



HÖGSKOLAN
DALARNA

Dalarna Doctoral Dissertations 10

Multisensor data-driven methods for automated quantification of motor symptoms in Parkinson's disease

SOMAYEH AGHANAVESI

Microdata Analysis
School of Technology and Business Studies
Dalarna University, Borlänge, Sweden
2019

Dissertation presented at Dalarna University to be publicly examined in Clas Ohlson, Borlänge, Friday, 26 April 2019 at 13:00 for the Degree of Doctor of Philosophy. The examination will be conducted in English. Opponent: Associate professor Shahina Begum (Mälardalen University).

Abstract

Aghanavesi, S. 2019. Multisensor data-driven methods for automated quantification of motor symptoms in Parkinson's disease. *Dalarna Doctoral Dissertations* 10. Borlänge: Dalarna University. ISBN 978-91-88679-00-0.

The overall aim of this thesis was to develop and evaluate new data-driven methods for supporting treatment and providing information for better management of Parkinson's disease (PD).

This disease is complex and progressive. There is a large amount of inter- and intra-variability in motor symptoms of patients with PD (PwPD). Current evaluation of motor symptoms which is done at clinics by using clinical rating scales provides limited and only part of the health status of PwPD. PD requires an accurate assessment that is approved by clinics. Therefore frequent evaluation of symptoms at micro-level is required.

Sensor systems including smartphone and motion sensors were employed to collect data from PwPD and the recruited healthy controls. Repeated measures consisting of subjective assessment of symptoms and objective assessment of motor functions were collected.

First, the smartphone-based data-driven methods were developed to quantify the dexterity presented in fine motor tests of spiral drawing and alternate tapping. The upper extremities temporal irregularity measure presented in spiral drawing tests of PwPD was further analyzed by the approximate entropy (ApEn) method. Second, tri-axial motion sensor data were collected from various tests like leg agility, walking, and rapid alternating movements of hands of PwPD during a full cycled levodopa challenge. Data driven methods for quantification of leg agility tests and a combination of multiple motor tests were developed. The clinimetric properties of the methods such as reliability, validity, and responsiveness were evaluated. In addition, the feasibility of using smartphone inertial measurement unit (IMU) sensors in comparison to motion sensors for quantifying the motor states in PD during rapid alternating movements of hands tests was investigated.

Results of the developed methods for quantification of PD motor symptoms via dexterity tests in a smartphone can be used for measuring treatment related changes in PwPD. Investigation of the ApEn method showed good sensitivity and test-retest reliability indicating that this method is useful in measuring upper limb temporal irregularity at micro-level. High convergence validity resulted from using motion sensors during leg agility tests which led to valid and reliable objective measures of PD motor symptoms. The results of fusion of sensor data gathered during standardized motor tests were promising and led to highly valid, reliable and sensitive objective measures of PD motor symptoms. The results of the analyzing acceleration IMU data showed that smartphone IMU is capable of capturing symptom information from hand rotation tests. It can provide sufficient data for quantification of the motor states.

The findings from the data-driven methodology in this thesis can be used in development of systems for follow up of the effects of treatment and individualizing treatments in PD.

Keywords: Parkinson's disease, motor symptoms, motion sensors, smartphone, microdata, multivariate analysis, data-driven, support vector machine, stepwise regression, predictive models

Somayeh Aghanavesi, School of Technology and Business Studies, Microdata Analysis,

© Somayeh Aghanavesi 2019

ISBN 978-91-88679-00-0

urn:nbn:se:du-29583 (<http://urn.kb.se/resolve?urn=urn:nbn:se:du-29583>)

Preface

The presented research in this thesis was carried out in the department of Computer Engineering, School of Technology and Business Studies, Dalarna University, Sweden during the years 2015-2019. Collection of data were done in a hospital in Uppsala, Sweden.

I'd like to express my sincere gratitude to all who contributed to the progress of my research work. Special thanks to my supervisors, Prof. Mark Dougherty, Doc. Jerker Westin, Dr. Mevludin Memedi, and Doc. Hasan Fleyeh for the continuous support of my PhD study. Mark, many thanks for all the encouragements and the motivations you gave me during past years. Your guidelines had always a positive effect on me and I'm grateful to have you as my advisor. Jerker, your immense knowledge and great effort in obtaining the fund for the project allowed me to have an opportunity to join the team and continue for my PhD studies. I learned from your way of looking at the problems and enjoyed the discussions with you during my research. Mevludin, I thank you for being collaborative and professional during different moments of my studies from the development of the methods till presentation of the results. You showed flexibility and care which helped me to go forward and succeed in the work. Hasan, thank you for your constant positive attitude, proofreading the papers and the support.

My sincere thanks also go to my co-authors: Filip Bergquist, Dag Nyholm, Marina Senek who provided me with their insights and valuable opinions in the research work. Dag, thanks a lot for your contributions to all papers and providing me with your neurological point of views when revising the papers. Filip, I'm grateful to have your comments from the neurological perspective incorporated in my research studies. Thank you for your collaborations and I enjoyed learning from your comments in the papers. Marina, you helped a lot with collecting the data and setting the stage for performing good research. Thanks for your collaboration and providing your comments in the papers.

Last but not least, I would like to thank my family: my parents for supporting me spiritually in my life. My mother has had a significant effect on me for perusing this work and look at things positively. My father gave me valuable lessons in life. My special thanks go to my dear husband Alireza and my daughter Yasmin for their love and the joy they have brought to my life. Alireza thank you for the support, love and the power you gave me during the difficult times of my work where my mind was continuously busy with work-

related questions. This work could have been incomplete without your all supports, sympathy, and encouragements.

February 27, 2019
Somayeh Aghanavesi

List of Papers

This thesis is based on the following papers. They are referred in the text by their Roman numerals.

