3.4 Windowing

The next step of the processing pipeline, after the movement detection, is the windowing.

The movement detection could be considered as a data cleaning because the resting periods are not useful for the study of bradykinesia.

In this phase, all the detected movements will be divided into fixed temporal windows; these windows will be the basic element for further analysis in this work.

According to several published works on bradykinesia severity analysis using IMU signals during ADL, the most used window's length is 5s [36]. This period is used also to estimate the bradykinesia severity during MSD-UPDRS standardized tasks. To get a greater number of windows, a 50% overlap among consecutive windows is applied. This strategy has been widely used in many works both for standardized tasks and ADL [18, 19, 36].

At the end of this procedure, the basic elements of the dataset is a 5s window, where a movement has been detected.

The windowing is applied to the laboratory and the apartment data; for what concern the laboratory data, after the windowing, each window is associated with the label given to the task that the window belongs.

The window labeling could be regarded as a limit of the analysis because there is no availability of the severity score for each movement performed inside a task. This limit introduces a mislabeling problem that could be handled using data cleaning techniques, to polish the laboratory dataset and get a good training set.

3.5 Feature extraction

The feature extraction is done on the 5s windows. At the end, each 5s windows will be represented by a feature array, which will create the feature matrix.

The feature extraction process is the same for the signal windows belonging to the laboratory and the apartment. The computed features derive from the accelerometer signals, but also the velocity, the position, and the jerk signals will be exploited in this phase. The first two quantities are respectively the integration and the double integration of the accelerometer signals. On the other hand, the jerk is the derivative of the accelerometer signals.

Since the accelerometer data consists of a triplet of numbers, matching the three channels; also the signals resulting from the accelerometer signals consist of three components for each time step.

At the beginning, the feature set is chosen according to previous studies in bradykinesia severity prediction [18, 36]; the derived feature set from these works could be categorized into three groups:

- time domain features;
- frequency domain features;
- segment velocity features.

Besides, other features are added according to other works in motor fluctuations analysis and HAR [13, 39, 40, 41], attempting to recognize the bradykinesia severity more accurately.

Finally, the last feature does not come from the sensor signals but it is a piece of information gathered by the clinician or by the subject. This feature is called time since medication intake and measures the temporal distance from the time of the medication intake and the current window.

During the laboratory data collection, this information is gathered by the clinicians, instead, during the apartment assessment, this data is collected according to the videotapes or the subject diary.

3.5.1 Time domain features

The time-domain features are statistical features computed on the temporal signals. All the time domain features exploited are listed with the corresponding signals.

- **Arithmetic mean:** it is defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (3.3)$$

where x_i is the single element of the array x and n is the length for the array. This feature is computed for $\|a\|$, $\|v\|$, $\|x\|$, and $\|j\|$ that correspond to acceleration magnitude, velocity magnitude, position magnitude, and jerk magnitude.

- **Root mean square (RMS):** it is defined by equation 3.32. This feature is extracted from $\|a\|$, $\|v\|$, $\|x\|$, $\|j\|$, and the three channels of the accelerometer sensor a_x , a_y , and a_z .
- **Standard deviation:** it is defined as

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3.4)$$

The magnitude of the four exploited signals is the input to computed this feature.

- **Range:** it is defined as

$$r = \max(x) - \min(x), \quad (3.5)$$

where \max finds the largest value of x and \min finds the smaller value of x . The range is extracted from $\|a\|$, $\|v\|$, $\|x\|$, $\|j\|$, and the three channels of the accelerometer sensor a_x , a_y , and a_z .

- **Entropy:** it is defined as

$$S = - \sum_{i=1}^n P(x_i) \ln P(x_i), \quad (3.6)$$

where P is the probability mass function of the discrete variable x . The entropy is extracted from $\|a\|$, $\|v\|$, $\|x\|$, $\|j\|$, and the three channels of the accelerometer sensor a_x , a_y , and a_z .

- **Cross-correlation peak and lag:** the cross-correlation function of two random sequences is defined as

$$R_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m} y_n^*, & m \geq 0, \\ R_{yx}^*(-m), & m < 0 \end{cases} \quad (3.7)$$

In this case the two sequences are all the combination among the three channels of the accelerometer sensor; $a_x a_y$, $a_x a_z$, and $a_y a_z$. After the estimation of the cross-correlation the maximum value of R_{xy} and the corresponding lag time value are extracted.

3.5.2 Frequency domain features

The frequency-domain features are features that derive from the frequency representation of the signals. The power spectral density (PSD) is extracted from the accelerometer signals: according to this quantity, some shape and statistical features are considered.

In this project, the PSD is estimated using the periodogram technique, which consists of square magnitude computing of the Discrete-Time Fourier Transform of the signal auto-correlation function (DTFT).

Obviously, for each channel, an estimate of the PSD could be computed. During the explanation of the frequency domain features the term PSD_i , with $i = x, y, z$, will refer to the power spectral density of a single channel; on the other hand, the term PSD will refer to the sum of the three spectral density estimations.

- **Dominant frequency amplitude:** it is defined as

$$PSD_{max} = \max(PSD) \quad (3.8)$$

This quantity is the peak value of the PSD computed by a 5s windows signal.

- **Dominant frequency:** it is defined as the frequency related to the peak value of the PSD.
- **Power sum:** it is defined as

$$p_{tot} = \sum_{i=1}^n PSD_i \quad (3.9)$$

This value is the total power contained in the accelerometer signals in the frequency range 0 Hz to 16 Hz.

- **Standard deviation:** it is defined as equation 3.4; in this case, the input signal is the PSD.
- **Skewness:** it is defined as

$$s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\sigma^3} \quad (3.10)$$

This feature carries the information of the amount of asymmetry of the PSD distribution.

- **Kurtosis:** it is defined as

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\sigma^4} \quad (3.11)$$

This feature is extracted from the PSD and gives the information about the shape of the PSD distribution.

- **Entropy:** it is defined as 3.6; the input signals to extract this feature are PSD_x, PSD_y, PSD_z .
- **Energy around the peak:** it is defined as the energy contained in the peak of the PSD, considering the width of the peak 0.5Hz.

- **Energy ratio:** it is the ratio between the energy associated to the peak of the PSD and the total energy contained in the PSD.
- **Energy ratio without peak:** it is the ratio between the energy associated to the peak of the PSD and the total energy contained in the PSD without the energy included in the peak.
- **Ratio max peak mean:** it is the ratio between the PSD peak value and the arithmetic mean of the PSD.

3.5.3 Segment velocity features

The segment velocity features are extracted from the segment velocity signals. These signals are the magnitude of the windowed accelerometer signals derivative; basically, they are an estimation of the velocity during a 5s window.

These kinds of signals are derived from two accelerometer signals; accelerometer signals bandpass filtered between 3 Hz to 8 Hz and 0.5 Hz to 3 Hz. The extracted velocity from the first ones is called segment velocity and the others are called segment velocity no tremor.

- **Max:** it is defined as the maximum value of the sequence. This feature is computed for the segment velocity and the segment velocity without tremor.
- **Arithmetic mean:** it is defined as equation 3.3; it is extracted from the segment velocity and the segment velocity without tremor.
- **Standard deviation:** it is defined as equation 3.4; it is extracted from the segment velocity and the segment velocity without tremor.
- **Entropy:** it is defined as equation 3.6; it is extracted from the segment velocity and the segment velocity without tremor.
- **Movement percentage:** it is the percentage of the signal, which is above the movement threshold. The movement threshold value is 0.05m/s. This feature is extracted from the segment velocity without tremor.
- **Intense movement percentage:** it is the percentage of the signal, which is above the threshold of intense movement. The threshold movement value is 0.75m/s. This feature is extracted from the segment velocity without tremor.
- **Arithmetic mean when movement:** it is the arithmetic mean (equation 3.3) of the signal above the movement threshold. This feature is computed only for the segment velocity without tremor.

Besides, other four features have been extracted exploiting the first four features in the list. These new features are the respective ratio between segment velocity no tremor and segment velocity.

3.5.4 Additional features

Attempting to depict in a better way the main clinical features of bradykinesia other features are added to the feature set, after an analysis of the literature and a study on the gathered signals.

All the proposed features belong to the category of the time domain features.

These new features will be explained showing their mathematical definition and the choice reason.

- **Signal magnitude area (SMA)** [41]: it is defined as

$$SMA = \sum_{i=1}^n |a_x| + \sum_{i=1}^n |a_y| + \sum_{i=1}^n |a_z| \quad (3.12)$$

This quantity carries the information of the total energy of the signal bringing together the three channels. The reason of its selection relies on the loss of energy experiencing a more severe bradykinesia manifestation. According to this, the quantity should decrease when the severity of the symptom increases. Besides, the SMA shows how intense is a movement during a certain period of time. In this case this factor at the beginning of the equation is omitted because this features is computed in a fixed window's length. This feature is extract only from the accelerometer signals.

- **Auto-correlation range** [13]: it is defined as

$$r_{range} = range(R_{xx}) \quad (3.13)$$

where R_{xx} is the auto-correlation function defined by equation 3.7 and $range$ is the mathematical operation defined by equation 3.5.

This feature gives the information of the modulation of the movement according to the work of Patel et al. [13].

This feature is extract only from the three channels of the accelerometer sensor.

- **Mean absolute deviation (MAD)** [39]: it is defined as

$$MAD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n |x_i - \bar{x}|} \quad (3.14)$$

It is a different dispersion or variability measure respect to the standard deviation. The main difference is due to the absolute value instead of the square to compute the dispersion measure.

The change of the variability along the signal is a sign of increased severity of bradykinesia. This feature is computed from $\|a\|$, $\|v\|$, $\|x\|$, $\|j\|$, and the three channels of the accelerometer sensor a_x , a_y , and a_z .

- **Median absolute deviation (MEAD)**: it is defined as

$$MEAD = median(|x_i - \tilde{x}|) \quad (3.15)$$

where \tilde{x} is the median of the signal. The median is the value that split in two halves, the higher and the lower, the data sample. Similarly to the MAD, this

quantity is a dispersion measure, but is it more robust respect to the MAD because the median is less sensitive to outlier values.

This feature is extracted from $\|a\|$ and the three channels of the accelerometer sensor a_x , a_y , and a_z .

- **Inter-quartile range (IQR)** [40]: it is defined as

$$IQR = Q_3 - Q_1 \quad (3.16)$$

where Q_3 is the third quartile, which is the middle value between the median value and the maximum value in a data sample, and Q_1 is the first quartile, defined as the middle value between the minimum value and the median value in a data sample.

This measure is a spread indicator and it is more robust to outlier respect to the standard deviation.

This feature is computed from $\|a\|$, $\|v\|$, $\|x\|$, $\|j\|$, and the three channels of the accelerometer sensor a_x , a_y , and a_z .

- **Inter-quartile mean (IQM)**: it is defined as

$$IQM = \frac{1}{n} \sum_{i=1}^n x_i, \text{ with } x \in [Q_1, Q_3] \quad (3.17)$$

This value carries the information of central tendency of a data distribution, which shows the typical value of the data sample ignoring the outlier values.

This feature is computed only from $\|a\|$.

- **Mid-hinge**: it is defined as

$$midhinge = \frac{Q_1 + Q_3}{2} \quad (3.18)$$

This measure gives the information about the location of the data distribution or the shift of the distribution.

This feature is computed only from $\|a\|$.

- **Zero crossing rate**: it is defined as

$$zerocrossing = \frac{zX}{t} \quad (3.19)$$

where zX is the number of zero crossing of the signal and t is the observation time of the signal; in this case the observation time corresponds to the window's length.

The zero crossing rate is a measure of fundamental frequency. If the bradykinesia score is severe the zero crossing rate value should be low respect to a bradykinetic movement scored with a less severity.

This feature is computed from the three channels of the accelerometer sensor a_x , a_y , and a_z .

- **Range signal mean**: it is defined as the arithmetic mean (equation 3.3) of the range signal. The range signal is a new signal obtained from the three channels of the accelerometer sensor; this signal is obtained sliding a window

on the accelerometer signal and computing the range (equation 3.5) of this small portion of signal. In this way a new signal is extracted containing the information of the range variation along the 5s window of signal. The length of the sliding window is 0.5s; this choice is due to have a good time resolution to follow the trend of the range over the accelerometer signal.

- **Range signal standard deviation:** it is defined as the standard deviation (equation 3.4) of the range signal.

At the end of this step, the total number of features extracted from the acceleration, velocity, position, and jerk signals is 96. Besides, to the feature set the time since medication intake is added even if it is not possible to extract this information from the gathered signals.

3.6 Laboratory dataset analysis

After the feature extraction, the feature matrix is ready for the analysis, to understand how the instances, 5s windows during movements, are placed in the feature space and to observe how the bradykinesia severities are separated among each other, or whether another information is predominant.

According to the outcome of this analysis, the proper solution will be implemented to maximize the results in the bradykinesia severity prediction.

To visualize the multi-dimensional dataset in a 2D or 3D scatter plot a dimensionality reduction method is necessary to summarize all the information in all the features into two or three dimensions. There are many techniques for dimensionality reduction: the most common are Principal Component Analysis (PCA), Sammon mapping and t- distributed Stochastic Neighbor Embedding (t-SNE). The choice of the analysis is the t-SNE because maintains the relative position of the data points that they have in the non-reduced feature space, retaining the greater part of the original information. This is an important advantage to other techniques like PCA. The main hypothesis of PCA is the information is in the variance of the data; this method project the data points maximizing the variance among the data.

The analysis of the data using t-SNE projections is conducted only for the laboratory dataset.

3.6.1 t-distribution Stochastic Neighbor Embedding

t-distribution Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction algorithm belonging to unsupervised learning.

This method was proposed by van der Maaten and Hinton [42] in 2008 as an improved method of the traditional projection approach called Stochastic Neighbor Embedding (SNE) by Hinton and Roweis [43]. The main goal of this technique is to project high-dimensional data into a low-dimensional space, typically two or three dimensions, retaining most of the original information and preserving the original clustering, the distances among near points, in the reduced space.

The steps of this technique are the following [42, 44]:

1. Compute the similarity matrix of the high-dimensional data points; this matrix is designed computing firstly the conditional probability (equation 3.20)

the joint probability (equation 3.21) of the data points using a Gaussian distribution.

$$p_{i|j} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2/2\sigma^2)} \quad (3.20)$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \quad (3.21)$$

2. Generate random points in the reduced space.
3. Calculate the similarity matrix for the low-dimensional data as for the high-dimensional one, but using a Student t-distribution for the joint probability (equation 3.22).

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \quad (3.22)$$

4. The low-dimensional point positions are updated iteratively, minimizing the Kullback-Leibler (equation 3.23) divergence between the two probability distribution, respectively the Gaussian for the high-dimensional space and the t-distribution for the low-dimensional one.

$$KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (3.23)$$

At the end of the process, equation 3.23 is minimized and the similarity matrix of the low-dimensional points should be similar to the high-dimensional data.

3.6.2 Data visualization

Using the feature matrix extracted by the 5s windows of the gathered accelerometer sensors different projections are computed to observe how the data points are located in the feature space.

The first observation is about the movements that involve the lower and upper limbs; since the type of motion are extremely different between these body area, also the kind of signals should be different. In figure 3.7 the t-SNE projection shows the instances marked according to the sensor position. It is clear that for further analysis different approaches should be implemented for these two body areas.

Secondly, it is interesting to see how the same projection looks like marking the points according to the bradykinesia score. In figure 3.8 it is possible to note that the different classes are cumbersome overlapped among each other. This means the complexity of the problem is high and it could be reduced splitting the main problem into subproblems with a low degree of complexity.

Since the tasks of the protocol are various in terms of range of motion and repeatability, the massive variability in figure 3.8 could be due to these reasons.

In figure 3.9 the projection is marked in accordance with the tasks; it is noted, although very distinct, that the tasks are grouped into clusters. Under this information in figure 3.10 is reported the same projection marking the possible clusters. Based on the nature of the movements involved during the data collection in the laboratory the decision was to split the range of tasks into four clusters:

- **Lower limbs:** in this particular case, the only available movement is walking;

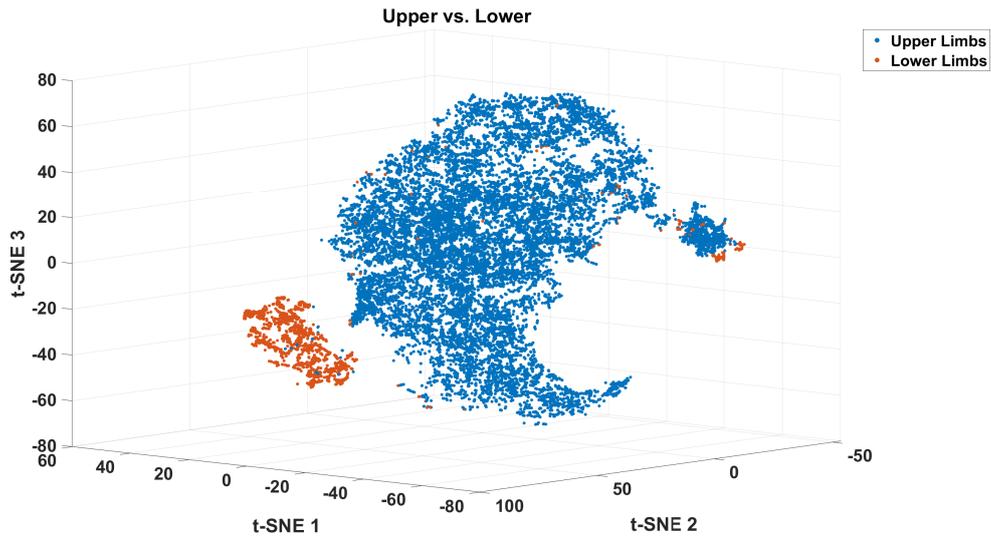


Figure 3.7: Visualization by t-SNE of the laboratory feature matrix reduced from 97 features into three dimensions. The instances are marked based on the sensor positions; the gathered data from the ankle sensors are called lower limbs, instead, the collected data from the wrist sensors are called upper limbs. In accordance with the nature of the movements, the data belonging from the two body areas are well separated, suggesting distinct analysis for them.

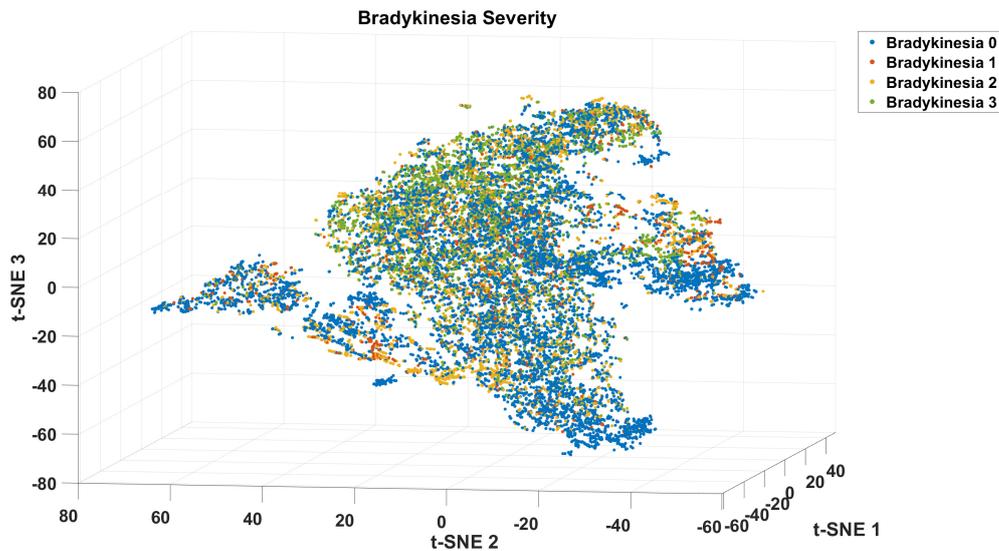


Figure 3.8: t-SNE projection of the laboratory feature matrix marking the data points according to the bradykinesia severities. It is possible to note that the variability among the classes is significant and a clear separation among the different scores is not noticeable.

- **Upper limbs during walking:** this movement is associated with the swing of the arms during walking;
- **Upper fine movements:** these exercises have a low range of motion and the specific tasks involving this movement category are using a remote control, writing a sentence, and writing the word "elelelel";

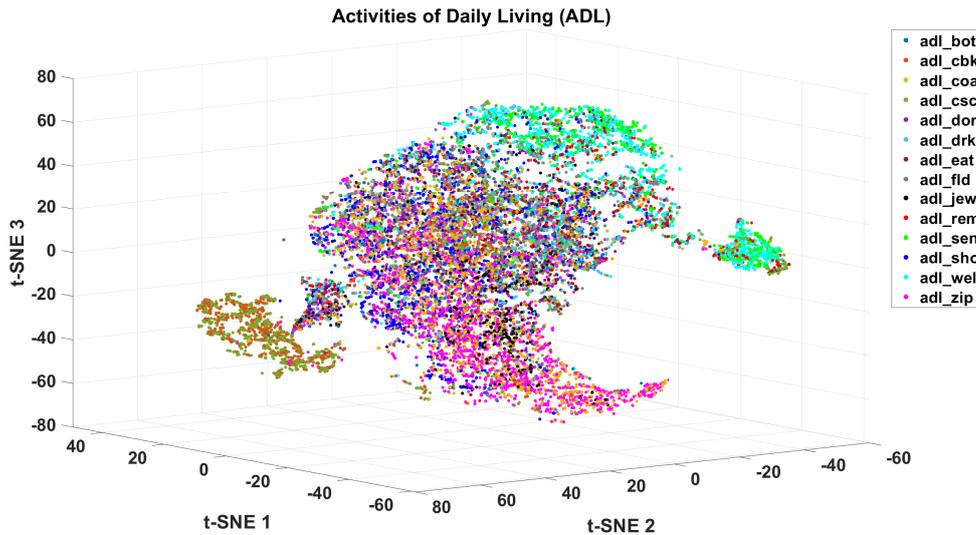


Figure 3.9: Representation of the t-SNE projection of the laboratory feature matrix, where the data points are colored based on the task’s membership. It should be noted there is a pattern according to the task typology; in fact, in the bottom left the instances belong to walking tasks, in the middle, where the density of the point is higher, the samples are associated to gross movement tasks, and at the top of the projection the points are part of activity, such as writing, hence fine movement tasks.

- **Upper gross movements:** this category covers the wide range movements and some examples in the task list are open and close a door, eating, put clothes on and off.

As specified by the projections a good way to analyze the instances is to split the entire laboratory dataset into four subsets; this approach might put aside the variability introduced by the different task nature and have a better focus on the bradykinesia severity. Besides, it is reasonable to consider that inside each subset the movements are invariant respect to the task typology (figure 3.11).

Finally, looking at the figure 3.12 it is noticeable, although the variability carried by the different tasks is removed or partially mitigated, there is still divergence among the severity scores. To get respectable results in the apartment data it is reasonable to clean the data collected in the laboratory to have a neat training set for the learning algorithm.

In agreement with the projections of the laboratory data points, the choice of splitting the dataset into subsets is followed (figure 3.13). In further analysis like the data cleaning and the bradykinesia severity estimation, each subset is considered separately.

This setting will be used for the apartment analysis too using different trained models for the different typology of samples collected in the natural environment.

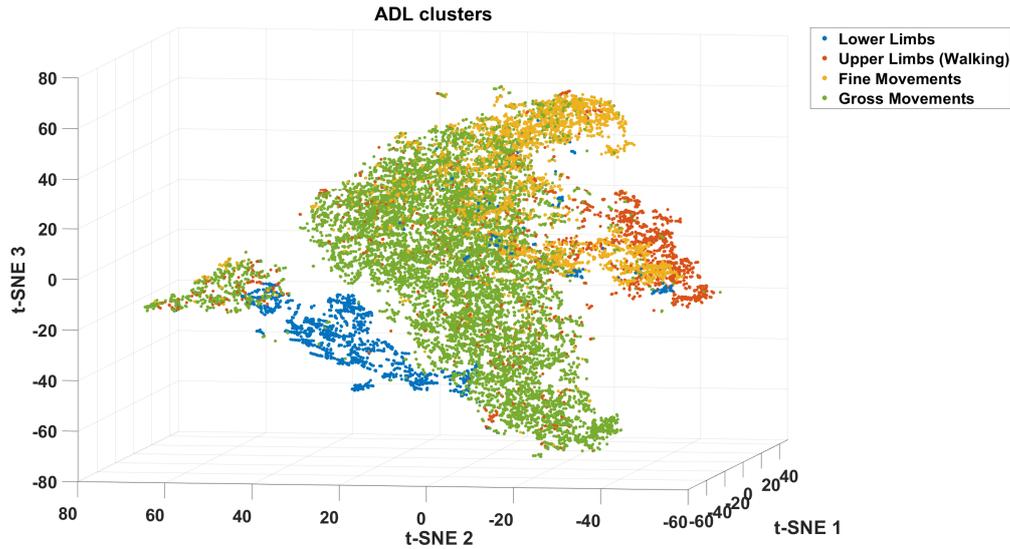


Figure 3.10: t-SNE projection of the laboratory data points, where each point is marked based on the membership to the task cluster. The clusters have been determined according to the main characteristics of the tasks which constitute the data collection protocol. The clearness of the projection means that the manual clustering is acceptable. This data pattern suggests splitting the dataset into four subsets to reduce the complexity of the problem.

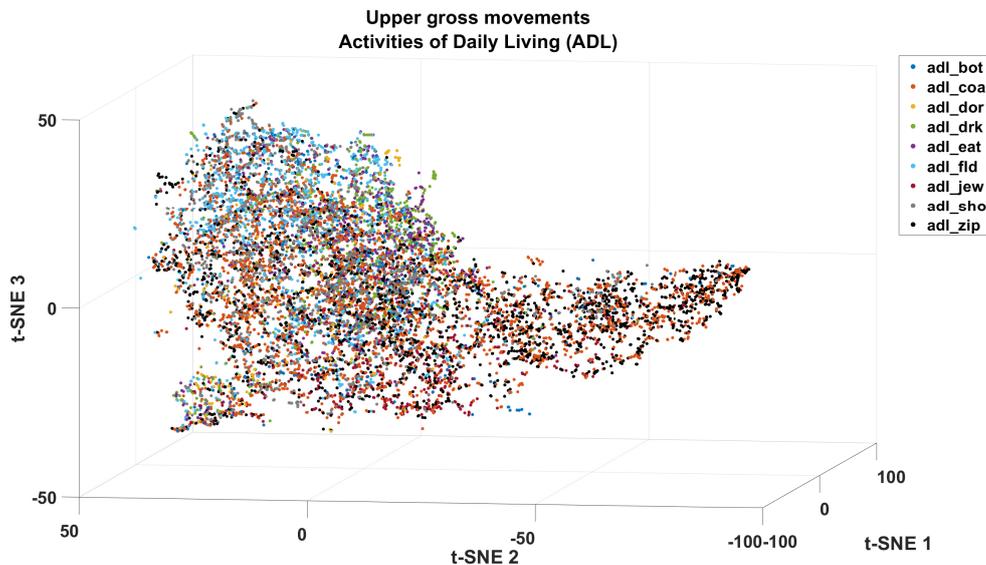


Figure 3.11: t-SNE projection of the points belonging to the gross movement cluster. The instances are marked according to the task typology. As it is possible to notice, the movement inside the tasks could be regarded as invariant respect to the category of the task, because there is no significant distinction in the proposed projection.

3.7 Laboratory dataset cleaning

Following the extracted information by the t-SNE projections, even if partitioning the entire laboratory dataset, there is a significant variability among the bradykinesia severity in each subset.

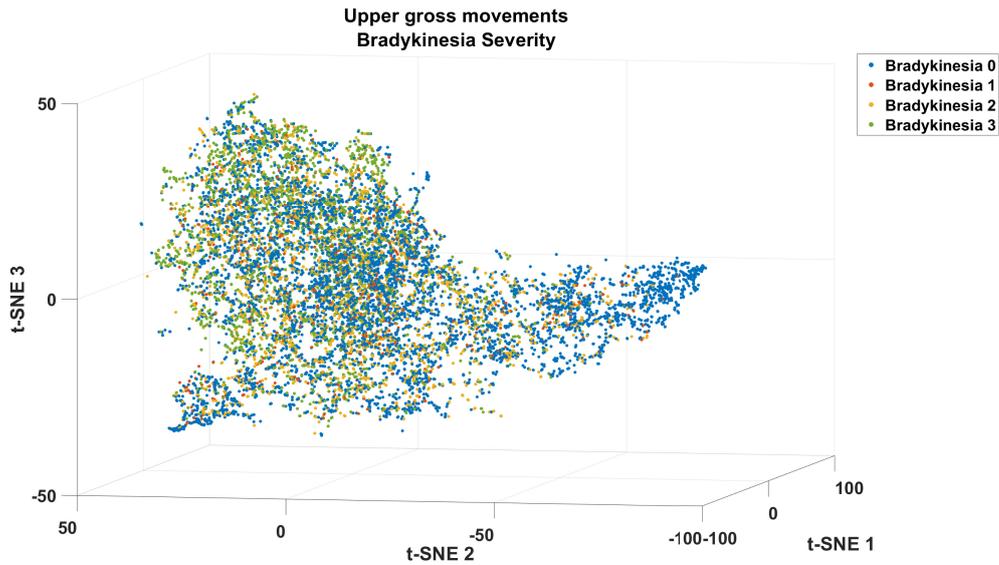


Figure 3.12: t-SNE projection of the points belonging to the gross movement cluster. The data points are labeled following the bradykinesia severity score. As it is clear, there is still overlap among the classes and a cleaning step is necessary to build a robust training set.

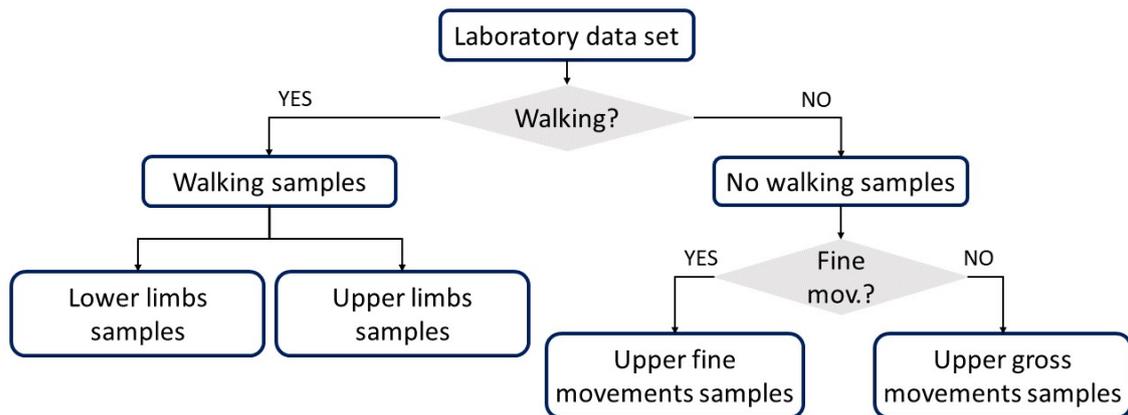


Figure 3.13: Laboratory dataset partitioning driven by the information collected analyzing the projections. The walking movements are analyzed in the lower and upper limbs, instead, the other movements involving the upper limbs are divided into fine and gross motions.

This behavior is the result of labeling errors and how the labels match with the 5s windows.

Assess bradykinesia severity is not an easy task during standardize movements and besides, along with ADL motions. Some errors have occurred during labeling because the scores are given according to the worst severity over the entire task. This means if the task was long enough a chance of severity could have happened causing a mislabeling error, pairing the labels with the windows, because all the movements inside the tasks have been scored as the worst severity score experienced by the subject.

This reason justifies the intention to clean the laboratory dataset. The cleaning is partitioned into two different phases:

1. **Redundant feature removal:** high correlated features are one of the main issues in ML problem because the information carried by these variables is just repeated;
2. **Data cleaning:** removing data points in accordance with their location in the feature space and the membership class using only the feature selected after the irrelevant features removal.

The data cleaning is performed for each dataset partitions adjusting the parameters for every partitions.

3.7.1 Redundant feature removal

The main reason to remove high correlated features, or redundant features, is to decrease the chances of overfitting during the training of the learning algorithm, feeding it with noise data points. Besides, redundant features could lead to low prediction accuracy.

To recognize and then discard redundant features an approach that merges feature importance and correlation among variables is proposed.

In accordance with this method, the final outcome is a feature subset with a high capability to predict the bradykinesia severity and a low correlation among each other features.

The proposed method consists of the following steps:

1. Compute the feature importance using the ReliefF algorithm and rank the variables according to its outcome. In this case the algorithm computes the quality of the features using as target values the bradykinesia severity scores;
2. Calculate the linear correlation value among each pair of features using Pearson's linear correlation coefficient [44] (equation 3.24);

$$\rho(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{j=1}^n (y_i - \bar{y})^2}} \quad (3.24)$$

3. Check, for each pair of features, if the correlation coefficient is higher than the fixed threshold; if this control is true the less important feature will be discarded.

At the end of this procedure, a feature subset with relevant and low correlated features is generated. The size of this subset is not fixed because depends on the correlation threshold value.

The parameter k of the ReliefF algorithm is set to 20 to have enough data points to estimate the feature importance of the predictors.

The correlation threshold value is 0.95 because the aim is to remove only redundant features.

Since this method to remove redundant features is proposed in this work in the result chapter a validation using other feature selection methods is presented to prove that this method is comparable to other already implemented approaches.

ReliefF algorithm

ReliefF is a supervised learning algorithm to estimate the relevancy of the features in a feature matrix. It is defined as supervised because the estimation is according to the labels of the data points, which form the rows of the feature matrix.

This method proposed by Kononenko et al., in 1997 [45] is an improvement of the method called Relief suggested by Kira and Rendel [46] in 1992.

Using all the instances in the input dataset to ReliefF, the algorithm looking at the neighbor of each point establishes if a feature is important or not following this criterion:

"The algorithm penalizes the predictors that give different values to neighbors of the same class and rewards predictors that give different values to neighbors of different classes." [44]

The only parameter to set before running the ReliefF is the size of the neighborhood or the number of nearest neighbors.

Based on [45] increasing the number of nearest neighbors the reliability of the importance estimation rises.

The pseudocode of ReliefF algorithm is the following [47]:

1. *Input*: training set $D = \left\{ (x^{(i)}, y^{(i)}) \right\}_{i=1}^m$;
2. *Output*: the vector W of feature importance estimates;
3. set all weights $W[A] := 0$;
4. for $i := 1$ to m do begin
 - (a) randomly select an instance R_i ;
 - (b) find k nearest hits H_j ;
 - (c) for each class $C \neq \text{class}(R_i)$ do
 - i. from class C find k nearest misses $M_i(C)$;
5. for $A := 1$ to a do

$$(a) \quad W[A] := W[A] - \sum_{j=1}^k \frac{\text{diff}(A, R_i, H_j)}{m \cdot k} + \sum_{C \neq \text{class}(R_i)} \frac{\frac{P(C)}{1-P(\text{class}(R_i))} \sum_{j=1}^k \text{diff}(A, R_i, M_j(C))}{m \cdot k};$$

3.7.2 Data cleaning

The main reason to delete noisy data points is to not compromise the ability to design a proper predictive model.

The applied method in this work is a revised approach used by Lee et al. [48] in 2015 solving the same problem about the mislabeling data points due to the windowing of the task.

Lee used a clustering technique called Expectation-Maximization (EM) clustering algorithm to recognize the data point correctly labeled. In this work, the procedure is quite similar except for the clustering algorithm; in this case, instead of using EM the clustering algorithm employed is k-means. This choice is because the main aim is to remove points according to the place in the feature space and the k-means uses the distances in the feature space to perform the clustering.

The outcome of the k-means is compared to the labels of the instances; the retained points are those that are clustered in the corresponding cluster to the right bradykinesia severity or in the adjacent clusters. Based on this approach of data cleaning, the points surrounded by many instances of a far class are removed (figure 3.14).

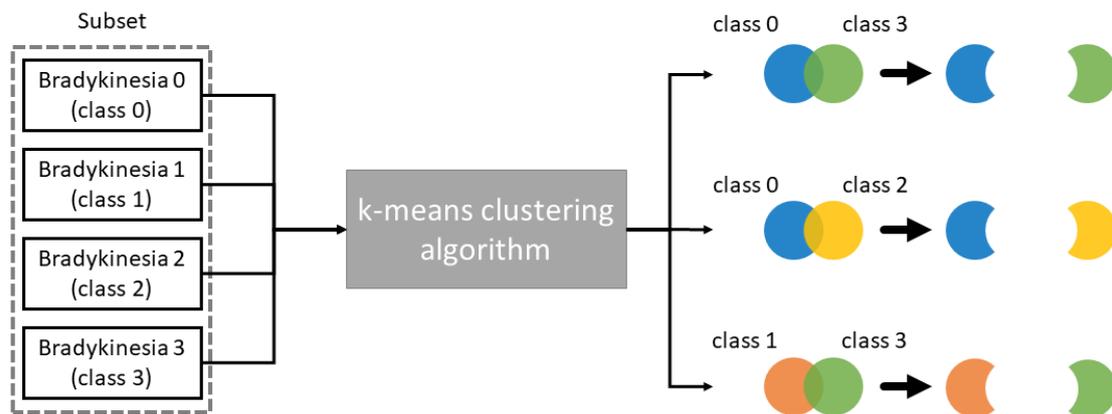


Figure 3.14: Scheme of the approach to remove noisy data points. The subset of data is processed using the k-means clustering algorithm. The clustering outcome is exploited to remove points that are between non-consecutive classes.

The final result is the decreasing of the overlap among the bradykinesia scores, keeping a partial overlap between adjacent classes.

Since the k-means algorithm needs initialization to start the clustering process, usually it is random, in this case the initialization corresponds to the centroids of the bradykinesia severities. In addition, the choice of k is based on the number of bradykinesia classes in each subset.

K-means clustering algorithm

K-means is an unsupervised learning algorithm, which partitions a dataset into k clusters according to the distance between the instances and the cluster centroids. A centroid is described as a representative point of a cluster; usually, the centroid is the arithmetic mean of all the points that belong to the cluster.

The algorithm is iterative and converges to the final solution step by step assigning the points to the proper cluster. K-means algorithm needs two initialization to run:

- The number of clusters k ; usually this value is defined by the user.
- The centroids used for the first iteration; the algorithm picks randomly k points from the input data points or the centroids could be another input for the algorithm forced by the user.

The steps of the k-means algorithm are the following [49, 44]:

1. Select k centroids randomly from the input instances;
2. Compute the distance between each point and each centroid;
3. Assign each instance to the cluster with the nearest centroid;
4. Update the value of the centroids;
5. Iterate from 2. to 4. till a stop condition is reached.

The outcome of this algorithm is k clusters, where the intra-cluster variability, or inner spread, is minimized and the inter-cluster variability is maximized. Ideally, the final clusters should have a low spread and far from each other.