- I Aghanavasi, S., Nyholm, D., Senek, M., Bergquist, F., and Memedi, M. (2017). A smartphone-based system to quantify dexterity in Parkinson's disease patients. *Informatics in Medicine Unlocked*, 9 (2017) 11–17. (Published)
- II Memedi, M., Aghanavasi, S., and Westin, J. (2016). A method for measuring Parkinson's disease related temporal irregularity in spiral drawings. *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Feb. 2016, pp. 410–413. (Published)
- III Aghanavasi, S., Memedi, M., Dougherty, M., Nyholm, D., and Westin, J. (2017). Verification of a Method for Measuring Parkinson's Disease Related Temporal Irregularity in Spiral Drawings. *Sensors (Basel)*. 17(10): 2341. (Published)
- IV Aghanavasi, S., Bergquist, F., Nyholm, D., Senek, M., and Memedi, M. (2019). Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge. *Journal of Biomedical and Health Informatics*. (Published as a preprint)
- V Aghanavasi, S., Bergquist, F., Nyholm D., Senek, M., and Memedi, M. (2019). Treatment response index from a multi-modal sensor fusion platform for assessment of motor states in Parkinson's disease. (Manuscript)
- VI Aghanavasi, S., Fleyeh, H., Memedi, M. and Dougherty, M. (2019). Feasibility of using smartphones for quantification of Parkinson's disease motor states during hand rotation tests. *41st International Engineering in Medicine and Biology Conference*. (Submitted)

Reprints were made with permission from the respective publishers.

My contributions to the included papers were:

Paper I – Data processing, methodology development, data analysis, results interpretation, writing the first version of the manuscript and revising the manuscript.

Paper II – Partly involved in method development and data analysis, writing parts of the manuscript and reviewing the rest.

Paper III – Data processing, methodology development, data analysis, results interpretation, writing the first version of the manuscript and revising the manuscript.

Paper IV – Data processing, methodology development, data analysis, results interpretation, writing the first version of the manuscript and revising the manuscript.

Paper V – Data processing, methodology development, data analysis, results interpretation, writing the first version of the manuscript and revising the manuscript.

Paper VI – Data processing, methodology development, data analysis, results interpretation, writing the first version of the manuscript and revising the manuscript.

Contents

Introduction.....	11
Motivation.....	11
Research questions.....	13
Quantification of fine motor symptoms using smartphones.....	13
Quantification of motor states using motion sensors.....	14
Research approach.....	16
Background.....	20
Parkinson’s disease.....	20
Subjective assessment of PD.....	21
Objective assessment of motor symptoms.....	22
Data collection and preparation.....	25
Subjects and experiment.....	25
Sensor measurements.....	29
Methods.....	31
Approximate entropy.....	31
Discrete wavelet transform.....	32
Principal components analysis.....	34
Linear mixed effect modelling.....	35
Cross-correlation.....	35
Linear regression.....	36
Stepwise regression.....	37
Least absolute shrinkage and selection operator.....	37
Decision trees.....	38
Multilayer perceptrons.....	38
Support vector machines.....	39
Results.....	40
Paper I.....	40
Paper II.....	41
Paper III.....	42
Paper IV.....	43
Paper V.....	48
Paper VI.....	52

Discussion and conclusions53
 Limitations55
 Future studies56
References.....58

Abbreviations

Anova	Analysis of variance
AS	Arm Swing
AUC	Area Under the receiving operating characteristics Curve
ApEn	Approximate Entropy
CI	Confidence Interval
DBS	Deep brain stimulation
DT	Decision Trees
DWT	Discrete Wavelet Transform
Dys	Dyskinesia
FDA	US food and drug administration
HE	Healthy Elderly
HY	Hoehn & Yahr Scale
Hz	Hertz
ICC	Intra-Class Correlation coefficient
IMU	inertial measurement unit
JSON	JavaScript Object Notation
LCIG	Levodopa-Carbidopa Intestinal Gel
LME	Linear Mixed-Effects models
LR	Linear Regression
Macc	Magnitude of acceleration
Mgyr	Magnitude of orientation
MDS	Movement Disorder Society
MEMS	MicroElectroMechanical Systems
MLR	Dyskinesia
MLP	Multi-Layer Perceptron
NK	Not Known
PC	Principal Component
PCA	Principal Component Analysis
PD	Parkinson's Disease
PDQ-39	Parkinson's Disease Questionnaire 39-item
PwPD	Patients with Parkinson's Disease
RMSE	Root mean square error
RT	Regression Trees
SD	Standard Deviation value

SD-DV	Standard Deviation of Drawing Velocity
SumUPDRA	Summation of UPDRS Scale items
SVM	Support Vector Machine
TIS	Temporal Irregularity Score
TRIMMS	Treatment Response Index from Multimodal Motion Sensors
TRS	Treatment Response Scale
UPDRS	Unified Parkinson's Disease Rating Scale
WSTS	Wavelet Spiral Test Score
Xacc	Acceleration in X axis
Xgyr	Orientation in X axis
Yacc	Acceleration in Y axis
Ygyr	Orientation in Y axis
Zacc	Acceleration in Z axis
Zgyr	Orientation in Z axis

Introduction

Motivation

The human motor system controls body posture and its movements. Parkinson's disease (PD) affects the motor control of humans and quantification of motor symptoms of PD is complex and challenging. The complexity is related to the variabilities present in the manifestation of symptoms between patients and within patients. PD is chronic and over time as the disease progress, motor fluctuations develop.

One of the currently available therapies which has been used for over 40 years is oral Levodopa [1] that is symptomatic but does not stop the progression of the disease. The clinical assessment of PD motor symptoms is by using clinical rating scales. The most commonly used clinical rating scales is Unified Parkinson's Disease Rating Scale (UPDRS) [2] that is based on observations and judgments of clinicians. PD treatment is guided by recording patients history which is based on patients recall and it is sometimes documented in paper home diaries [3].