3.8 Data Balancing

The last phase, before the training of the learning algorithm and the evaluation of the performance in the two settings, is the balancing of the cleaned subsets to reduce the bias towards the majority class during the training of the models.

In accordance with figure 3.6 the dataset of the laboratory is clear imbalanced to the bradykinesia severity 0. This pattern is similar looking at the cleaned subsets too.

To solve this issue there are many techniques to balance a dataset; the most common are based on undersampling the majority class or oversampling the minority classes creating synthetic samples.

Obviously, there are pros and cons for both the approaches because the undersampling allows to use only real data but some information is lost; on the other hand, the oversampling retains all the information of the dataset but generating synthetic instances which do not have a link with the reality.

Analyzing the advantages and the drawbacks of these approaches the choice is to oversample the minority classes to not get rid of important information retained by the majority class.

The used technique is called Adaptive Synthetic (ADASYN) sampling [50]. This method is an improvement of another approach called Synthetic Minority Over-sampling Technique (SMOTE) proposed by [51].

3.8.1 Synthetic Minority Over-sampling Technique

SMOTE [51] oversamples the minority class points generating synthetic instances of k nearest class neighbors [52].

After the detection of the nearest class neighbors, q points are randomly picked; the artificial instances are generated along the segments that link the selected points via linear interpolation.

The number q is decided based on the amount of the desired oversampling.

3.8.2 Adaptive Synthetic sampling

ADASYN [50] is an extension of SMOTE, synthetizing minority points near the boundary between two classes, rather than in the interior of the minority class as SMOTE does.

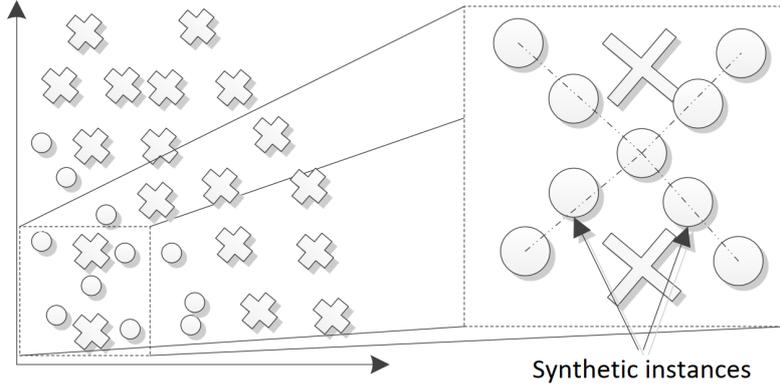


Figure 3.15: Example of SMOTE applied in a dataset, where the crosses are the majority class and the dots are the minority class. The artificial instances lie on the lines that link the real data points of the same class. Retrieved from [52].

The basic idea of this technique is not to decide apriori the number of artificial samples to generate but set this parameter according to the density distribution of the k nearest neighbor of each minority instance.

This strategy ensures to increase the samples in the proximity of the boundaries and enhance the ability to learn from those samples [50].

The steps of the ADASYN algorithm are the following [50]:

1. Input D with m samples; each row of D is $\{x_i, y_i\}$ and m_s is the number of minority instances and m_l is the size of the majority class.
2. Compute the degree of class imbalance as

$$d = \frac{m_s}{m_l}, \quad (3.25)$$

where $d \in (0, 1]$.

3. if $d < d_{th}$, where d_{th} is the threshold for the maximum degree of class imbalance ratio:

- (a) Compute the amount of synthetic points needed (G):

$$G = (m_l - m_s)\beta \quad (3.26)$$

where $\beta \in [0, 1]$ is factor to set the level of balance at the end of the process.

- (b) $\forall x_i \in \text{minorityclass}$, find k nearest neighbor and calculate r_i defined as:

$$r_i = \frac{\Delta_i}{k} \text{ with } i = 1, \dots, m_s, \quad (3.27)$$

where Δ_i are the number of $x_i \in \text{majorityclass}$, so $r_i \in [0, 1]$.

- (c) Normalize r_i as

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_i} \quad (3.28)$$

where now \hat{r}_i is a density distribution.

- (d) Compute the number of artificial (g_i) instances to be generated for each $x_i \in \text{minorityclass}$:

$$g_i = \hat{r}_i G \quad (3.29)$$

- (e) $\forall x_i \in \text{minorityclass}$ generate g_i synthetic data points (s_i):

$$s_i = x_i + \lambda(x_{z_i} - x_i), \quad (3.30)$$

where x_{z_i} is a random data sample that belongs to the x_i k nearest neighbor and $\lambda \in [0, 1]$ is a random number.

3.9 Learning Algorithm

Predict bradykinesia is not an easy task, hence a robust learning algorithm is necessary. A proper learning algorithm should not overfit and should generalize while new data points are processed.

Besides, in contrast to the most works in literature, which usually consider the bradykinesia severity prediction as a classification task, in this work a different approach is applied. In this work, the bradykinesia severity is handled as a continuous variable between 0 to 3 despite the gathered clinical labels are discrete.

This idea, about considering the problem as a regression task, relies on the notion that discrete bradykinesia scores can not catch the little shades that people with PD experience.

The MDS-UPDRS score is discrete to help the clinician during the assessment; moreover, in some cases, the physicians have troubles for the decisions also with only 5 categories of severity.

According to all these considerations, the choice of the regressor learning algorithm is the Random Forest (RF), which has a good resilience to overfitting, a few numbers of tunable parameters and it does not require data normalization.

The tuning of the parameters, numbers of trees and minimum leaf size, is done using the out-of-bag (OOB) error which is a sort of validation of the model exploiting the OOB samples.

Using the RF model, the prediction of bradykinesia is performed both for the laboratory data and for the apartment ones. To achieve this task using the laboratory data two cross-validation (CV) approaches are used. These CV methods are the k-fold and the leave-one-subject-out (LOSO).

Regarding the bradykinesia severity estimation using in the apartment setting four RF have been trained, one for each subset of tasks, using the laboratory data.

Finally, to evaluate the performance both in the environments some measures are used. First of all, since the estimation is a regression task the root mean square error (RMSE) and the mean absolute error (MAE) are used to measure the quadratic and the absolute distance between the estimates and the clinical scores.

In addition, some metrics used in classification are adopted, rounding to the nearest integer value the continuous estimates and comparing them with the clinical scores. These measures are the accuracy, the specificity, the sensitivity, and the F1 score.

3.9.1 Random Forest

Random Forest (RF) is an ensemble learning algorithm used for regression and classification tasks. It is defined ensemble because it is made up of independent trees.

A tree is a flowchart composed of a root, the starting point, internal nodes, where a decision is performed, branches, they are the outcome if the nodes and leaves, which are the outcome of the tree. Typically, trees are binary, it means that each decision node can have only two branches. The process of decision, regression or classification is a specific path over the tree flowchart.

Although the trees tend to overfit and to have a low generalization level for unseen instances, they can catch complex structures inside the data samples; for this reason, trees are the basic elements of the RF.

Besides, to improve the prediction accuracy, RF exploits the bagging technique to generate independent trees applying the "Wisdom of Crowds" concept [53], which states:

"the collective knowledge of a diverse and independent body of people typically exceeds the knowledge of any single individual, and can be harnessed by voting." [54]

Bagging attempts to reduce the high variance of the single trees averaging the outcomes of them to have a more robust prediction.

Another point of strength of the RF is that, during the creation of each node of each tree, only a subset of features picked randomly is used to decide the best feature for the split. This approach allows to design not only independent trees, but also uncorrelated ones.

The final advantage of the RF is the number of parameters to tune are few compared to other learning algorithms, and the tune could be very slight getting good performance.

Bagging

To generate independent trees the RF generates a number of subsets equal to the number of weak learners. In this way, each tree has its training set for learning allowing to capture different characteristics from the data.

To generate each subset a random sampling with replacement is performed using all the instances in the training set of the RF. Sampling with replacement means that each example could be picked more than once for the same subset.

Exploiting the bagging every tree has a different perception of the dataset catching different information and following the concept of the Wisdom of Crowds.

Besides, another feature of the bagging method is that for each tree some samples will be not picked; these samples are called out-of-bag (OOB) samples because they are not in the subset of a particular tree.

These special examples are fundamental to validate the trained RF estimating the OOB error because the OOB samples are unseen instances for a specific tree.

Random Forest algorithm

The pseudocode of the regression RF is the following [54]:

1. for $n = 1$ to N , with N as the number of weak learners:
 - (a) Generate the subset S sampling with replacement the dataset D .
 - (b) Grow a random-forest tree T_n to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size l_{min} is reached.
 - i. Select randomly m features from the total number of features M .
 - ii. Pick the best variable/split-point among the m .
 - iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_n\}_1^N$.

After the training phase the prediction of a new instances the following:

$$\hat{f}_{RF} = \frac{1}{N} \sum_{n=1}^N T_n(x) \quad (3.31)$$

3.9.2 Cross-validation approaches

CV is a validation method to exploit efficiently the available data for the testing phase. CV validates the learning method and not the learning model.

The CV test a partition of the dataset using the remaining one as a training set. The two most used CV approaches are k-fold and leave-one-out (LOO).

- **k-fold** consists of partitioning the dataset into k subsets; iteratively one of the k subsets is the test set and the others are the training set (figure 3.16). The value of k could be chosen according to the amount of available data. Increasing the value of k the result is enlarging the training set and decreasing the size of the test set.
- **Leave-one-out (LOO)** is a particular case of the k-fold when the value of k is equal to the number of data points in the dataset. A different version of the LOO is the LOSO cross-validation. This method works leaving as a test set not a single example, but the entire examples that belong to a specific subject. The other subjects' instances are used to train the learning algorithm. The main advantage of the LOSO is the removal of the subject bias because all the instances of the are the test set.

3.9.3 Performance measures

To measure the quality of the estimates during a validation or a testing phase some performance measures were implemented both for regression and classification tasks.

Regression

Since the true output values in regression tasks are continuous the best way to estimate the quality of the model output is to measure the distance between the true values and the predictions and compute the average to get a single value.

The two common measures for regression tasks are:

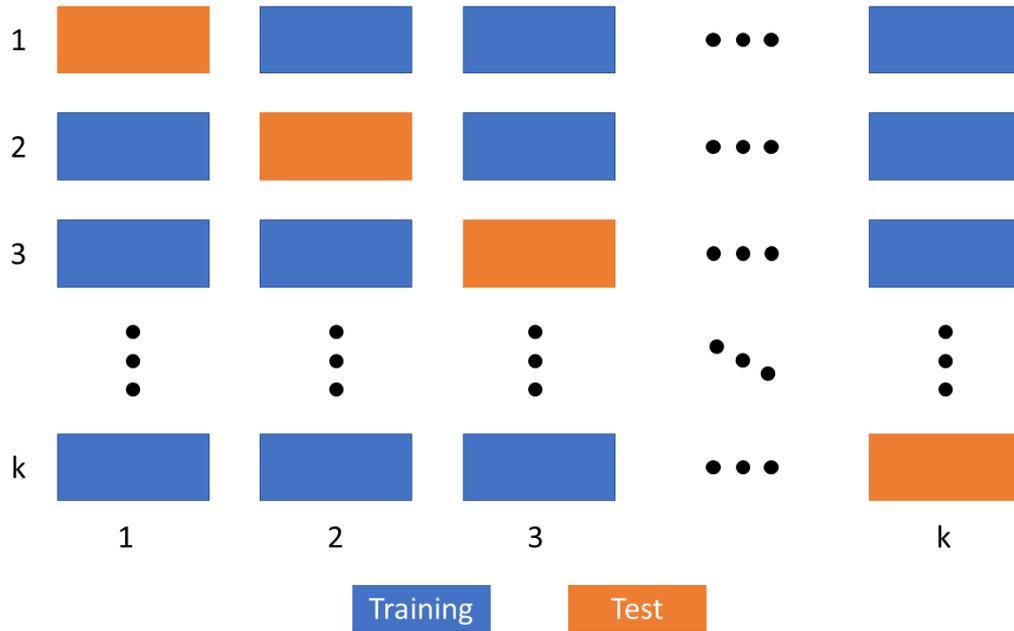


Figure 3.16: k-fold cross-validation concept illustration. Each row represents an iteration of the CV; in blue, there are the subsets labeled as training set and in orange, the subset defined as test set for the i^{th} iteration. At the end of the CV, each subset is tested getting the outcome, that could be compared to the real outputs of the instances.

- **Root mean square error (RMSE):** it is defined as

$$RMSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.32)$$

where y_i is the real value for the i^{th} instance and \hat{y}_i is the correspondent predicted value.

A derived measure from the RMSE is the mean square error (MSE) which is the square of the RMSE.

RMSE penalizes large errors because the distance between them is squared; lower is this value and better is the quality of the estimation.

- **Mean absolute error (MAE):** it is defined as

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.33)$$

This indicator calculates the absolute error between the real values and the prediction without penalizing large errors.

Classification

Since the labels in classification tasks are discrete or categorical, the performance measure relies on the concept of confusion matrix, which express graphically the prediction performance. Moreover using the entries of the confusion matrix some values could be extracted.

The confusion matrix is a square matrix ($n \times n$), where n is the number of classes of the problem. The columns of the confusion matrix corresponds to the predicted values and the rows to the real values.

According to this layout, the principal diagonal of the matrix corresponds to the hits, meaning that the estimation is the same as the true label; outside the principal diagonal there are the misses, meaning that the prediction differs to the true label.

True class	0	True negative (TN)	False positive (FP)
	1	False negative (FN)	True positive (TP)
		0	1
		Predicted class	

Figure 3.17: Confusion matrix representation of a binary problem. Along the rows are the true labels, on the other hand along the columns are the predicted scores.

In binary classification problem, the confusion matrix is 2×2 and each element has a specific name (figure 3.17):

- **True negative (TN):** number of negative cases correctly identified;
- **False positive (FP):** number of negative cases identified as positive;
- **False negative (FN):** number of positive cases identified as negative;
- **True positive (TP):** number of positive cases correctly identified.

This notation could be generalized to multi-classification problems.

Combining this information it is possible to extract some performance measure:

- **Accuracy:** it is defined as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.34)$$

In other words, it is the sum of the principal diagonal divided by the number of the predictions. This is the measure of correct prediction respect to the entire tested samples.

- **Specificity:** it is defined as

$$specificity = \frac{TN}{TN + FP} \quad (3.35)$$

It measures the amount of negative samples that are correctly classified.

- **Sensitivity** or recall: it is defined as

$$sensitivity = \frac{TP}{TP + FN}. \quad (3.36)$$

It is an indicator of the percentage of positive samples correctly recognized by the classifier.

- **Precision**: it is defined as

$$precision = \frac{TP}{TP + FP}. \quad (3.37)$$

It is the proportion of the positive samples recognized as positive against the number of positive prediction of the classifier.

- **F1 score**: it is defined as

$$F1\ score = \frac{2TP}{2TP + FP + FN}. \quad (3.38)$$

It is the harmonic mean between precision and recall.

3.10 Apartment data analysis

The final purpose of this work is to predict the severity of bradykinesia and the motor fluctuations in community settings, like the patient's home.

Following the choices used for the laboratory data analysis, it is necessary to design a processing pipeline for the accelerometer signals gathered in the apartment.

The pipeline reviews the analysis done using the laboratory data; the main difference is in the movement patterns recognition, because in the laboratory there are labels to define them, instead in the apartment this information is not available.

The flowchart of the apartment analysis from the raw data to the final bradykinesia prediction is shown in figure 3.18. The first block is the filtering stage, the movement versus rest detection and the windowing of the movements. The second block is the feature extraction and it is the same as described in section 3.5.

The remaining part of the pipeline is inspired to figure 3.13. Firstly, it is necessary to know if the patient is walking or not. If the subject is walking the partition between upper and lower limbs is according to the sensor location. Based on this information, the sample is processed to the properly trained model, returning the bradykinesia severity prediction.

If the subject is not walking, the movement involves only the upper limbs, but it is necessary to know if the motion pattern is fine or gross. To know that a classifier to recognize fine versus gross movement is implemented. According to the outcome of the classifier, the sample is processed by the correct trained regressor model.

To decide if the subject is walking the assumption is that all the movements in the lower limbs are walking. To recognize the swing of the arms, it is necessary to know the starting and the ending time stamps of the walking movement.

The learning algorithm for this classification task is RF. It is trained using the data gathered in the laboratory and as targets, the labels associated with gross or fine movement.

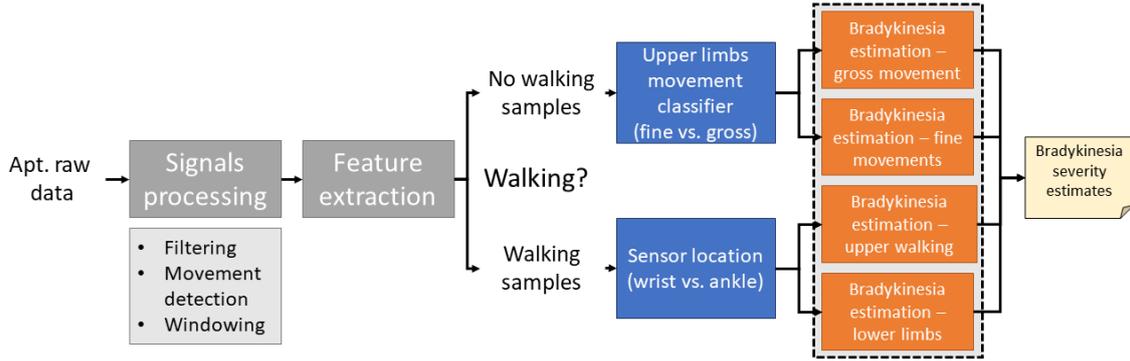


Figure 3.18: Illustration of the processing pipeline for the apartment signals. The starting inputs are the raw accelerometer signals. After the signal processing and the feature extraction the first split partitions the samples during walking and during motions that involve only the upper limbs. The branch of the walking samples is further partitioned according to the sensor position. The other instance branch has a split based on the outcome of the classifier to recognize if the pattern of movement belongs to gross or fine movement. The final part is the bradykinesia severity estimation using one of the four trained regressor RF.

The optimal number of trees is chosen according to the OOB error and removal of the redundant features, as in section 3.7.1 is performed before the training phase. The trained model is validated using the laboratory data using the k-fold and the LOSO cross-validation.

The performances of bradykinesia severity estimates are extracted comparing the estimates with the clinical score of the physician looking at the recorded videotapes of the apartment assessment.

Each clinical score is the overall bradykinesia score over 30s; hence it is necessary to aggregate the outcomes of the model to be able to do a fair comparison.

For the aggregation of the estimate scores, the median is used following the work of Pulliam et. al [32], where the median is used to smooth the outputs to build a clear render of the predicted score trend. Besides, the median operator is robust to the noisy output and the application of this mathematical operation may be considered as a post-processing to reduce noise. The aggregation of the predicted scores works as follow:

1. Detect a clinical score c_i ;
2. Examine 30s window starting from the timestamp t_i of the clinical score c_i ;
3. Consider all the predicted score \hat{c} included in the 30s window and computing the aggregation using the median.

After the aggregation, the estimation errors in terms of RMSE is computed. Moreover, only for a graphical reason, the raw outputs over the entire apartment assessment are smoothed using the median filter with a sliding window of 10s minimize the filter transient because the outputs are not continuous over the entire examination.

3.11 Conclusions

To summarize this chapter all the methods used are to handle the intrinsic complexity in the collected data in the laboratory. After the traditional operation, like filtering, windowing, and feature design and extraction other processes are necessary. The distinction among different movement patterns is the key to the analysis. The idea of the analysis of the laboratory data is replicated for the apartment data, designing different paths for the movement patterns and building four regressor RF to estimate the bradykinesia severity in the distinct cases.

In the next chapter, the results obtained in the different parts of the pathway are shown.

Chapter 4

Results

The most significant results obtained during the analysis are shown following the path of the processing pipeline described in the previous chapter.

For what concerns the results of the bradykinesia severity prediction and the detection of the motor fluctuations, the comparison between two configurations is proposed. The two setups differ by the feature set, indeed the first one includes the medication intake information, on the other hand, the second feature set consists of only variables extracted by the accelerometer signals.

This choice is justified since the medication intake is a strong predictor for the motor fluctuations, but unfortunately, it can not be gathered using wearable sensors, however, it could be recorded using a drug manager application. For this reason, the approach without the medication intake is proposed as well.

4.1 Filtered signals

The filtering stage is the starting point of the analysis of the accelerometer signals. The raw signals are filtered to isolate the useful information and discarding the frequency components related to other PD symptoms.

The signals before and after the filtering are represented in figure 4.1. The raw signals have a strong DC component due to the orientation of the limbs, and it is clear to notice some high frequencies especially when the signals are flat. This manifestation is the tremor at resting, which is very easy to recognize using accelerometer sensors.

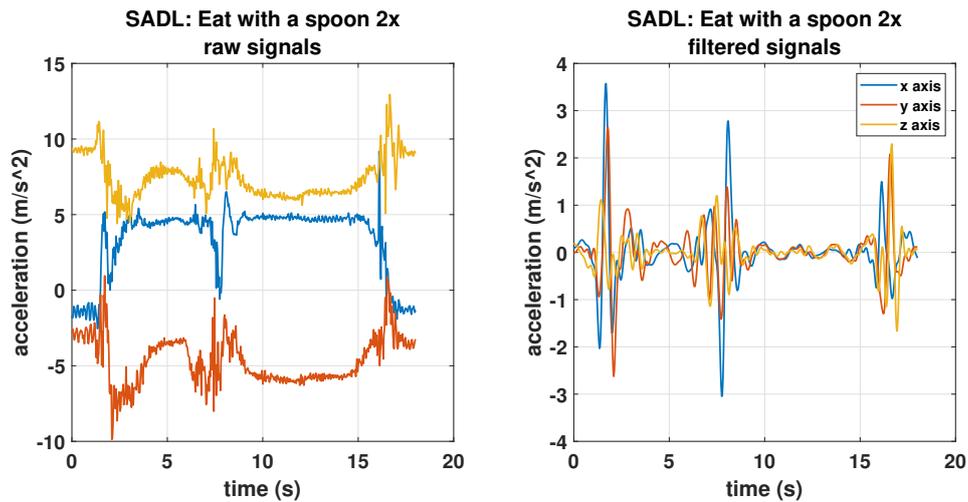


Figure 4.1: Illustration of the raw and filtered signals gathered during the eating with a spoon ADL. On the left, the raw signals appear noisy; the noise derives by the orientation of the limbs, and the presence of tremor movements during the resting periods. On the right, the signals are filtered to reduce the two principal noise sources. In this way, the analysis is focused only on the bradykinesia symptom.

On the other hand, the filtered signals fluctuate around the zero line, and the frequency components related to the tremor are significantly attenuated thanks to the high attenuation of the LP filter.

To check the correctness of the filtering, frequency analysis of the signals is necessary. Estimating the PSD, it is possible to observe the design of the filters is correct or not.

The periodograms of the raw and filtered signals are shown in figure 4.2. The frequency components related to the bradykinesia are retained, and the power carried by the other frequency is almost zero.

4.2 Movement detection validation

For the validation of the motion detector, the entire dataset collected during the Blue Sky project is exploited to get more reliable results.

The validation consists of attempting different levels of threshold applied to the envelope of the acceleration magnitude to obtain the best separation between resting periods and movements.

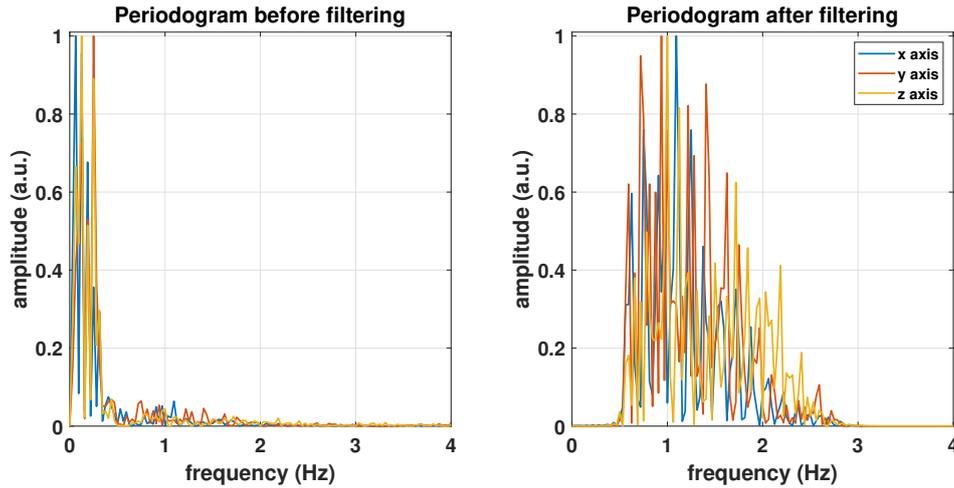


Figure 4.2: Representation of the periodograms estimated from the raw and the filtered signals. On the left, the periodograms of the raw signals are shown. Most of the power is toward the DC frequency component and going to the high frequencies the amount of power decrease dramatically. On the right, the periodograms of the filtered signals show that the designed filters accomplished to attenuate the useless information for the bradykinesia prediction.

Since the threshold is based on the RMS of the sensor noise, which is very low, the starting point for the validation was 15 times the noise level. The first attempt is from 15 to 50 times the noise level by steps of 5; afterward, the grid is thickened between 20 and 25 to find out the best threshold level.

To extract the performance for each threshold level labels collected during the laboratory data collection, which states if in the task there is or not a voluntary movement.

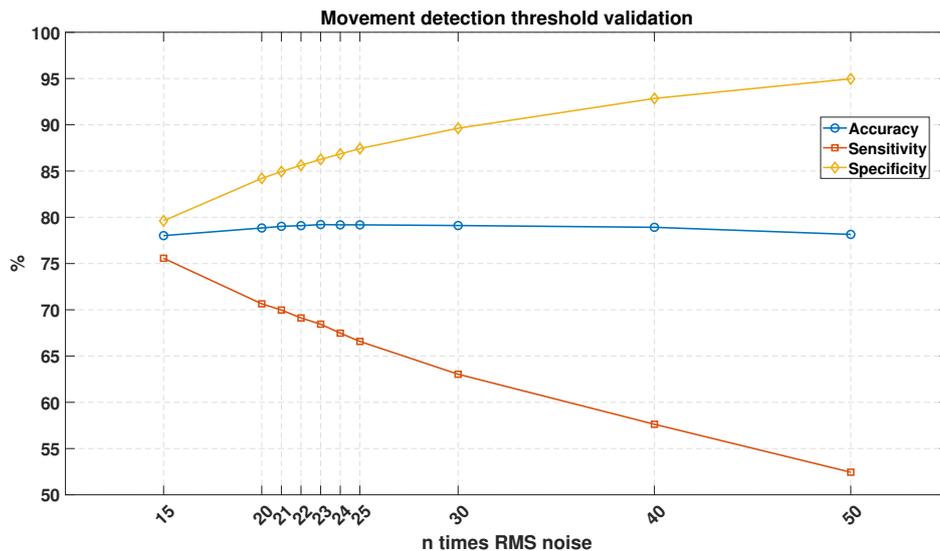


Figure 4.3: Threshold validation of the movement detector. An investigation about the best threshold level to separate the resting periods from the movements is done. An initial grid is explored, once determined the best interval a new denser grid is used to find the best threshold. A trade-off among the performance measure is necessary for the final choice.

The results of the first validation part are shown in figure 4.3. The best results are between 20 and 25. The best results considering a trade-off among accuracy, specificity, and sensitivity corresponds to 22 times the noise level.

This value of threshold is used to detect the movements inside the tasks that constitute the bradykinesia dataset. This threshold is suitable for the removal of the rest periods in the apartment setting as well.

The final result of the movement detection in the laboratory dataset is shown in figure 4.4(b), while the starting phase is in figure 4.4(a)

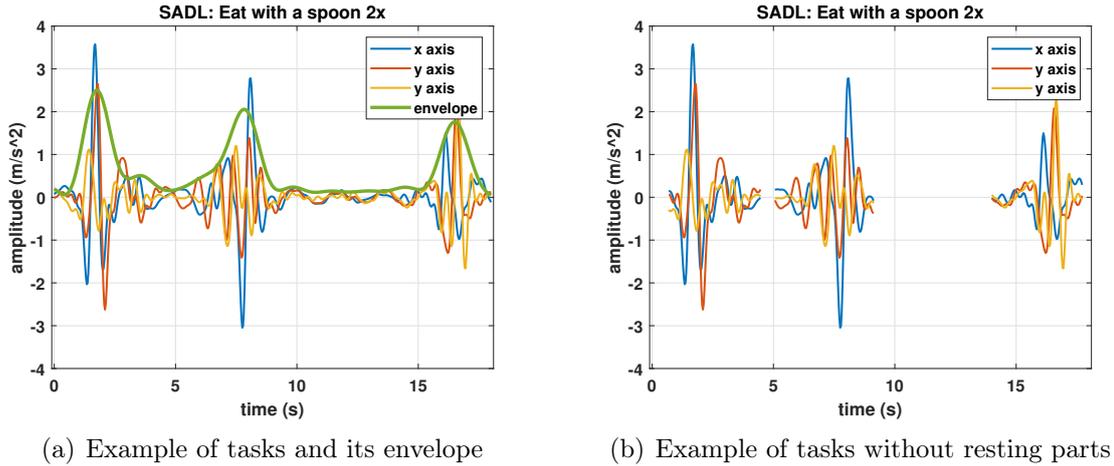


Figure 4.4: An example of signal processing before and after the movement detection is shown. On the left side, the three channels and the extracted envelope are illustrated. On the right side, only the movement segments are retained.

4.3 Feature selection validation

Since the method used for the feature selection during this work, is an approach implemented for this project, it is compared with other feature selection methods implemented in the software used for the project.

The method used for the comparison are Correlation Feature Selection (CFS) [55], ReliefF [45], and minimum Redundancy Maximum Relevance (mRMR) algorithm [56] and the comparison is for each subset of movement patterns.

Since the outcome of the methods is the ranking of the features, to select the right size of feature subset a ranking approach is exploited.

The ranking approach starts evaluating the most important feature according to the feature selection algorithm and computes a quality measure; iteratively all the features are added according to the ranking and the performance measure is computed. The best subset is identified according to the best quality measure. In this case, the quality measure is the RMSE of bradykinesia severity prediction using 10-fold CV and as learning algorithm the RF.

At the end of this procedure for each subset of movement pattern, a feature subset for each feature selection method is retrieved.

The sizes of the feature subset are shown in figure 4.5(a). The different bars are the results of the applied methods. The RMSE for each movement pattern is shown in figure 4.5(b).

The results using the proposed method of feature selection outperform the other methods both for the number selected features and in terms of prediction error.

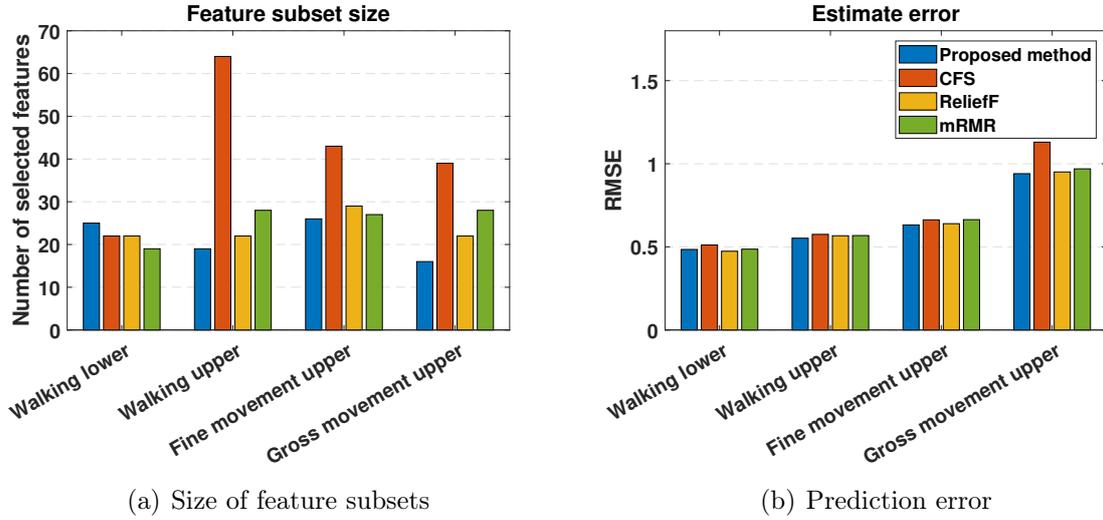


Figure 4.5: Validation results of the proposed feature selection method. On the left side, the number of features selected by the ranking algorithm using the different feature selection approaches are shown. Compared to the other methods, the proposed one selects fewer features. On the right side, the RMSE between the UPDRS score and the predictions using the selected feature subsets is shown. The proposed method based on feature correlation and ReliefF outperforms the other approaches making it a reliable feature selection algorithm.

4.4 Data cleaning results

In this section, the principal results of the cleaning will be explained following the steps described in section 3.7.

In the first part, the feature selection results will be shown, presenting the step-by-step outcomes only for the gross movement pattern; on the other hand, the final result of the redundant feature removal will be shown for every movement pattern. In the second part, the results of the noisy instances will be presented only for the gross movement pattern, but the others have similar results.

4.4.1 Selected features

Since the first step of the feature selection algorithm proposed in this work is the estimation of the feature importance, figure 4.6 represents the bar diagram of the importance for the gross movement cluster of movement patterns. According to the feature importance, most of the proposed features in section 3.5.4 carry out essential information for the bradykinesia prediction.

The second result of the procedure is the correlation matrix presented in figure 4.7. It may be observed that a significant number of features are highly correlated, hence the removal of those features is necessary. The final result of the redundant feature removal is in figure 4.8, where the black spots in the correlation matrix means that the feature is discarded.

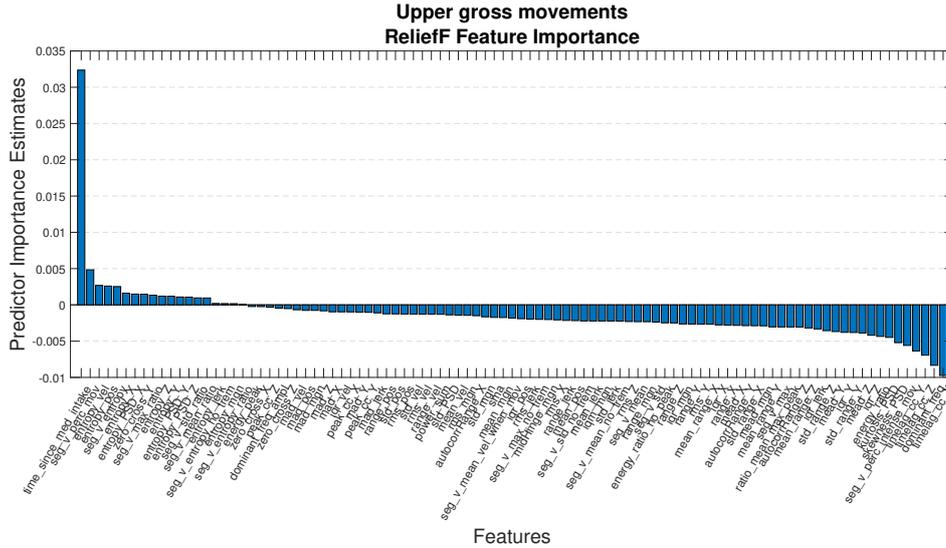


Figure 4.6: Feature importance estimates computed using ReliefF on the gross movement cluster. The outcome of the algorithm shows as the most important features to predict the bradykinesia severity is the medication intake, the segment velocity features, and the entropy. On the other hand, the less important for the bradykinesia estimates are the cross-correlation time lag and the dominant frequency.

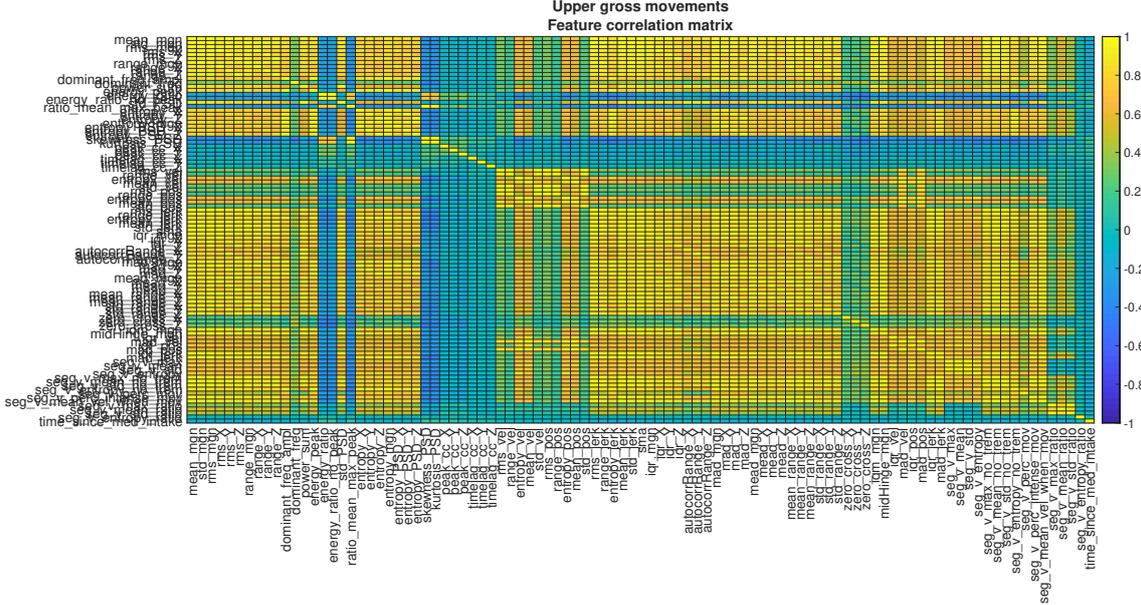


Figure 4.7: Feature correlation matrix of the gross movement cluster. Most features are highly correlated among each other and this justify the removal of the redundant ones. Meanwhile, a small amount of features are very low correlated like the energy ratio, the features derived by the channel cross-correlation, the zero crossing rate and the medication intake.

Besides, a general overview of the results in the different movement patterns is in figure 4.9, where the features are ranked according to their importance. The features selected in all the clusters are invariant respect to the movements, instead, the less selected features depend on the patterns of movement.