However, using clinical ratings at clinics needs trained experts, and requires PD patients to visit the clinics. Frequent evaluation of symptoms at micro-level by clinicians is not feasible since it is time consuming and costly [4]. There is inter- and intra-observer variability when using the scales [5]. Since the clinical visits are not frequent, the assessment only provides limited and only part of the health status of the PD patients. This limits the healthcare providers to manage the symptoms, improve the clinical outcomes, and offer treatment at an individualized level. For those reasons, repeated, unbiased, and observer-independent measurements are necessary to accurately measure a more complete extent of motor symptoms in PD [6].

Sensor technology consists of connected and sensing components embedded in a smartphone and wearables. It provides means for repeated, long term, remote, convenient and cost-effective motor symptom assessments which enhances the quality of life of PD patients [7], [8].

In this thesis, the quantification of PD motor symptoms using sensor technology is done. The data was collected from various parts of the body of PD patients over a course of tests when they were performed before and after levodopa medication. Data-driven methods were developed and their clini-

metric properties were evaluated to objectively estimate the states of PD patients. The overall aim was to investigate new methods for quantifying the motor states of PD patients. It can be observed in two sub-objectives.

First, to investigate the use of smartphone-based data-driven methods in quantifying the dexterity of the motor symptoms present in fine motor tests. It was to examine the combination of spiral drawing and alternate tapping tests data used in methods for quantification of the motor symptoms. In addition, a part of the first aim was to investigate the upper extremities temporal irregularity measure presented in spiral drawing tests of PD patients. Subsequently, it was aimed to assess the distinguishability of a score generated by an entropy-based method for separating PD patients at different stages of the disease and healthy controls.

Second, the aim was to develop data-driven methods and evaluate their clinimetric properties when motion sensor data of various tests such as leg agility, walking, and rapid alternating movements of hands of PD patients were gathered during a levodopa challenge. It was first to develop and evaluate the methods for quantification of the leg agility test and to estimate the presence and the severity of PD motor states. Then it was intended to develop data-driven methods using combined feature sets calculated from multiple tests and then to evaluate the clinimetric properties of the methods. Various features were calculated from sensor data in relation to what kinematics could be important from a clinical view, such as walking speed during the walking test, number of foot taps in the leg agility test, and amount of acceleration of arm swing during walking and etc. The relevance of the calculated and selected features from sensor data to clinical ratings, the reliability of the scores generated by the developed methods, the responsiveness of the methods to levodopa treatment, and differentiate ability of the methods between healthy controls and PD patients were examined. In addition, it was intended to investigate the feasibility of using smartphone inertial measurement unit (IMU) sensors in comparison to motion sensors for quantifying the motor states in PD during rapid alternating movements of hands tests.

The presented work is part of the projects “Flexible Levodopa Optimizing Assistive Technology” (FLOAT) and “Multimodal motor symptoms quantification platform for individualized Parkinson’s disease treatment” (MUSYQ) that were performed during 2015-2018. The overall aim of the project was to develop a system to improve treatment for PD patients who consume Levodopa in microtablets [9]. Micro tablets allow finer control of dosage levels, but the effect of dosage variation needs to be explored and dosage protocols to be optimized. The project mainly focused on measuring the symptoms accurately and objectively and simulating models of pharmacodynamics to predict the effect of a given dosage, both in terms of blood concentration levels and also resultant symptoms. It was intended to develop a new type of sensor technology platform that can quantitatively evaluate symptoms related to PD. This work was also financed by the project “Remote monitoring of Parkinson's

disease - Empowerment of patients and improved treatment using ICT-based tools” (EMPARK).

The research profile of the thesis is microdata-analysis. Microdata can be described as data at an individual level and it contains data about individuals at low level e.g. responses to a questionnaire designed for a specific research question. Whereas macrodata refers to data that are aggregated in many different forms. Macrodata is at a system-level which cannot be divided to lower level units. E.g. the number of people grouped by age or gender, or income. In this thesis, microdata was collected from each individual through sensors. The microdata was collected from multiple body segments, throughout a day, and at different levels of medication, together with the symptom rating scores that were gathered from neurological experts about each individual. Microdata can include a large amount of data from each individual. Having access to microdata allows investigation and analysis at a detailed level. Microdata analysis focuses on processing and performing the analysis on microdata. Microdata analysis can be considered as a kind of data analysis in data science which is a multidisciplinary field with mathematics and statistics as the key subsets.

Research questions

Quantification of fine motor symptoms using smartphones

Dexterity manifests as a result of disruptions in normal motor control processes that are affected by PD [10]. The ability to perform small movements requiring hands, wrists, fingers and eye contact is defined as fine motor movements. PD affects the control of movement variables such as finger position and speed. The reduction of the ability in performing fine motor activities affects their overall quality of life. The assessment procedure is done by neurological experts using Unified Parkinson’s Disease Rating Scales (UPDRS). However, this scale is subject to limitations like inter- and intra- rater variability, the impossibility of long term continuous assessment at clinics, and recall bias. Therefore assessment of fine motor impairments in PD needs to be done when repeated measures are collected by sensors. First, the focus of the thesis was to employ smartphones for quantification of fine motor impairments. PD patients performed two fine motor tests, a spiral drawing test, and an alternate finger tapping test. The use of smartphones for quantification of motor dexterities have been the focus in the first parts of this thesis.

The following research questions were investigated specifically:

- How to develop a method for quantification of motor symptom dexterity where the data from spiral drawing and tapping tests were combined?