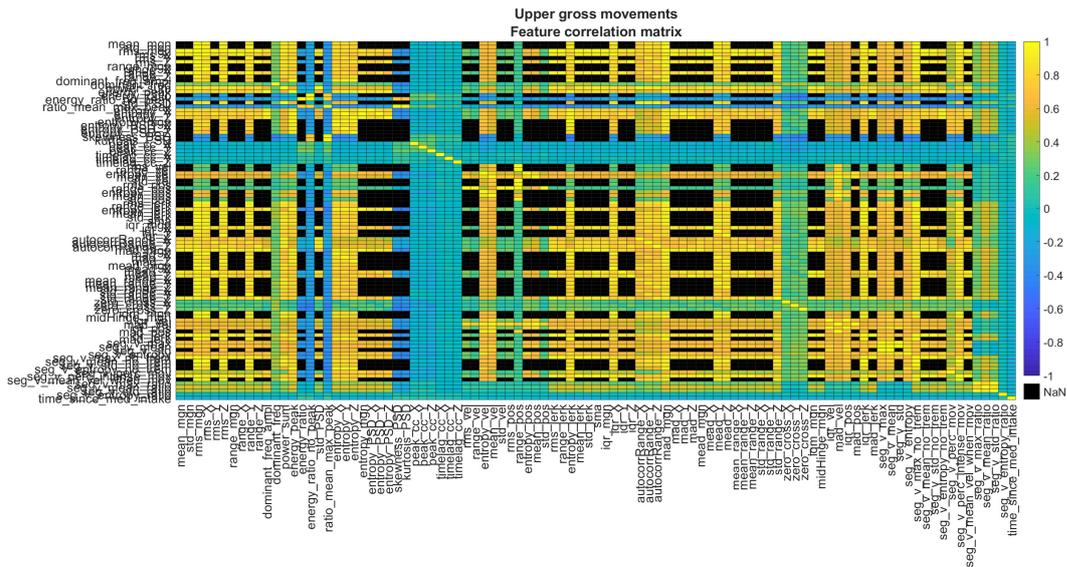


Figure 4.8: Feature correlation matrix of the gross movement pattern after the removal of redundant predictors. The black squares are the removed features; in general, there are still some correlation among the features but the retained features carry out the most information for the bradykinesia estimates.

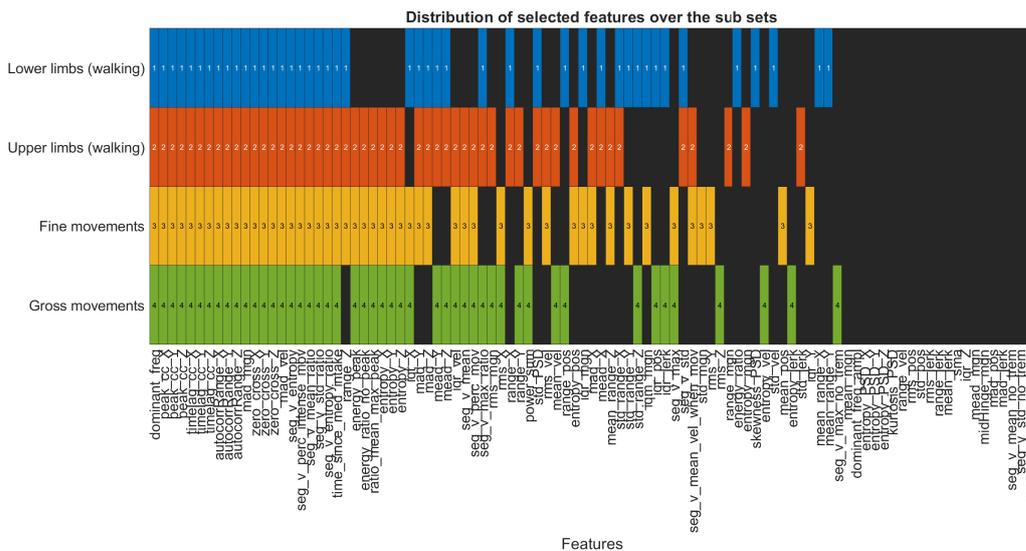


Figure 4.9: Retained features of the different movement patterns. The movement clusters have some features in common, such as the dominant frequency, the auto-correlation range, and the zero crossing, but different other features can detect the bradykinesia only in particular movement patters.

4.4.2 Data cleaning

The results are shown in a qualitative way using t-SNE projections and the numerical results of this section will be described in section 4.7.1.

Figure 4.10 illustrates the projection of the gross movement cluster before the removal of noisy and outlier data points. It may be noticeable that the class separation is not significant.

Then, the result of the k-means clustering algorithm is shown in figure 4.11, where

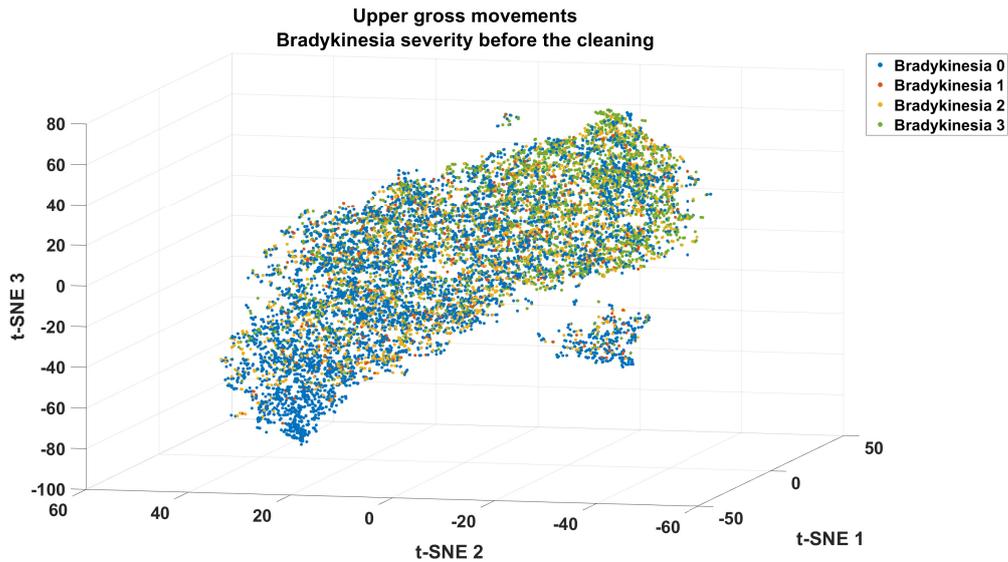


Figure 4.10: t-SNE projection of the gross movement cluster before the cleaning. The overlap among the classes is serious and the variability is still high.

the clusters should correspond to the classes of bradykinesia severity.

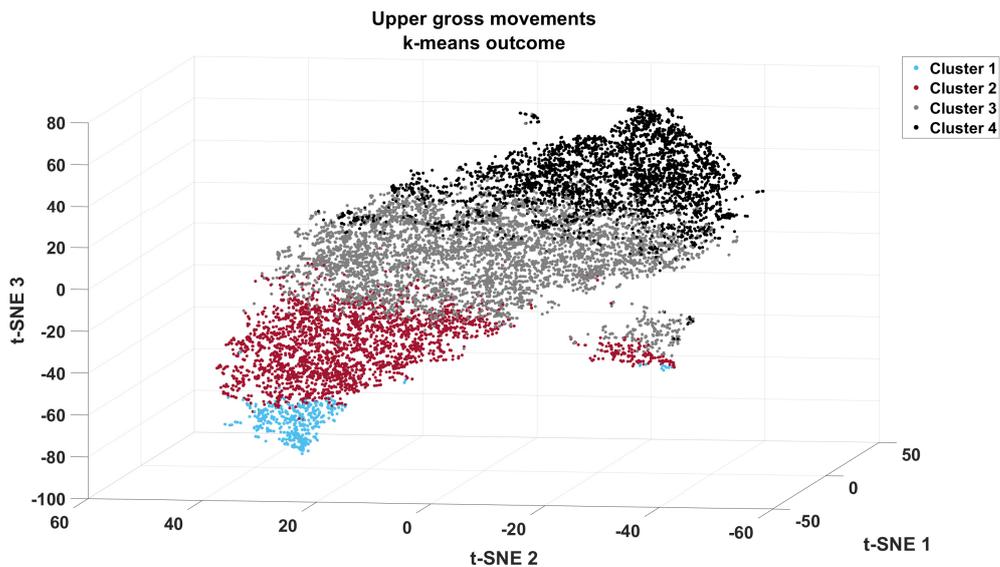


Figure 4.11: k-means clustering outcome of the gross movement pattern.

Finally, in figure 4.12 the projection after the cleaning is illustrated, showing an enhancement of the class separation compared to figure 4.10.

4.5 Data Balancing results

The results of the balancing using the ADASYN algorithm are represented in terms of class size before and after the oversampling of the minority classes. This result is shown for each cluster of movement patterns in figure 4.13.

The balancing is not perfect because the choice is to have a trade-off among the classes since generate a significant amount of synthetic data could be lead to errors

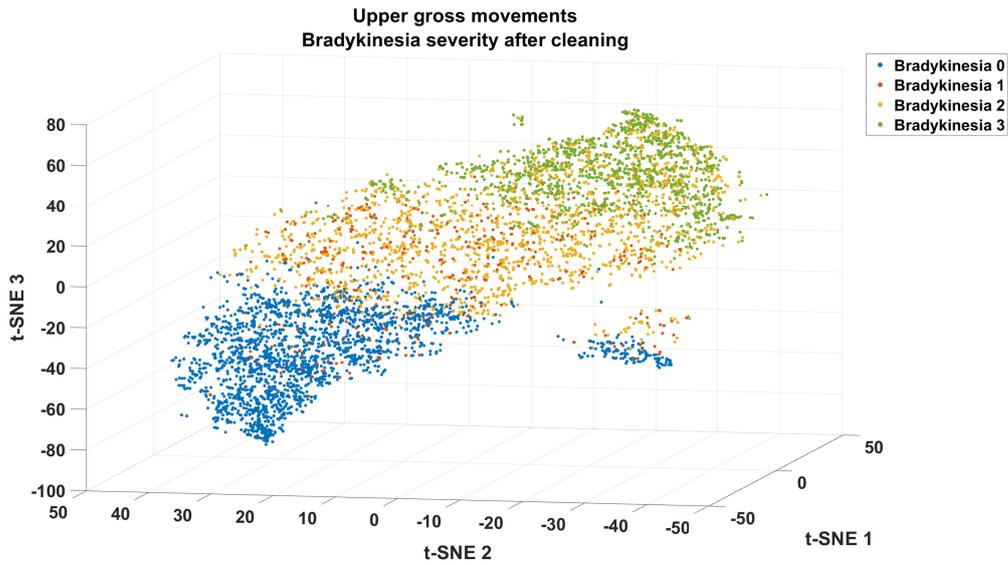


Figure 4.12: Example of the outcome after the data cleaning miming the procedure proposed by Lee et al. [48]. The projection of the gross movement cluster after the cleaning step. A clear distinction of the classes and at the same time a decreasing of the variance are noticeable.

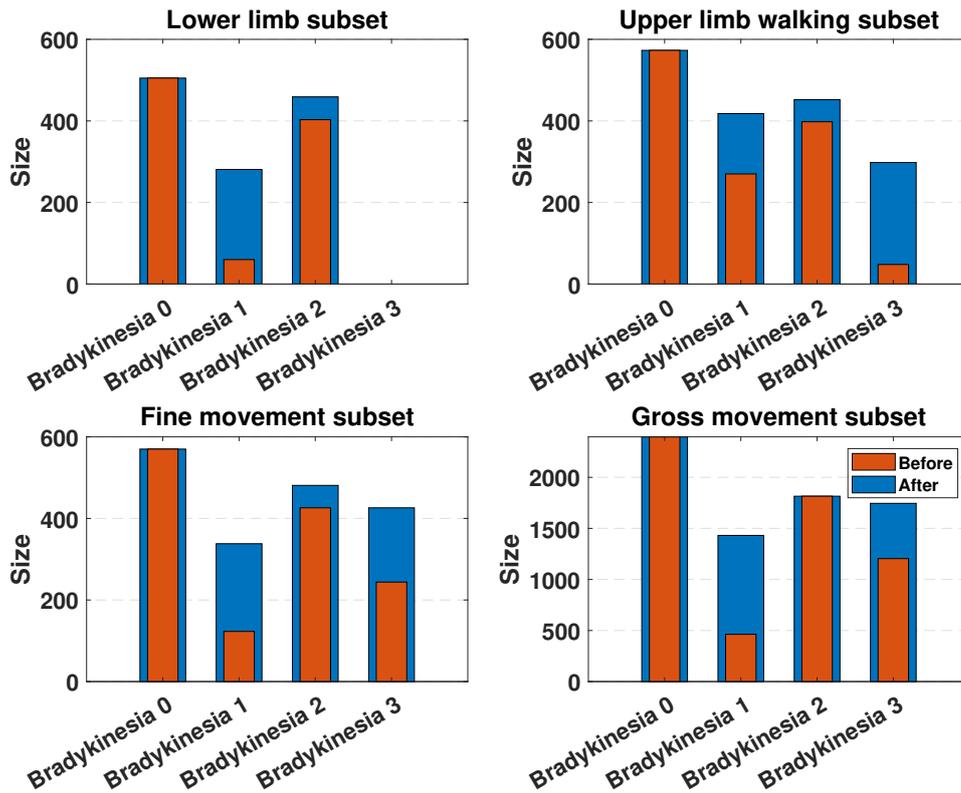


Figure 4.13: Illustration of the data balancing comparing the distribution of the classes before and after the application of the ADASYN algorithm. The severity 0 is not oversampled in every cluster, instead the class 1 is the minority one and the oversampling is significant. The severity 2 is slightly oversampled and the class 3 is oversampled severely only in the movement pattern upper limb walking.

during the training phase.

An example of how the synthetic instances are placed in the reduced feature space is in figure 4.14; the circles show the synthetic points that are generated near the boundaries of the classes.

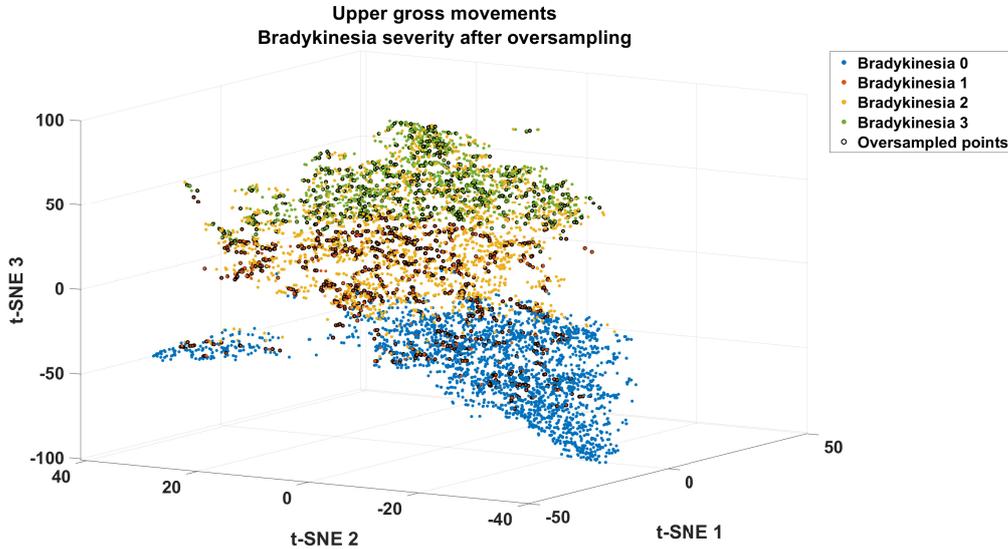


Figure 4.14: Projection of the gross movements cluster after the class balancing using the ADASYN algorithm. The black circles highlight the synthetic instances generated by the oversampling method; the artificial points are mainly places near the class boundaries increasing the chances to separate more the bradykinesia severities.

4.6 Learning algorithm parameters

An important phase of the training is the choice of the learning algorithm parameters for each movement cluster pattern. The two parameters to tune are the number of weak learners and the minimum leaf size. For the latter, the choice is to use the default value of the algorithm equal to 5 because the aim is to minimize the deepness of the single tree to avoid overfitting.

To decide the optimal number of trees, the OOB samples and the OOB error are used during the training. For each cluster, the OOB error curve is computed using a large number of trees, to see at which number of trees the model stops to learn or to reduce in a significant way the validation error.

Table 4.1: Parameters used for the training of the RF algorithm.

Random Forest parameters				
	Lower limbs	Upper limbs (walking)	Fine movements	Gross movements
Trees	20	25	25	25
Leaf size	5	5	5	5

In table 4.1 the parameters of each model are shown; besides, in figure 4.15 are represented the OOB error curves and marked the number of trees chosen.

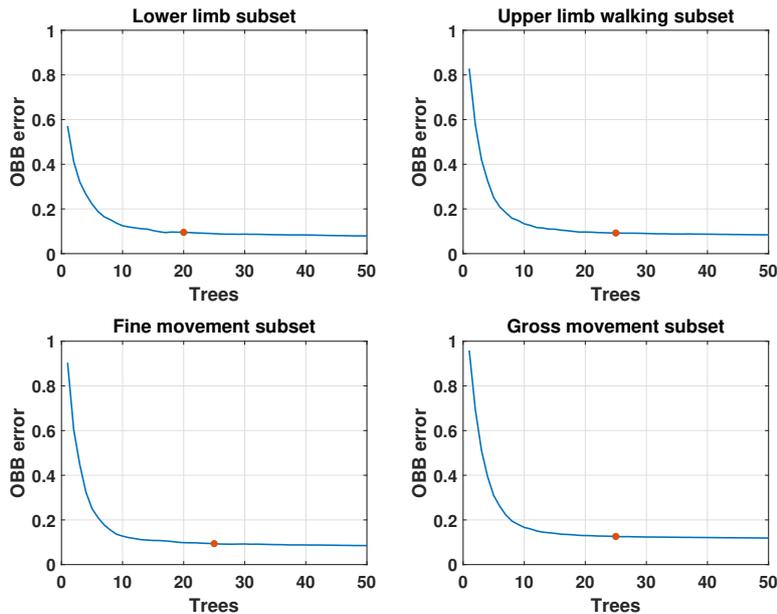


Figure 4.15: OOB error curves for each movement pattern. The maximum number of tree is 50, but the curves become flat after 20 or 25 trees. When the curves are flat means that the learning algorithms stop to acquire information from the training data points. The red dots on the OOB curves detect the optimal number of trees for each pattern.

4.7 Laboratory bradykinesia prediction

The section concerning the results of the laboratory data is partitioned into two parts; the first is about the results of the bradykinesia severity before and after the data cleaning, while the second part regards the validation result of the bradykinesia severity for each movement patterns including or not in the feature set the medication intake variable.

In both cases, two cross-validation methods are used, the k-fold with k equal to 10 and the LOSO.

4.7.1 Bradykinesia prediction after cleaning

The following results are including the medication intake in the feature set. In figure 4.16 it may be noticeable as the regression error decreased after the cleaning step. This is justified by the removal of noisy points and the resultant increasing of the class separation allowing the learning algorithm to learn from only reliable instances.

4.7.2 Bradykinesia prediction

The result will be shown for each movement pattern, in terms of confusion matrix and performance metrics. Since the task is regression, the proper measures should be RMSE and MAE, but rounding the continuous outcomes to the nearest integer value, it is possible to generate the confusion matrix and compute the typical classification performance measures.