- What measures from these methods would be closely related to the clinical evaluation of the tests?
- What are the clinimetric properties of the developed methods? (Properties such as validity of the scores to clinical measures; reliability of the scores and their responsiveness to treatment)
- How to develop an entropy-based method to measure the temporal irregularity in spiral drawing tests?
 - What are the clinimetric properties of this method?
 - What are the results achieved by this method compared to other methods?

Quantification of motor states using motion sensors

It is important for clinicians to evaluate the motor symptoms of PD patients to be able to make decisions for the medication dose and the treatment. Continuous evaluation for prolonged periods with a wider and more accurate picture of PD symptoms and their fluctuations is needed. Evaluation of experts are based on observations and some test performances containing information that is not evident to the naked eye. The quantification of PD motor states where the data is collected from lower extremities e.g. leg agility tests could be a potential for this, as there have been only a few studies [11], [12], [13] analyzing this test. Quantification of lower extremities has mostly included walking tests in previous studies [14], [15], [16]. Using the leg agility test could be needed for severely disabled PD patients, e.g. at the stage of 4 and 5 of the Hoehn and Yahr scale [17], where there is a high chance of falling during walking.

Furthermore, in order to have a better understanding of PD as a multidimensional disease, the assessment methods should address different aspects of the disease. The motor symptoms are variable between and within PD patients where they manifest at different parts of the body of PD patients. Recent review studies [18] suggested including multimodal fusion systems to be able to provide an accurate assessment of PD states. The motion sensor data from multiple upper and lower extremities need to be collected and quantification methods based on the fusion of motion sensor data should be developed.

Based on the point mentioned, the following research questions related to methods for quantification of PD motor states during leg agility tests and quantification methods using multisensory data were identified and addressed as follows:

- What quantities should be calculated using motion sensor data collected from leg agility tests?
- Which features would be the most relevant measures to clinical ratings?
- How to develop methods based on extracted relevant features to quantify the PD motor states?

- Which machine learning method provides better validity to clinical scales?
- What are the properties of the generated scores from the developed method in terms of reliability, and responsiveness?
- At what level the collected data from multiple extremities should be combined?
- What are the important quantities when the fusion of the sensor data is going to be used for the development of the quantification methods?
 - What would be the validity results of the scores generated by machine learning methods, compared to the clinical rating scales?
 - What would be the other clinimetric properties e.g. reliability and responsiveness, of the methods?

Furthermore, with regard to the objective assessment of PD motor states, the thesis performed a study with collected motion sensor data and smartphone inertial measurement units (IMU) data during the standardized motor test of rapid alternating movements of hands. Hand rotation tests were previously quantified using collected data of the motion sensors which were comprised of gyroscope and acceleration sensors [19]. The methods provided good validity, reliability, and dose-effect time profile. However, using motion sensors has limits since they are used for passive data collection where the data need to be transferred to a processor for analysis, and to a screen for visualization of the results. These sensors are not widely and personally available in the market as ready to use devices and wearing them is not recognized as comfortable for some PD patients [20]. On the other hand, the acceptability and sustainability of the employed sensor devices in PD patients' life is an important matter [20]. In contrast, smartphones are publically available, include embedded memories and processors, and enable active data collection where the tests can be objectively and continuously performed. Plus the interaction of PD patients and clinicians becomes possible using this device [21]. Gyroscopes consume much more power than accelerometers and therefore the periodic and continuous sampling of gyroscope data for the purpose of objective measurement faces difficulties. IMUs from a smartphone are available and the application of smartphones is expanding quickly in many directions of human life. For the mentioned reasons, providing objective measures using the only accelerometer of a smartphone are highly desired.

The research question at this point is: how feasible would it be to quantify a standardized UPDRS test like rapid alternating movements of hands with an IMU accelerometer embedded in smartphones. For this purpose, the following questions are highlighted:

- How to compare the signals recorded by the motion sensor and the smartphone IMU accelerometer where rapid alternating movements of hands data was collected simultaneously?

- How similar would be the kinematic features extracted from these two sensors data?
- Using a validated machine learning method, will it be possible to identify the motor state of the subject who performed the tests when smartphone data was used?

Research approach

The above stated research questions require to express the dominant concepts of the research domain that is microdata analysis and the research approaches in order to meet the research objectives. Microdata were collected at different time points and different motor and medication states from PD patients together with the microdata of healthy individuals as a control group. The overall interpretive scheme of the thesis architecture is depicted in Figure 1. It is illustrated from bottom to top into four main methodological parts. Data preparation and processing, development of data-driven methods and evaluation of the methods.

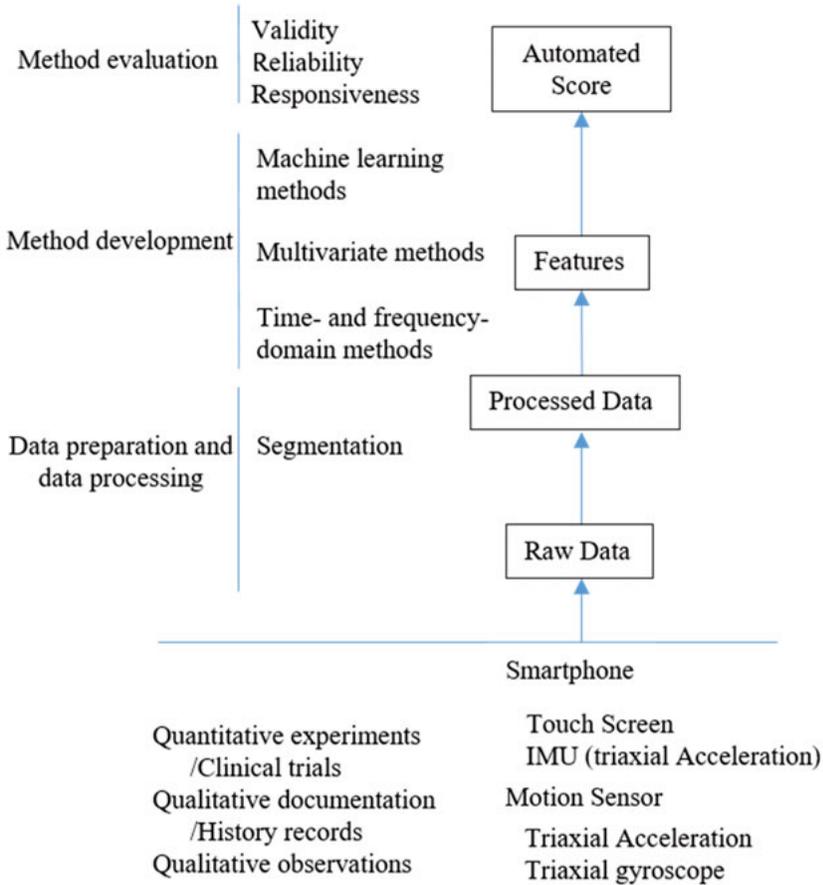


Figure 1. Interpretive scheme of the thesis.

At the lowest level, it is shown that the objective and subjective measures of fine motor and gross motor functions were gathered using the smartphone touch screen, smartphone IMU accelerometer, motions sensor triaxial accelerometer, and triaxial gyroscope. Data was collected quantitatively from repetitive experiments of the PD patients and healthy controls. The health-related history of the subjects were also included in the data together with an observational evaluation of the neurological experts who rated the symptoms of PD patients from different aspects of UPDRS, Treatment Response Scale (TRS), and dyskinesia rating scales. The preparation of the data included the motion sensors data segmentation and smartphone touch screen data type transferring e.g. identify objective movement series of data and handle the missing/duplicate entries; JavaScript Object Notation (JSON) data type transfer to string format, respectively. The processed data were then used in multivariate data analysis methods to extract spatiotemporal features. Multivariate

statistical methods involve more than two variables and permit simultaneous analysis of multiple dependent and independent variables [22]. Multivariate methods tend to analyze multiple measurements of individuals or objects under investigation [23]. Spatiotemporal features extracted by time-domain, wavelet-domain, and frequency-domain analysis methods included the statistical moments, time series trend and irregularity components, similarity measures and kinematic quantity measures. Features were used as inputs to machine learning methods and statistical models for modeling, regression, and classification of dependent (response variable) and independent variables (calculated features). As part of the analysis in the thesis (Paper I, III, IV, V, and VI) the importance of the features in relation to clinical experts scoring were investigated. The purpose of employing these approaches was to avoid overfitting and improve model performance. Two feature selection approaches employed in this thesis were stepwise selection [24] and lasso regression [25] models. There are basically two models of the methods, the knowledge-driven approach and the data-driven approach. As it is clear from their names, the knowledge-driven approach is based on the domain knowledge and derived from the facts, information, structure, and skills of the field experts about the problem which can include errors related to their measurements limitations. Whereas the data-driven approach is based on the analysis of the characters extracted from data of the problem. This approach is free from unrelated and intuitive or emotional assumptions to be made as the underlying process of the problem. In fact one important purpose of the data-driven methods were to model the relationship between a set of independent quantitative measures and the dependent variable that was obtained by clinical ratings. The selection of the type of the method depended to the desired type of outcome. For example regression (numeric prediction) was applied on numeric outcomes and classification was applied on nominal or ordinal outcomes. The data in this thesis was collected, accessed, processed and modelled with this approach. Some instances of such approach includes principal component analysis (PCA), factor analysis, statistical regression analysis, and artificial neural networks. PCA mathematically rearranges the variables into components on the basis of their variance. Components explain the variances whereas factor analysis explain the covariances in variables [26].

The focus of this thesis is the development and evaluation of data-driven methods to automate PD motor states assessments using data collected during a day from multiple extremities of PD patients and healthy controls. Together with the various types of motor symptoms in PD, the motor symptom fluctuations differ between and within PD patients. For this reason, data need to be collected accurately specific to objective and subjective fine motor and gross motor symptoms, and to be analyzed using data-driven methods incorporating the facts that are extracted from the data.

Given the fact that PD is multidimensional, multiple measures were gathered and calculated from multiple extremities on PD patients. In the assessment of fine motor symptoms data collected by smartphone, the positions extracted from coordinates of x and y and the time stamps, of spiral drawing and alternate tapping in millisecond were recorded. In the assessment of gross motor symptoms like walking, leg agility, and rapid alternating movements of hands, the triaxial coordination data of acceleration and gyroscope sensors as well as their time stamps in millisecond were collected. Understanding the natural phenomena of the motor symptoms from the data in that level was difficult e.g. finding the meaning of data recorded during different severity levels of PD patients only by looking at the x and y coordinates in tapping data. The data needed to be processed using mathematical equations to be meaningful and related to the motor functions. Even thereafter the calculated motor functions alone could not be related to the motor symptoms without incorporating the knowledge of domain experts. Therefore the developed methods were validated using specialists scoring of the motor symptoms that were formulated by rating scales.

The evaluation of the metrics from quantitative and automatic methods was done by assessing their validity, reliability, and responsiveness. Validity in simple words refers to how sound is the research and whether the findings from the developed method represent the phenomenon related to the problem. The type of validity evaluated in this thesis is the convergent validity, where the correlation of the scores from methods to the scores of the clinical ratings was examined (Paper I, III, IV, and V). Reliability refers to the consistency of methods outcomes over time, across items, and the consistency of the scores between researchers. In this thesis, three types of test-retest reliability, internal consistency, and inter-rater reliability were evaluated (Paper I, II, III, IV, and V). Sensitivity refers to the ability of the methods in detecting the changes in motor symptoms at different medication levels over time. Over the development of various data-driven methods in this thesis, the sensitivity of the produced scores was examined against the treatment interventions over time (Paper I, II, III, IV and V). In addition, part of the thesis covered the correlation of the method to other methods (Paper II) and the ability of a method in differentiating the PD patients at different stages of the disease with respect to the healthy controls (Paper III). This was done using a linear mixed effect (LME) models with restricted maximum likelihood estimation methods. LME model is a developed version of simple linear models with incorporating allowance of fixed and random effects. A fixed effect is a parameter that does not vary whereas the random effects are random variable parameters.