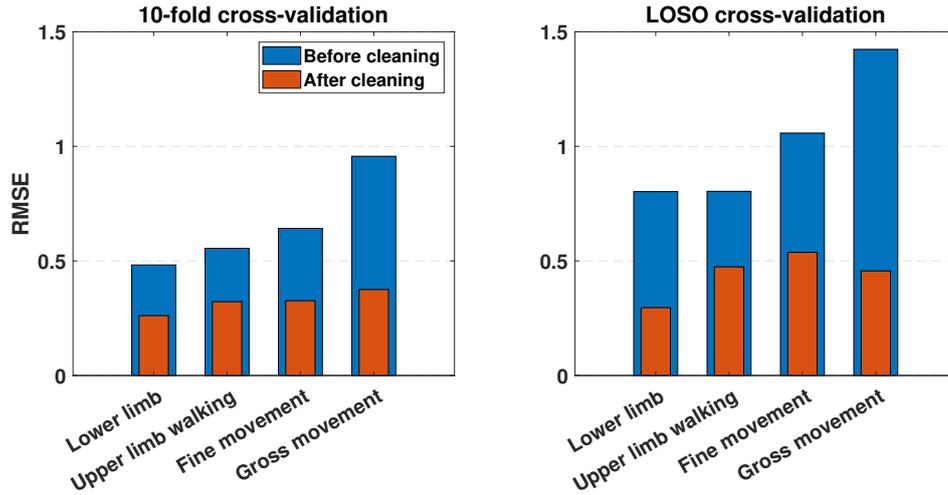


Figure 4.16: Bradykinesia severity prediction results in terms of RMSE before and after the data cleaning. In both the cross-validation methods the overall trend is a decrease of the error. The most significant improvement is for the gross movement subset, indeed the RMSE drops from 1 to 0.4 and from 1.4 to 0.45 using the k-fold and the LOSO respectively.

For the next sections, the results will be shown as follows: a comparison between the results with and without the medication intake for each cross-validation method and a table that sums up the performance in each case.

Lower limbs

In figure 4.17 the confusion matrices of the bradykinesia severity prediction in the lower limbs using the 10-fold CV are shown. The comparison is between the usage or not of the medication intake feature. The same comparison using the LOSO CV is proposed in figure 4.18.

The overall performance is listed in table 4.2 for the four cases. According to the performance measures, there is no significant discrepancy between the two feature sets suggested.

Table 4.2: Lower limbs performance with and without medication intake predictor using 10-fold and LOSO cross-validation.

Lower limbs				
	10-fold		LOSO	
	Medication	No medication	Medication	No medication
RMSE	0.26	0.27	0.30	0.30
MAE	0.10	0.11	0.12	0.12
Accuracy (%)	93.5	94.0	92.4	92.9
Sensitivity (%)	94.8	95.3	94.1	94.4
Specificity (%)	97.6	97.7	97.1	97.1
F1 score (%)	86.1	87.4	85.0	86.0

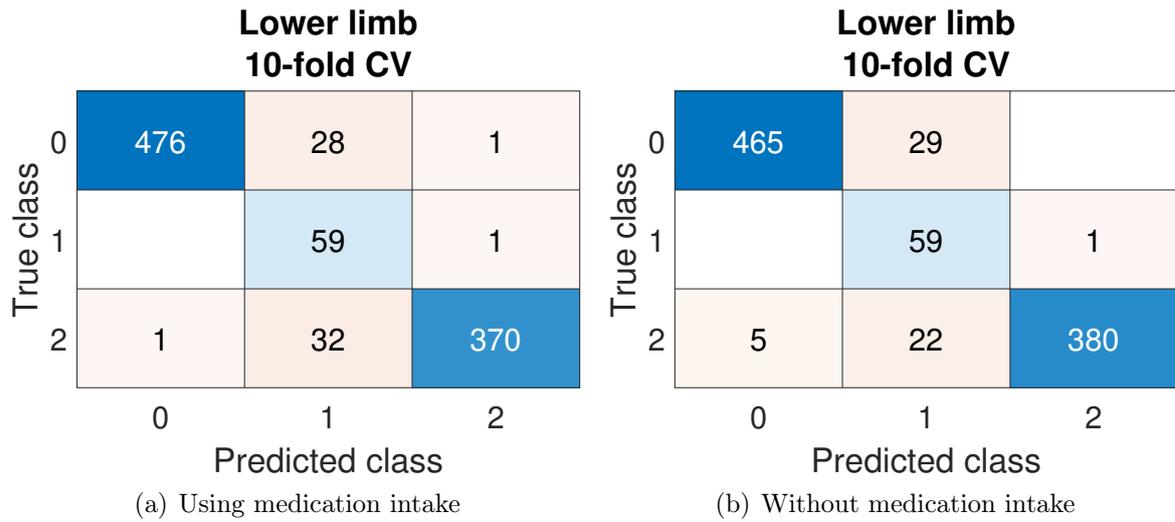


Figure 4.17: Confusion matrices of the lower limbs subset using the 10-fold CV. On the left side the result regards the usage of the medication intake, instead on the right side the results are without this feature.

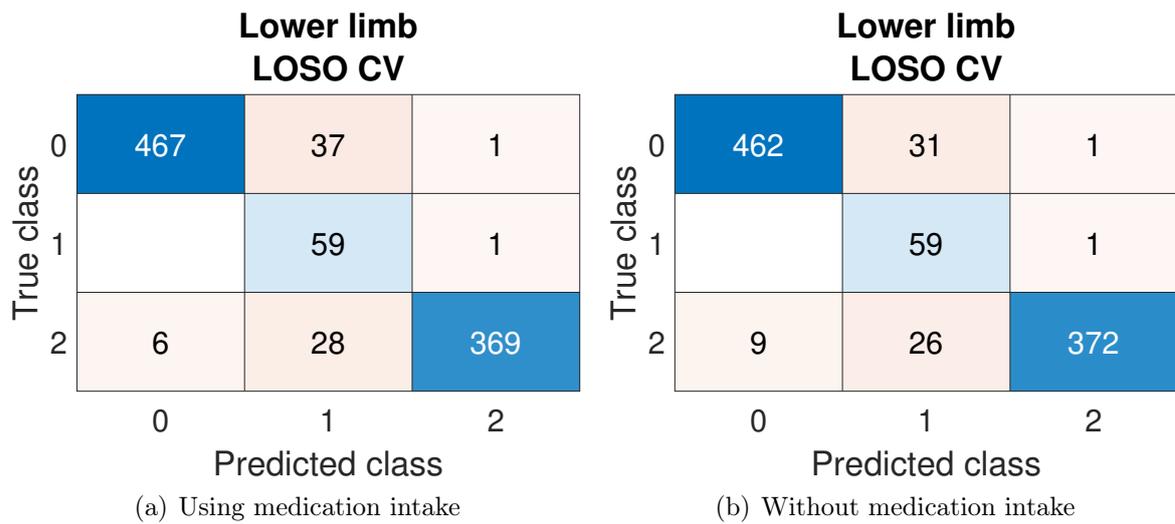


Figure 4.18: Confusion matrices of the lower limbs subset using the LOSO CV. The results including or not the medication intake are proposed on the left and right side respectively.

Upper limbs during walking

The confusion matrices of the bradykinesia severity prediction using the upper limbs during walking are shown in figure 4.19 applying the 10-fold CV; meanwhile, the results using the LOSO CV are illustrated in figure 4.20. The comparison between the two feature sets is proposed highlighting as the inclusion of the medication intake predictor leads to a slight improvement in the bradykinesia severity prediction. In table 4.3 the prediction error and the other performance measures are summed up.

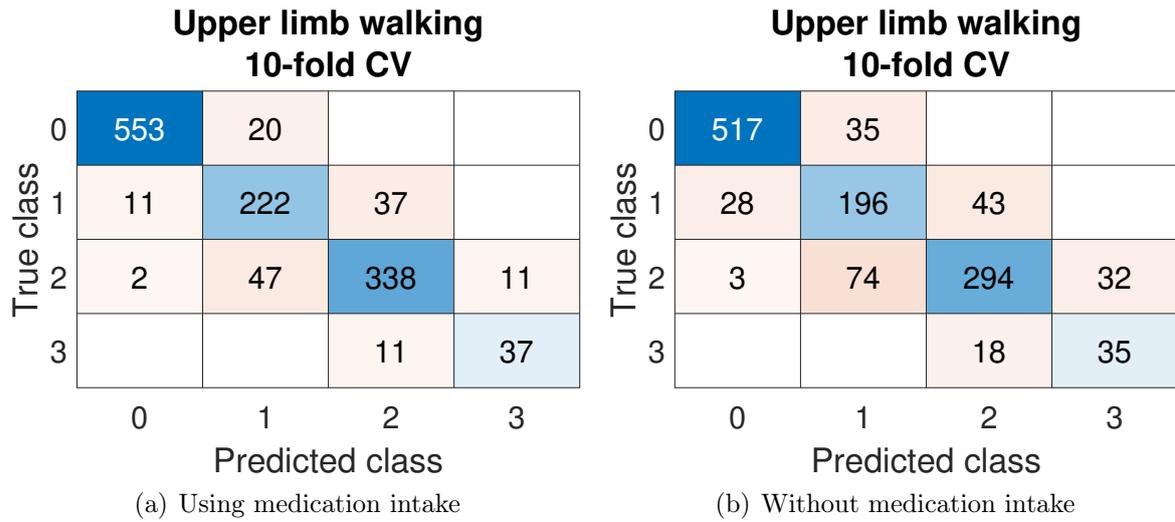


Figure 4.19: Prediction results of the upper limbs during walking subset using the 10-fold CV.

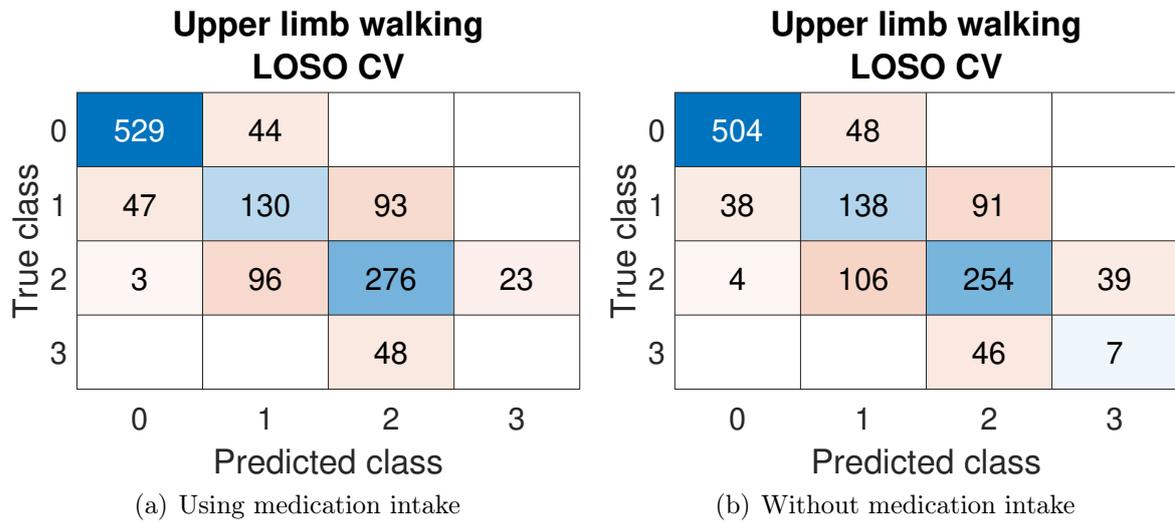


Figure 4.20: Confusion matrices of the upper limbs during walking subset using LOSO CV.

Fine movements

The results of the bradykinesia severity during fine movement actions are in figure 4.21 using the 10-fold CV, whereas in figure 4.22, there are the bradykinesia prediction applying the LOSO CV.

In table 4.4 the performances are listed; it may be noticed that during fine movements, the bradykinesia severity prediction does not change whether the medication intake feature is included or not in the feature matrix.

Gross movements

The confusion matrices related to the bradykinesia estimation during gross movement using 10-fold CV are in figure 4.23, while the results of the prediction applying LOSO CV are shown in figure 4.24.

Table 4.3: Upper limbs during walking performance with and without medication intake predictor using 10-fold and LOSO cross-validation.

Upper limbs walking				
	k-fold		LOSO	
	Medication	No medication	Medication	No medication
RMSE	0.32	0.38	0.47	0.47
MAE	0.22	0.27	0.35	0.36
Accuracy (%)	89.2	81.7	72.5	70.8
Sensitivity (%)	85.1	76.5	52.4	54.8
Specificity (%)	96.3	93.8	90.4	90.0
F1 score (%)	85.0	74.6	51.9	54.8

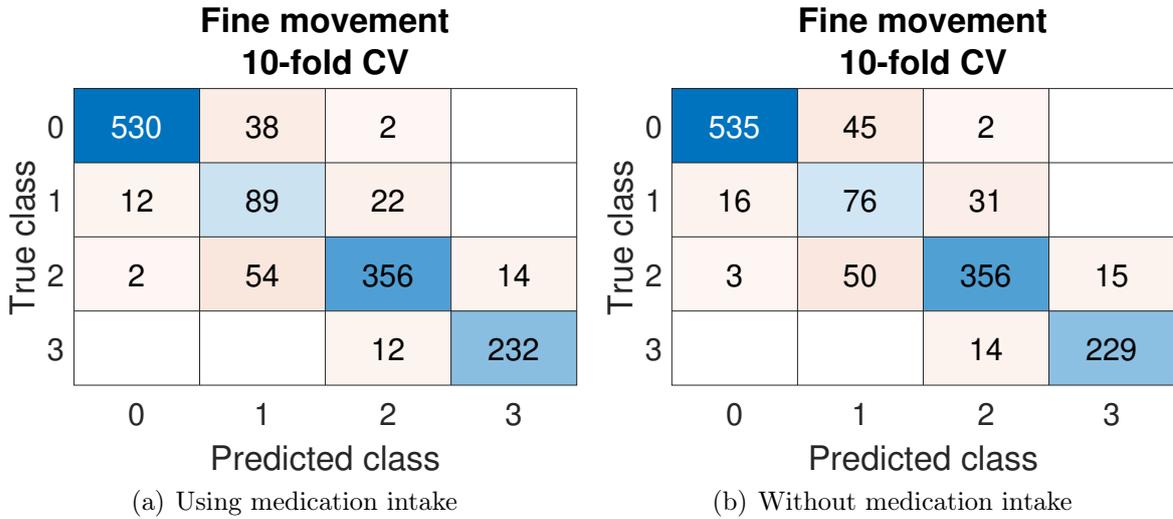
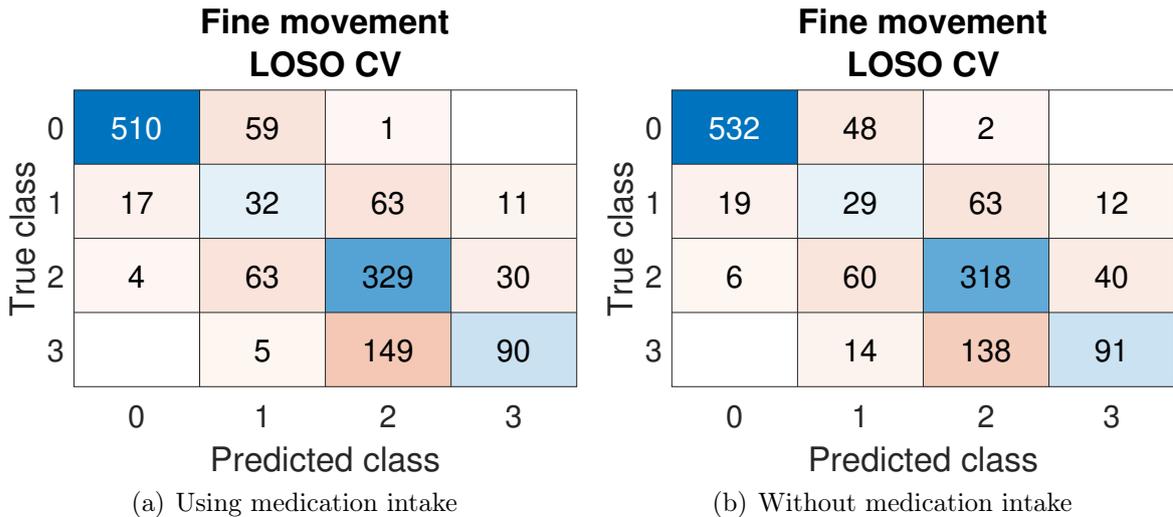
**Figure 4.21:** Confusion matrices of the upper fine movements applying the 10-fold CV.**Figure 4.22:** Prediction results of the upper fine movements using the LOSO CV.

Table 4.4: Upper fine movement performance with and without medication intake variable using 10-fold and LOSO cross-validation.

Fine movements				
	k-fold		LOSO	
	Medication	No medication	Medication	No medication
RMSE	0.33	0.34	0.53	0.54
MAE	0.18	0.20	0.36	0.35
Accuracy (%)	88.5	87.2	70.5	70.7
Sensitivity (%)	86.0	83.0	57.4	56.9
Specificity (%)	96.4	95.9	90.2	90.3
F1 score (%)	83.8	81.5	57.8	57.3

The behavior during the gross movement is different compared to the other movement patterns; using the 10-fold CV the performance is better using the medication intake information, on the other hand, the LOSO results show that the bradykinesia estimates are superior without the medication intake variable.

In table 4.5 are summarized the performance obtained in the four conditions.

Gross movement 10-fold CV					Gross movement 10-fold CV					
True class	0	2291	104	1		0	2292	114		
	1	88	216	157	2	1	122	124	216	1
	2	14	223	1244	334	2	11	250	1146	405
	3		1	120	1083	3			150	1047
		0	1	2	3		0	1	2	3
		Predicted class					Predicted class			
		(a) Using medication intake					(b) Without medication intake			

Figure 4.23: Prediction results of the upper gross movement subset using the 10-fold CV.

Nevertheless, the discrepancy in the bradykinesia estimates between the usage of the medication intake feature or not is not so significant; the two approaches will be carried on for the estimates of bradykinesia severity using the apartment data.

The choice could be justified by the fact that the medication intake is a good prediction for the motor fluctuations and they will be detected in the apartment analysis; keeping this feature could lead to monitor the fluctuations with a better accuracy.

4.8 Apartment bradykinesia prediction

The results regard the validation of the movement classifier, to recognize if the movement belongs to fine or gross class and the bradykinesia severity prediction in

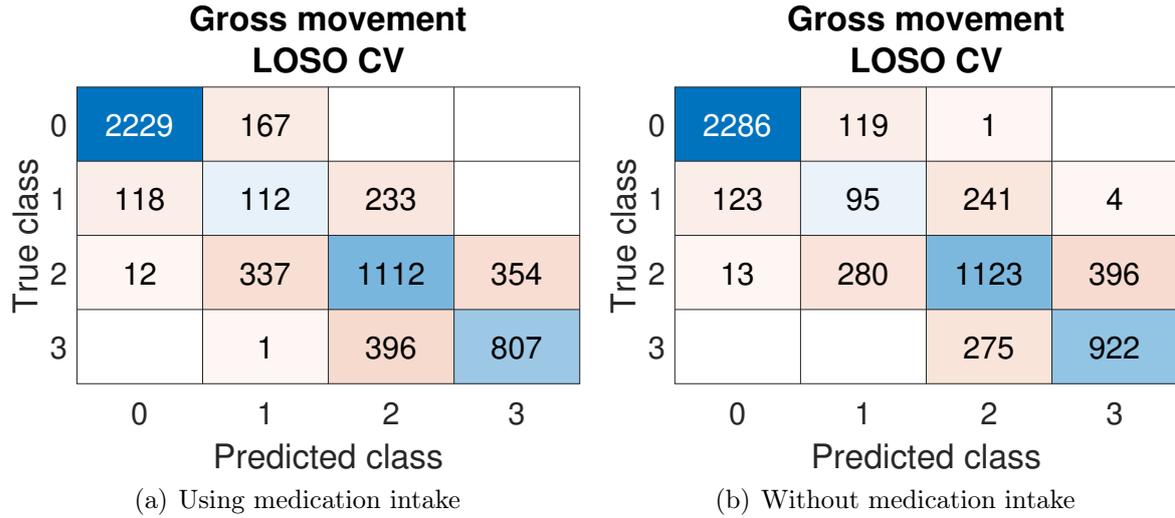


Figure 4.24: Confusion matrices of the upper gross movements applying the LOSO CV.