Background

Parkinson's disease

Parkinson's disease (PD) is a chronic degenerative disorder of the central nervous system and progresses slowly. Several factors contribute to the risk of developing PD. This disease is caused by biopsychosocial influences including genetic, nutritional, neuroanatomic, neurochemical abnormalities and dopamine deficiency in substantia nigra [27]. The substantia nigra is a basal ganglia structure located in the central parts of the brain (Figure 2).

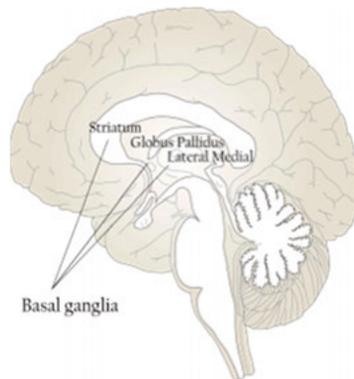


Fig 2. Illustration of brain structure related to Basal Ganglia.

The first medical description of the disease was given by James Parkinson in 1817 [28]. It is slowly progressing neurodegenerative condition of the disease affecting 1% of the population over the age of 60, and about 90% of the cases are sporadic [29]. In 2004, Germany contained the largest number of people with Parkinson's within Europe [30] and the cost of PD in Europe in 2010 was estimated for about 13.9 billion euro [31].

PD is characterized by a large number of motor symptoms that can impact on the function to a variable degree. The four cardinal symptoms of PD are tremor, rigidity, bradykinesia and postural instability [32] [33]. Tremor starts in one hand, foot or leg and eventually can affect both sides of the body. It is like oscillating movements and appears when a person's muscles are relaxed, disappears when the person starts an action. It's the most apparent well-known symptom. Rigidity causes stiffness of the limbs, neck or trunk and results in

inflexibility. Bradykinesia means slow movement and in general is the reduction of spontaneous movements like abnormal stillness and a decrease in facial expressivity. Bradykinesia causes difficulties with repetitive movements. It can cause walking with short and shuffling steps and can also affect the speech. Postural instability is a tendency to be unstable when standing upright, rising from a chair or turning [33]. In addition, among classic features of Parkinsonism, flexed posture and freezing are included with PD as the most common form [32].

Diagnosis of PD is done by clinically evaluating the presence of cardinal symptoms and based on the response to medication. Treatment of PD consists of Levodopa-Carbidopa oral tablets [34], continuous infusion pumps [35] and deep brain stimulation (DBS) [36]. Levodopa as a dopamine precursor is in fact considered as a gold standard oral treatment for PD [37]. In the early stages of PD Levodopa works well and helps the motor symptoms to improve. Over the progression of the disease and long-term therapy, the motor fluctuations and the disease complication appear. Fluctuations mainly appear due to the impaired absorption of oral levodopa therapy [38]. Motor states fluctuate between the two states of Off and On. Off state refers to when the medication concentration in blood is not enough causing the motor movements to fluctuate whereas the On state refers to a state that PD patients experience normal functioning with less disturbance. A side effect of long-term therapy is a Dyskinesia state in which PD patients experience uncontrolled movements and difficulties in performing voluntary movements [33]. Dyskinesia can occur both in response to excessive concentrations of medication in the blood and to decreasing concentrations when they are suddenly followed by severe Off. The quality of life of PD patients is affected by the above-mentioned difficulties.

Subjective assessment of PD

Basically, the need for a metric to assess therapeutic interventions was one of the key driving reasons for the development of rating scales. Evaluation of the severity of PD symptoms at clinical settings is done using two main subjective approaches. The first is based on observation and examination of the experts in the field using rating scales at clinics. The second is based on patient self-assessment and questionnaires.

For the first approach, there are UPDRS [39], Hoehn & Yahr scale (HY) [17], Treatment Response Scale [35] and dyskinesia scales. UPDRS is the most widely used multidimensional clinical scale for assessing PD motor impairment and disability. It mainly includes four chapters of mentation, behavior and mood (I), activities of daily living (II), motor examination (III), Complexions of therapy (IV). Chapter III which is the focus of this thesis requires the physical examination of PD patients. It contains 14 items, each of which

scored on a five-point scale ranging from 0 (normal function) to 4 (severe function). The items are speech (#18), facial expression (#19), tremor at rest (#20), action or postural tremor of hands (#21), rigidity (#22), finger taps (#23), alternate hand movements (#24), rapid alternating movements of hands (#25), leg agility (#26), arising from chair (#27), posture (#28), gait(#29), postural stability (#30) and body bradykinesia and hypokinesia (#31). In this thesis the data of #23, #25, #26, #27, #29 and #31 were included in the analysis. In addition, “SumUPDRS” score is a combined sum of the included items were used for presentation of total impairment. UPDRS scale is revised under the sponsorship of Movement Disorder Society (MDS) and the new version is MDS-UPDRS [2]. HY scale [17] is one common scale which is used to describe symptom progression of PD on a scale of 1 to 5. HY scale was modified by addition of two more stages 1.5 and 2.5 to take the intermediate course of PD into account. The higher stages of HY scale correlate with neuroimaging studies of dopaminergic loss, and some standardized scales of motor impairment. The weakness of the HY scale is mixing of impairment and disability and its non-linearity. The primary index of disease severity in HY scale is heavily weighted toward postural instability. For this reason, it does not completely capture impairments from other motor features of PD [40]. In this thesis TRS was used for evaluating overall mobility of PD patients [35]. It was used by neurological experts to score the physical examination of PD patients over the course of tests. TRS is on a scale of -3 to +3 and ranges from -3 (very “Off”, severe Parkinsonism) to 0 (On, normal mobility) to 3 (Severe dyskinesia, severe choreatic dyskinesia). In addition, Dyskinesia scale was used to score presence of dyskinesia on a scale ranging from 0 to 4 [2]. For the second approach, Parkinson’s disease questionnaire (PDQ-39) is the most widely used measurement tool used by PD patients. It assesses how often people affected by PD experience difficulties across some dimensions of daily living [41].