Table 4.5: Upper gross movement performance with and without medication intake feature using 10-fold and LOSO cross-validation.

Gross movements				
	k-fold		LOSO	
	Medication	No medication	Medication	No medication
RMSE	0.38	0.41	0.45	0.43
MAE	0.27	0.31	0.35	0.34
Accuracy (%)	0.82	78.4	72.5	75.3
Sensitivity (%)	75.1	68.2	61.4	63.6
Specificity (%)	94.3	92.9	90.9	91.7
F1 score (%)	73.9	67.2	61.3	63.2

the apartment setting.

The estimate errors will be presented for each subject and with the two feature sets, one including the medication intake and the other excluding it.

4.8.1 Movement classifier

Before the training of the RF learning algorithm, the redundant features are removed following the procedure described in section 3.7.1. The retained features after the removal are in figure 4.25.

The choice of the number of trees is using the OOB error attempting a range of trees from 1 to 100. The optimal number is 50 (figure 4.26), while the minimum leaf size is set to 5 to reduce the tendency to the overfitting. Before the optimization phase the minority class, fine movements, is oversampled using ADASYN increasing the instances from 3190 to 12000, which is the same size of the gross movement class. The validation of the learning algorithm is performed on the laboratory data applying the 10-fold and LOSO cross-validation methods. The results are shown in terms of confusion matrices in figure 4.27. In table 4.6 are summarized the performance measure extracted by the confusion matrices.

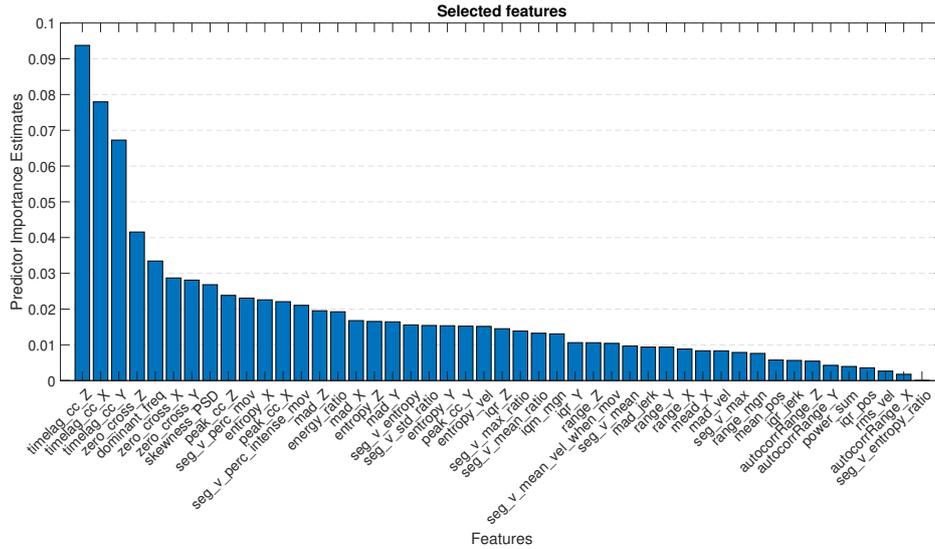


Figure 4.25: Feature importance of the selected variables after the removal of the redundant ones. The selected feature are 47 and the most significant for the classification task are the cross-correlation features, the zero crossing rate, and the dominant frequency.

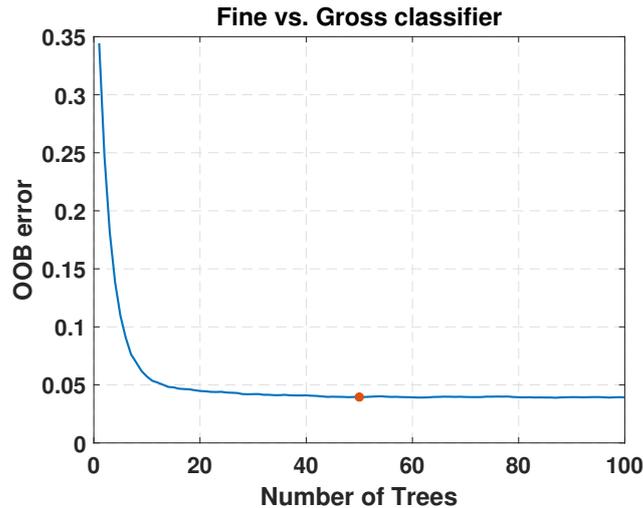


Figure 4.26: OOB error curve for the number of trees optimization. The error decrease significantly till 50 trees, then the curve is almost flat. The red dot shows the optimal number of weak learners for the model.

4.8.2 Longitudinal bradykinesia prediction

The results for the apartment are in terms of RMSE of the bradykinesia severity prediction, comparing the outcome of the models with the true labels given by the clinicians using the collected videotapes of the apartment assessment. Besides, the results regards the comparison between the two subsets of features: with medication intake and without.

Figure 4.28 shows the RMSE over the subjects involved in the apartment assessment in the two configurations of variables.

The overall estimation errors over the entire group of subjects is in table 4.7, where the error is shown in terms of average and standard deviation over the errors of the

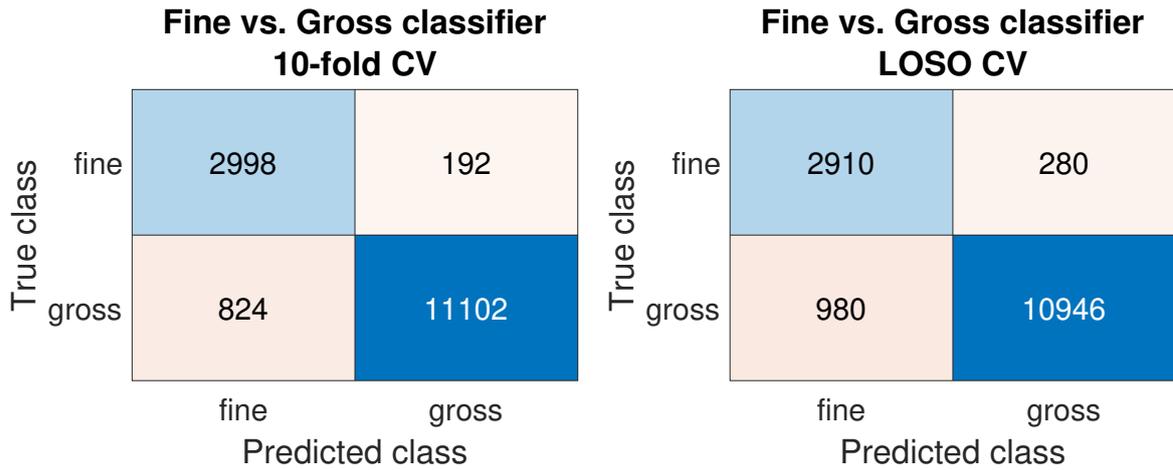


Figure 4.27: Validation results in terms of confusion matrices of the classifier to recognize fine and gross movements. On the left side the result is obtained using the 10-fold CV, while on the right side the results is applying the LOSO CV.

Table 4.6: Validation performance of the classifier for the recognition between gross and fine movements.

Fine vs. Gross classifier		
	k-fold	LOSO
Accuracy (%)	93.3	91.6
Sensitivity (%)	93.1	91.8
Specificity (%)	94.0	91.2
F1 score (%)	95.6	94.6

different subjects.

Table 4.7: Apartment estimation error in terms of average and standard deviation of the RMSE over the entire group of subjects.

Apartment error		
	medication intake	no medication
RMSE	0.90 ± 0.25	0.89 ± 0.25
MAE	0.65 ± 0.32	0.66 ± 0.32

In accordance with the results, the two configurations lead to the same bradykinesia severity estimation error, despite the medication intake appears more significant compared with the other variables.

Finally, an example of each configuration is illustrated in figure 4.29 and figure 4.30, where the blue line is the raw outcome of the model and the red one is the smoothed outcome using the median filter.

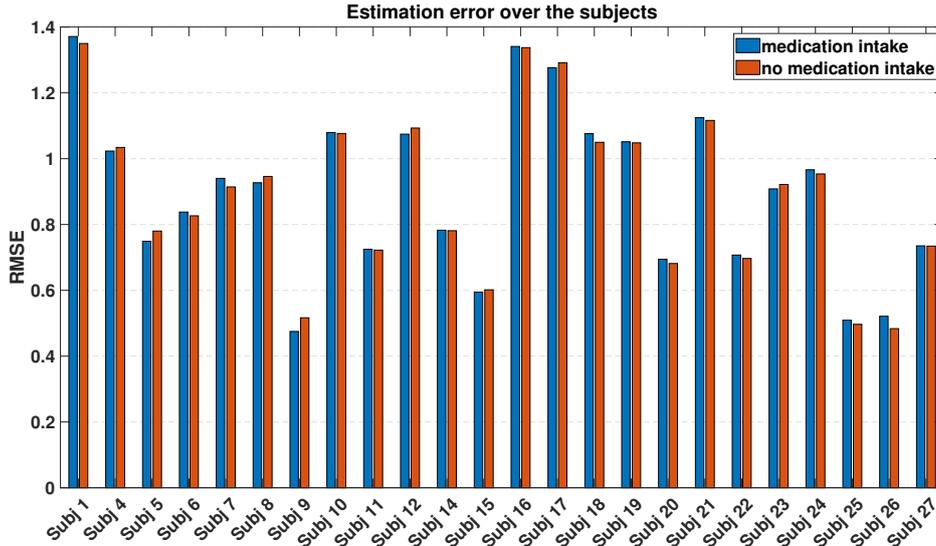


Figure 4.28: Apartment estimation error along the subjects. For each patient involved in the apartment assessment the error in terms of RMSE is shown, comparing the results on bradykinesia prediction with two different subset of features, one including the medication intake and the other excluding it. The results show that there is no difference between the two conditions.

4.9 Conclusions

To summarize the results chapter the overall result of the prediction of bradykinesia in the laboratory is 0.5 in terms of RMSE using LOSO; instead, the prediction error in the apartment is 0.90 using the medication intake variable and 0.89 without the use of this feature. These results are obtained treating the movement patterns separately, hence for the apartment analysis a classifier for the movement recognition, fine versus gross movements, is implemented with a validation accuracy of 91.6% using the LOSO. Other minor results are the optimization of the threshold for motion detection and the introduction of a method for removing redundant features based on the ReliefF method and the correlation between features.

Finally, the cleaning of the data, although fruitful for the prediction in the laboratory, has not proved efficient for the prediction of bradykinesia in an unconstrained environment.

In the next chapter, the discussion of the results presented in this chapter are discussed.

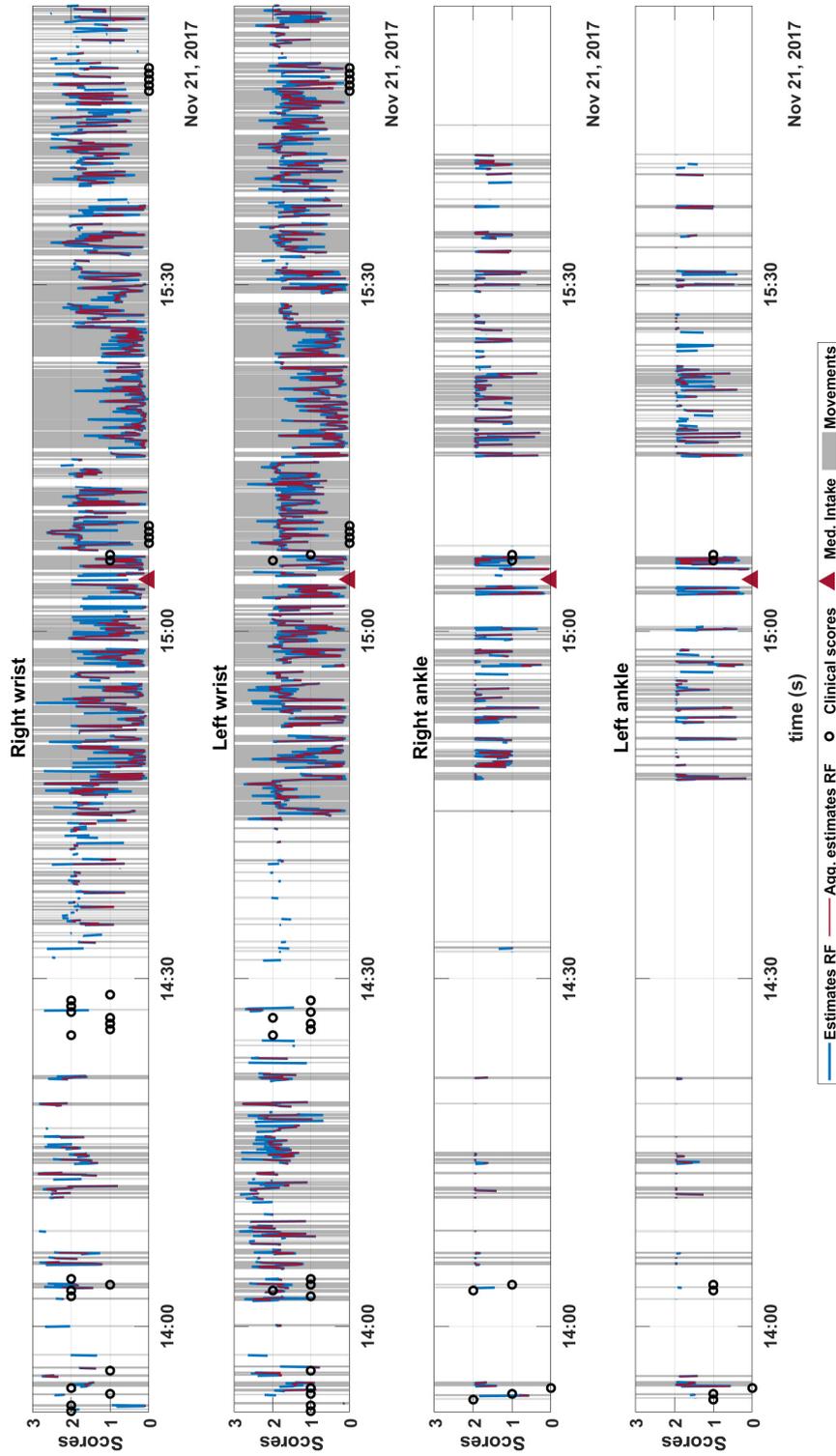


Figure 4.29: Longitudinal representation of the bradykinesia severity prediction of the subject 19 during the apartment assessment on each limb using the medication intake feature. The visualized time period is around 2 hours in which the subject took the medication, this information is marked using a red triangle on the temporal axis. The blue line is the raw output of the models, instead the red line are the smoothed predictions using a median filter with a windows of 10s. The clinical scores are the black circles, and each of them represents a 30s of examination. The fluctuation of the bradykinesia symptom is not significant, mostly on the lower limbs, and a greater oscillation is after the medication intake.

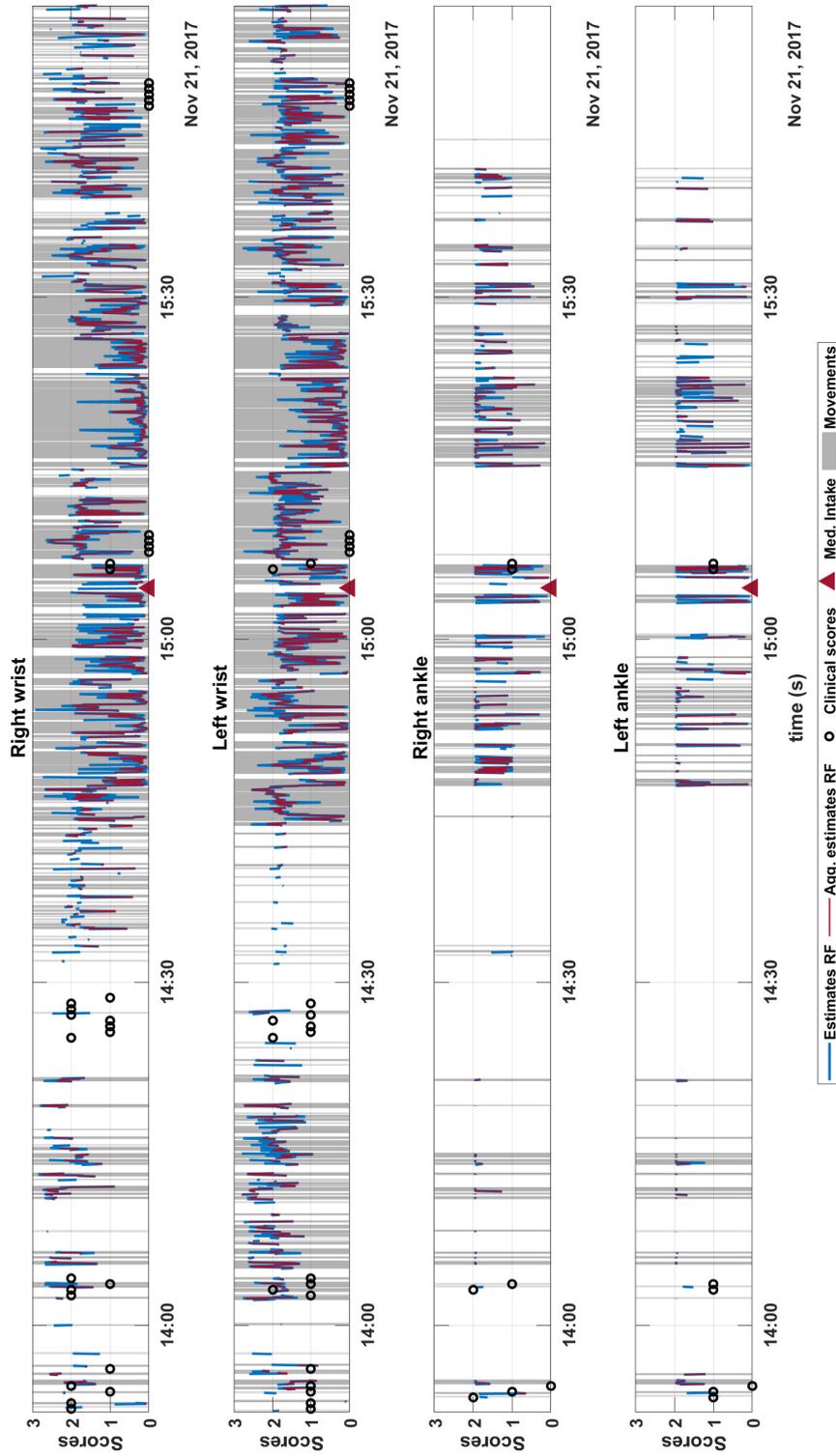


Figure 4.30: Longitudinal representation of the bradykinesia severity prediction of the subject 19 during the apartment assessment for each limb without the medication intake predictor. The visualized time period is around 2 hours in which the subject took the medication, this information is marked using a red triangle on the temporal axis, but this information is not used. The blue line is the raw output of the models, instead the red line are the smoothed predictions using a median filter with a windows of 10s. The clinical scores are the black circles, and each of them represents a 30s of examination. The trend of the fluctuations is slight, as in the previous case, and the greater fluctuation is after the medication intake.