However, the mentioned clinical approaches for assessment of the severity of PD motor symptoms include limitations which are the reasons for objectively assessing the motor symptoms or combining the objective and subjective methods. Subjective and objective measures in PD complement each other, each of them has strengths and weaknesses [42]. In the section below, we describe the objective measurements as it is the focus of this thesis. Thereafter, we highlight the limitations in coming sections.

Objective assessment of motor symptoms

Objective motor measurement is the process of monitoring and measuring PD motor functions with repeated measurements and from different kinematic aspects. Since PD is a progressive degenerative neural disease it is necessary to

develop an objective measurement (long-term) of motor functions of PD patients. The characteristics of sensitivity, accuracy, portability, and objectivity [43] are needed in technology-based objective measurements. Technology-based objective measurements are the results of instrumented clinical tests conducted at clinics to objectively measure specific behaviors to detect impairments of PD [7]. In order to improve the accuracy of PD diagnosis, detecting early PD, quantifying symptom severity, the progression of PD and the treatment responses, several objective measures have been proposed.

According to Odin et al. [44], performing objective measurement is necessary. The therapeutic target is often insufficiently defined and can vary during the course of the disease. An objective measure is most valuable when a change in therapy can alleviate the complications. A natural point for objective measures guided therapy target in PD is the normal range in a healthy age-matched population. It is important that the objective measure can separate the controlled and uncontrolled movements. The objective methods should be validated [7]. There is a large number of regulations required for the instruments to be added and be accepted in a clinical tool [45].

In our previous review study, we listed some studies which objectively assessed cardinal PD motor symptoms [33]. Studies which assessed single symptoms [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58] and multiple symptoms [59], [60], [61], [62], [63], [64], [65] were included. It was highlighted that machine learning techniques are potential approaches for the development of assessment systems to determine the effectiveness of drug dosing. Machine learning methods that are employed in this thesis will be described later in detail. Tools that can effectively characterize the severity of symptoms and can discriminate between bradykinesia and dyskinesia are needed. Depending on symptoms and utilized sensors various features have been calculated in past studies (e.g. entropy, spectral or fractal features). Sensor signals were analyzed using various methods (e.g. time series analysis, regression or classification methods) in order to find the relation to a known sample of symptoms and/or predict the symptoms states. The employed sensors for estimation of the motor symptom states collected data such as time, acceleration, gyroscope, force, angle, video/image, and voice data. In a review study by Fisher et al. [20] that evaluated the acceptability of wearable sensors by PD patients, it was suggested to employ instruments that are convenient, publicly wearable, and could be used by PD patients at daily basis since the motor functions need to be measured continuously. These sensors do not impose limitations on the movements of PD patients. Instances of those instruments are Micro-Electro-Mechanical Systems, wearable motion sensors that could be mounted on wrist or ankle, and smartphones. This thesis focused on using motion sensors and smartphones.

The further stages of PD motor symptom objective measurement involved machine learning methods which were employed after extracting features and selecting the most relevant ones to the targeted symptoms/clinical rating. At

this stage, the relation between features and symptoms are estimated. As in many studies, there are multiple features extracted from sensor data. For example, spatiotemporal features of speed, amplitude, and irregularity of spiral drawing test when it was performed by subjects using a smartphone in paper I, or the number of foot taps, amount of displacement and speed of tapping in leg agility test performed by PD patients while wearing motion sensors in paper IV. The role of data-driven methods such as principal component analysis when there are multiple related features to motor functions is to provide components which principally include the largest amount of feature information to be used in further analysis. The features could also be used in data-driven approaches of machine learnings as inputs for regression analysis or classification. Principal components and regression analysis are used for relating the sensor data to clinical scores. This is done for estimation of the validity of the scores derived from sensors data to clinical ratings that are in fact based on the experience and observation of the neurological experts. Whereas classification can be done to classify groups of data in relation to motor symptom states like On and Off states or estimation of separation ability of the methods when the data consists of data from PD patients and Healthy controls.

Despite all proposed objective measures, the proposed methods for PD have not widely been available, adopted in clinical tools, or clinically used. Therefore a combination of objective measures, sensor data that is collected from multiple body segments of the PD patients might add more symptom information. This was investigated in part of this thesis. The ultimate goal of our research is to individualize treatments in PD. In order to do so, first, the methods need to be developed to quantify the motor states of the PD patients and test their clinometric properties. In the next chapters we describe the data and subjects together with smartphone and motion sensor data analysis performed in this thesis. Thereafter the employed methods and their clinometric properties estimation will be discussed.

Data collection and preparation

Subjects and experiment

The main dataset included data from recruited 19 PD patients and 22 healthy controls in a single center, open label, single dose observational clinical study. The clinical trial performed in a hospital Uppsala, Sweden [66], and written consent was given. The study was approved by the regional ethical review board in Uppsala, Sweden (reference number: 2015/100). Out of 19 PD patients eighteen patients experienced wearing off fluctuations and 13 of them experienced dyskinesia and responded with dyskinesia [66].