Chapter 5

Discussion

The purpose of this work is to investigate the feasibility and reliability of a system able to track the severity of bradykinesia and motor fluctuation in individuals with PD, using wearable sensors and ML approaches for the analysis and processing of the data.

The motion detection is one of the fundamental steps of the processing since the bradykinesia involves slowness of movement. After a quantitative analysis to detect the movement using a threshold on the acceleration magnitude, the optimal threshold level is 22 times the noise level of the sensor (figure 4.3). These validation results have limitations related to the quality of the labels that identify if a task is categorized as rest or voluntary movement. Since the involved subjects have other symptoms, like dyskinesia it might be possible that a task categorized as rest has inside an involuntary movement that is detected by the threshold.

Despite these limitations, the overall result of the movement detection is 79% and it is suitable for the bradykinesia dataset (figure 4.4).

The analysis of the projections of the instances leads to consider different types of movements separately, to reduce the variability introduced by the amount of executed tasks in the laboratory. The result of this analysis is to recognize the movements into four groups of movement patterns: walking, the swing of the arms during walking, fine movements and gross movements.

Introducing this partition a limitation occurred about the walking data because in this cluster the bradykinesia severity scores are from 0 to 2 instead of severity 3. This restriction causes the inability to predict score greater than 2 in the apartment setting for what concern the walking task.

For the optimization of the dataset, a method to remove the redundant features is proposed applying the ReliefF predictor importance coupled with the correlation among the features. Since this method is proposed in this work, a quantitative validation is performed, comparing it with other feature selection methods. The result is promising because the proposed method can select a lower number of features maintaining the prediction error satisfactory (figure 4.5).

Once noted the performance of the proposed method, it is applied to the groups of movements to retain only the important features.

The outcome after the removal of irrelevant features is in figure 4.9. It is noted that

some features are invariant respect to the movement patterns, like the dominant frequency, the features extracted by the cross-correlation signals, and the zero-crossing rates, nonetheless the medication intake being the most important. Other features are selected only for certain kinds of movement groups meaning that these features are specific to predict bradykinesia using precise movement patterns. An additional consideration is about the features added in section 3.5.4, which are retained after the feature selection, however, the SMA, the MEAD of the acceleration magnitude, and the midhinge are discarded.

The data cleaning approach is encouraging because it can remove noisy points generated by the windowing approach for the analysis and the scoring method based on tasks. The latter is the main limitation of the laboratory analysis leading to have mislabeled windows.

The results of the cleaning are respectable because the variability is reduced inside every group of movements, increasing the bradykinesia prediction in the laboratory setting (figure 4.16).

The reduction of variability allows getting a cleaner training set for the bradykinesia severity prediction in the apartment setting and retaining only reliable points deleting the mislabeled ones.

The unbalancing of the dataset represents a limitation of the work and also a problem for the design of the predictive model. Before the training of the models, the oversampling of the minority class instances is necessary. The results of the balancing in figure 4.13, where class 0 is not oversampled since it is the biggest one; rather, classes 1 and 3 are the most oversampled, attempting to reduce the degree of unbalancing.

The training phase led to an optimization of the RF parameters using OOB error curves and choosing the best number of trees for each model of each movement pattern.

In figure 4.15, the OOB curves decrease quickly until there is no significant reduction of the error. In correspondence at the beginning of the plateau, the number of trees is identified for each model. Table 4.1 summarizes the results obtained for the optimization phase preceding the training of the models. The optimization phase reduces the possibility that the learning algorithm overfits the training set instances.

The bradykinesia severity prediction in the laboratory setting is promising since the cleaning of the data reduced the variability among the classes. The prediction error in terms of RMSE for all the movement patterns is under 0.5 using the LOSO CV and the overall RMSE, applying the LOSO, is 0.47 for the two feature configurations, with medication intake and without it.

The best result belongs to the walking tasks (table 4.2) since the characteristic of this movement is ideal for bradykinesia detection because this motion is periodical. Instead, the error of the gross and fine movements is slightly higher (table 4.5 and table 4.4), the motivations are due to the higher complexity and for the positioning of the sensors respectively; since for the recording of fine movements, the best placement of the sensors should be on the fingers.

Summarizing the performance in the laboratory, the prediction error is satisfactory considering that the prediction range is between score 0 and score 3. Besides, there

is no significant difference between the two subsets of features; this may be an advantage because the medication intake data could not be always available in natural settings and relying only on the accelerometer sensors does not change the severity prediction.

Despite this result, the investigation carries on the different feature subsets to understand if in the apartment environment the behavior is the same as in the laboratory or not.

Regarding the apartment results, variability on the prediction error along the subjects is present (figure 4.28), showing the trained models have difficulties to predict correctly the bradykinesia severity of some patients, such as subjects 1, 16 and 17. The overall RMSE over the entire group of subjects is 0.90 and 0.89 for the feature set with the medication and for the one without it respectively (table 4.7).

Unfortunately, these results are rough estimates, because the true labels in the apartment setting are very few for each patient over a recording time of more than 5 hours and the scores are given by the clinicians observing videotapes.

The aspect of the videotapes must not be overlooked, because for a clinician it is tough to score a subject during ADL and more so looking at a videotape, where the positions of the cameras might influence the perception of the movement quality.

The last aspect to take into account for the interpretation of the apartment results is the poor quality of the medication intake information during the apartment assessment because in some cases this information is not available.

It might be thought that the absence of a discrepancy between the two subsets of features can be attributed to the low accuracy of the medication intake data.

The attempt to monitor the motor fluctuation is challenging and in figure 4.29 and figure 4.30 the models can detect only small fluctuations, but not the overall trend described in figure 1.5.

Overall, taking into account the limitations of the work and the degree of freedom of the subjects during the apartment assessment the results can be considered as a starting point for further analysis in unconstrained environments. Putting more efforts to recognize firstly the movement category more accurately, though the performance of the classifier to recognize fine and gross movements is encouraging with an accuracy of 91.6% using the LOSO (table 4.6), and then predict the bradykinesia severity using trained models with more instances that represent better the different severities could increase the chance to improve the accomplishment in unconstrained settings.

Chapter 6

Conclusions

Although this work is a first investigation for the prediction of the severity of bradykinesia in unconstrained environments during ADL and using wearable sensors, the results are promising and offer insights for future improvements.

The prediction accuracy in the simulated apartment is still far from that achieved during the assessment in the laboratory, due to the difference in movements performed in the two settings, by the different methods of attribution of clinical scores, and the use of only accelerometer sensors.

This discrepancy can be thinned at first instance by selecting only the tasks most similar to those performed in the laboratory and increasing the number of clinical scores during the assessment in the apartment to more accurately estimate the performance of trained models and better comprehension, by the clinician, of the subject state to titrate the medication accordingly to the patient needs.

Furthermore, in this work, concepts introduced by other authors working in this field have been adopted and modified, such as the subdivision of motor tasks into patterns of movements and the method for removing noisy samples collected in the laboratory.

Nonetheless, a method for removing redundant features has been proposed, proving how it can be compared to feature selection methods already existing in the literature.

Overall, this work shows how the proposed and undertaken concepts could represent a concrete future solution for monitoring the severity of PD symptoms, despite the limitations and possible improvements regarding unconstrained environments.

The improvements are wide due to the complexity and the intrinsic nature of bradykinesia and the motor complications related to PD that are difficult to deal with little data availability and with specifications and constraints associated with the project to accomplish. In addition, this symptom is not easy to estimate during ADL and the clinician sometimes finds difficult to assess the subject state.

Nevertheless, future developments offer broad scenarios for improving the work proposed to achieve better reliability of the severity prediction in the community setting or patient's home.

New data analysis approaches to increase the class separation among the bradykinesia severities and the use of learning algorithms belonging to the branch of the DL, that consider the temporal sequences between one movement and another one, could exploit the information of symptom severity in the past for the next predic-

tions of the state improving the overall accuracy of the system especially during the longitudinal assessment during the daily life.

Bibliography

- [1] L. Borzì, M. Varrecchia, G. Olmo, C. A. Artusi, M. Fabbri, M. G. Rizzone, A. Romagnolo, M. Zibetti, and L. Lopiano, “Home monitoring of motor fluctuations in parkinson’s disease patients,” *Journal of Reliable Intelligent Environments*, vol. 5, pp. 145–162, Sep 2019.
- [2] J. Jankovic, “Parkinson’s disease: clinical features and diagnosis,” *Journal of neurology, neurosurgery, and psychiatry*, vol. 79, no. 4, 2008.
- [3] L. M. de Lau and M. M. Breteler, “Epidemiology of parkinson’s disease,” *Lancet Neurology*, vol. 5, no. 6, pp. 525–535, 2006.
- [4] C. A. Davie, “A review of Parkinson’s disease,” *British Medical Bulletin*, vol. 86, pp. 109–127, 04 2008.
- [5] “New sensor and wearable technologies to aid in the diagnosis and treatment monitoring of parkinson’s disease,” vol. 21, no. 1, pp. 111–143, 2019.
- [6] OpenStax, *Anatomy and Physiology*. Houston, Texas: OpenStax, 10 ed., 2013.
- [7] Wikipedia contributors, “Basal ganglia — Wikipedia, the free encyclopedia,” 2019. [Online; accessed 15-October-2019].
- [8] “Parkinson’s foundation.” <https://www.parkinson.org>.
- [9] O.-B. Tysnes and A. Storstein, “Epidemiology of parkinson’s disease,” *Journal of Neural Transmission*, vol. 124, no. 8, pp. 901–905, 2017.
- [10] J. E. Thorp, P. G. Adamczyk, H.-L. Ploeg, and K. A. Pickett, “Monitoring motor symptoms during activities of daily living in individuals with parkinson’s disease,” *Frontiers in Neurology*, vol. 9, p. 1036, 2018.
- [11] J. Jankovic, “Motor fluctuations and dyskinesias in parkinson’s disease: Clinical manifestations,” *Movement Disorders*, vol. 20, no. S11, pp. S11–S16, 2005.
- [12] L. V. Kalia and A. E. Lang, “Parkinson’s disease,” *The Lancet*, vol. 386, no. 9996, pp. 896–912, 2015.
- [13] S. Patel, D. Sherrill, R. Hughes, T. Hester, N. Huggins, T. Lie-Nemeth, D. Standaert, and P. Bonato, “Analysis of the severity of dyskinesia in patients with parkinson’s disease via wearable sensors,” in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN’06)*, vol. 2006, pp. 4 pp.–126, IEEE, 2006.

- [14] C. Ossig, A. Antonini, C. Buhmann, J. Classen, I. Csoti, B. Falkenburger, M. Schwarz, J. Winkler, and A. Storch, “Wearable sensor-based objective assessment of motor symptoms in parkinson’s disease,” *Journal of Neural Transmission*, vol. 123, no. 1, pp. 57–64, 2016.
- [15] Q. W. Oung, H. Muthusamy, H. L. Lee, S. N. Basah, S. Yaacob, M. Sarillee, and C. H. Lee, “Technologies for assessment of motor disorders in parkinson’s disease: a review,” *Sensors*, vol. 15, no. 9, pp. 21710–21745, 2015.
- [16] A. J. Espay, P. Bonato, F. B. Nahab, W. Maetzler, J. M. Dean, J. Klucken, B. M. Eskofier, A. Merola, F. Horak, A. E. Lang, *et al.*, “Technology in parkinson’s disease: challenges and opportunities,” *Movement Disorders*, vol. 31, no. 9, pp. 1272–1282, 2016.
- [17] E. Rovini, C. Maremmani, and F. Cavallo, “How wearable sensors can support parkinson’s disease diagnosis and treatment: A systematic review,” *Frontiers in Neuroscience*, vol. 11, p. 555, 2017.
- [18] 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, no. 6, pp. 864–873, 2009.
- [19] J. Cancela, M. Pansera, M. T. Arredondo, J. J. Estrada, M. Pastorino, L. Pastor-Sanz, and J. L. Villalar, “A comprehensive motor symptom monitoring and management system: The bradykinesia case,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, vol. 2010, pp. 1008–1011, IEEE, 2010.
- [20] Z. Lin, H. Dai, Y. Xiong, X. Xia, and S.-J. Horng, “Quantification assessment of bradykinesia in parkinson’s disease based on a wearable device,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 803–806, IEEE, 2017.
- [21] C. Gao, S. Smith, M. Lones, S. Jamieson, J. Alty, J. Cosgrove, P. Zhang, J. Liu, Y. Chen, J. Du, *et al.*, “Objective assessment of bradykinesia in parkinson’s disease using evolutionary algorithms: clinical validation,” *Translational neurodegeneration*, vol. 7, no. 1, p. 18, 2018.
- [22] L. D. Biase, S. Summa, J. Tosi, F. Taffoni, M. Marano, A. C. Rizzo, F. Vecchio, D. Formica, V. D. Lazzaro, G. D. Pino, and M. Tombini, “Quantitative analysis of bradykinesia and rigidity in parkinson’s disease,” *Frontiers in Neurology*, vol. 9, 2018.
- [23] S. Memar, M. Delrobaei, M. Pieterman, K. McIsaac, and M. Jog, “Quantification of whole-body bradykinesia in parkinson’s disease participants using multiple inertial sensors,” *Journal of the neurological sciences*, vol. 387, pp. 157–165, 2018.
- [24] H. Dai, H. Lin, and T. C. Lueth, “Quantitative assessment of parkinsonian bradykinesia based on an inertial measurement unit,” *Biomedical engineering online*, vol. 14, no. 1, p. 68, 2015.

- [25] N. Y. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plötz, “Pd disease state assessment in naturalistic environments using deep learning,” in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [26] A. Rodríguez-Molinero, C. Pérez-López, A. Samà, E. de Mingo, D. Rodríguez-Martín, J. Hernández-Vara, À. Bayés, A. Moral, R. Álvarez, D. A. Pérez-Martínez, *et al.*, “A kinematic sensor and algorithm to detect motor fluctuations in parkinson disease: validation study under real conditions of use,” *JMIR rehabilitation and assistive technologies*, vol. 5, no. 1, p. e8, 2018.
- [27] B. M. Eskofier, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz, *et al.*, “Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson’s disease assessment,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 655–658, IEEE, 2016.
- [28] J.-F. Daneault, S. I. Lee, F. N. Golabchi, S. Patel, L. C. Shih, S. Paganoni, and P. Bonato, “Estimating bradykinesia in parkinson’s disease with a minimum number of wearable sensors,” in *Proceedings of the second IEEE/ACM international conference on connected health: applications, systems and engineering technologies*, pp. 264–265, IEEE Press, 2017.
- [29] M. Delrobaei, S. Tran, G. Gilmore, K. McIsaac, and M. Jog, “Characterization of multi-joint upper limb movements in a single task to assess bradykinesia,” *Journal of the neurological sciences*, vol. 368, pp. 337–342, 2016.
- [30] A. Tzallas, M. Tsipouras, G. Rigas, D. Tsalikakis, E. Karvounis, M. Chondrogiorgi, F. Psomadellis, J. Cancela, M. Pastorino, M. Waldmeyer, *et al.*, “Perform: a system for monitoring, assessment and management of patients with parkinson’s disease,” *Sensors*, vol. 14, no. 11, pp. 21329–21357, 2014.
- [31] A. Samà, C. Pérez-López, D. Rodríguez-Martín, A. Català, J. M. Moreno-Aróstegui, J. Cabestany, E. de Mingo, and A. Rodríguez-Molinero, “Estimating bradykinesia severity in parkinson’s disease by analysing gait through a waist-worn sensor,” *Computers in biology and medicine*, vol. 84, pp. 114–123, 2017.
- [32] C. L. Pulliam, D. A. Heldman, E. B. Brokaw, T. O. Mera, Z. K. Mari, and M. A. Burack, “Continuous assessment of levodopa response in parkinson’s disease using wearable motion sensors,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 159–164, 2017.
- [33] “Opal data sheet.” <https://www.apdm.com/wearable-sensors/>.
- [34] D. Karantonis, M. Narayanan, M. Mathie, N. Lovell, and B. Celler, “Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 156–167, 2006.
- [35] R. San-Segundo, H. Blunck, J. Moreno-Pimentel, A. Stisen, and M. Gil-Martín, “Robust human activity recognition using smartwatches and smartphones,” *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 190–202, 2018.

- [36] L. Lonini, A. Dai, N. Shawen, T. Simuni, C. Poon, L. Shimanovich, M. Daeschler, R. Ghaffari, J. A. Rogers, and A. Jayaraman, “Wearable sensors for parkinson’s disease: which data are worth collecting for training symptom detection models,” *npj Digital Medicine*, vol. 1, no. 1, pp. 1–8, 2018.
- [37] J. I. Hoff, A. A. V/D Plas, E. A. H. Wagemans, and J. J. Van Hilten, “Accelerometric assessment of levodopa-induced dyskinesias in parkinson’s disease,” *Movement Disorders*, vol. 16, no. 1, pp. 58–61, 2001.
- [38] R. Dunnewold, C. Jacobi, and J. van Hilten, “Quantitative assessment of bradykinesia in patients with parkinson’s disease,” *Journal of Neuroscience Methods*, vol. 74, no. 1, pp. 107–112, 1997.
- [39] O. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [40] M. Hassan, S. Huda, M. Uddin, A. Almogren, and M. Alrubaian, “Human activity recognition from body sensor data using deep learning,” *Journal of Medical Systems*, vol. 42, no. 6, pp. 1–8, 2018.
- [41] V. Lugade, E. Fortune, M. Morrow, and K. Kaufman, “Validity of using tri-axial accelerometers to measure human movement—part i: Posture and movement detection,” *Medical Engineering and Physics*, vol. 36, no. 2, pp. 169–176, 2014.
- [42] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [43] G. Hinton and S. Roweis, “Stochastic neighbor embedding,” in *Proceedings of the 15th International Conference on Neural Information Processing Systems, NIPS’02*, (Cambridge, MA, USA), pp. 857–864, MIT Press, 2002.
- [44] MATLAB, *MATLAB documentation*. Natick, Massachusetts: The MathWorks Inc., 2019.
- [45] I. Kononenko, E. Simec, and M. Robnik-Sikonja, “Overcoming the myopia of inductive learning algorithms with relieff,” *Applied Intelligence*, vol. 7, pp. 39–55, 1997.
- [46] K. Kira and L. A. Rendell, “A practical approach to feature selection,” in *Machine Learning Proceedings 1992*, pp. 249–256, Elsevier, 1992.
- [47] M. Robnik-Šikonja and I. Kononenko, “Theoretical and empirical analysis of relieff and rrelieff,” *Machine learning*, vol. 53, no. 1-2, pp. 23–69, 2003.
- [48] S. I. Lee, J.-F. Daneault, F. N. Golabchi, S. Patel, S. Paganoni, L. Shih, and P. Bonato, “A novel method for assessing the severity of levodopa-induced dyskinesia using wearable sensors,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 8087–8090, IEEE, 2015.
- [49] S. Lloyd, “Least squares quantization in pcm,” *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

- [50] H. He, Y. Bai, E. A. Garcia, and S. Li, “Adasyn: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322–1328, IEEE, 2008.
- [51] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [52] N. Poolsawad, C. Kambhampati, and J. Cleland, “Balancing class for performance of classification with a clinical dataset,” in *Proceedings of the World Congress on engineering*, vol. 1, pp. 1–6, 2014.
- [53] J. Surowiecki, “The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business,” *Economies, Societies and Nations*, vol. 296, 2004.
- [54] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 ed., 2009.
- [55] M. A. Hall, “Correlation-based feature selection for machine learning,” 1999.
- [56] C. Ding and H. Peng, “Minimum redundancy feature selection from microarray gene expression data,” *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.