Patients were administered a dose of 150% of their individual levodopa-carbidopa equivalent morning dose. It was in order to induce dyskinesia (denoted as levodopa challenge in the text). Standardized motor tests according to the motor section of the UPDRS-III was performed. The tests comprised of rapid alternating movements of hands, reading a text, finger tapping, leg agility, and walking which were performed at different time intervals. The time intervals were once within 50 minutes before taking the dose, once at the time of dose administration (0 min), and then approximately at 20, 40, 60, 80, 110, 140, 170, 200, 230, 260, 290, 320 and 350 min after dose administration to follow the individual response of PD patients to their morning doses from Off motor state to good mobility and/or dyskinetic state, and regressing back to the Off state. Each patient performed the tests as long as they could, up to 15 trials. The characteristics of subjects who participated in data collection are shown in Table 1.

Table 1. Characteristics, mean (standard deviation) of patients and healthy subjects. In Gender: M is male, F is Female; PD side (most affected side): R is right, L is left, and nk is not known. HY is Hoehn and Yahr stage.

	Gender	Age (years)	Height (m)	Weight (kg)	Years with PD	Years on levodopa	PD side	HY	UPDRS IV
PD patients	14 M 5 F	71.4 (6.3)	1.75 (0.09)	75.4 (11)	9.7 (6.8)	9.5 (6.5)	9 R, 8 L, 2 nk	3.1(0.8)	6.21 (3.13)
Healthy controls	16 M 6 F	64.2 (7.4)	1.75 (0.1)	83.6 (13.8)	-	-	-	-	-

From Table .1 the studies in this thesis included all or part of the data for analysis. Three papers included the data of 19 PD patients and 22 healthy controls. Paper I and III where fine motor tests were collected by smartphone. Paper V where data were collected from multiple body segments by multiple motion sensors. During the study in paper IV, the thesis focused on analysis of the data of 19 PD patients only.

The second dataset was based on data from two clinical studies, both of which were approved by the relevant agencies and informed consent was given. 98 PD patients at different stages of PD and 10 healthy subjects were recruited for data collection. Out of 98 PD patients, 68 of them were advanced PD patients who were recruited in an open longitudinal 36-month study at nine clinics in Sweden [67]. 35 of these patients were treated with levodopa-carbidopa intestinal gel infusion (LCIG) and 30 patients were candidates for switching from conventional oral PD treatment to LCIG. In the latter group, the patients were LCIG treatment-naïve at study start.

In the second study, 38 patients with a clinical diagnosis of idiopathic PD in Milan, Italy participated [9]. These patients were divided into two groups: intermediate stage patients experiencing motor fluctuations ($n = 16$) and clinically stable, early PD patients ($n = 17$). Characteristics of the patients and HE subjects are shown in Table 2. This data set was used in paper II and III.

Table 2. Characteristics of PD patients and healthy elderly participants, presented as median \pm interquartile range.

	Subjects			
	Swedish study (Advanced PwPD)	Italian study (Intermediate PwPD)	Italian study (Early stage PwPD)	HE
Patients (n, gender)	65 (43m;22f)	17 (13m, 2f)	16 (13m,2f)	10 (5m;5f)
Age (Years)	65 \pm 11	65 \pm 6	65 \pm 6	61 \pm 7
Years with Levodopa	13 \pm 7	7 \pm 8.5	5.5 \pm 6	NA
Hoehn and Yahr stage at present	2.5 \pm 1*	2 \pm 0**	2 \pm 0.5	NA

* Assessment performed in the afternoon. ** Assessments performed in the On state. Abbreviation: HE, healthy subjects; NA, not applicable.

The third dataset was based on motions sensor and smartphone IMU data collected from a healthy subject during hand rotation hand movements. There were 12, 12, 13, 35 and 10 tests performed during 5 days and the tasks were repeated three times. In total there were 246 trials collected from each device. This data set was used in paper VI in addition to the main dataset.

Sensor measurements

During data collection by smartphone on each test occasion, subjects performed two upper limb fine motor tests consisting of alternate tapping and spiral drawing tests. Subjects were first instructed to sit on a chair and perform the test using an ergonomic pen with the device that was placed on a table. They were instructed to use neither hand nor arm for supporting the performance. The tapping test was shown to subjects on the screen of the smartphone and they were asked to alternately tap two fields as fast and accurate as possible using first right hand and then left hand. The time to complete a tapping test was 20 second. During the spiral test, subjects were instructed to trace a pre-drawn Archimedes spiral as fast (within 10 s) and accurately as possible, from the center out, using the dominant hand. The test was repeated three times per test occasion. The smartphone device had a 4" (86 × 53 mm) touch screen with a 480 × 800 pixels and recorded both position (x and y coordinates) and time-stamps (in milliseconds) of the pen tip. The device sampling was event-based, which means signals are sampled only when certain predefined events occur. In this case, a sensor event was generated every time the sensor values x and y were changed.

Data collection from smartphone and motion sensors were done during hand rotation tests which were investigated in paper VI. The collection of motion sensor data is explained in the next paragraph. The 3D acceleration data from smartphone IMU were collected after a clock drift correction. The IMU data was collected using "Sensor Kinetics Pro" application in a smartphone. The smartphone sampling rate was 100 Hz, with an accelerometer range of +/- 2 g.

Acquisition of motion sensors data was done from different motor tests such as leg agility test, rapid alternating movements of hands test and combination of leg agility, rapid alternating movements of hands, and walking tests. Motion sensors were worn on both wrists and ankles. During leg agility tests patients were instructed to sit on a straight-back chair and place both feet comfortably on the floor and then to raise and stomp each foot on the floor 10 times as fast as possible. They performed the test first with the right foot and then with the left foot. During rapid alternating movements of hands tests, subjects were instructed to sit on a chair with no armrest and perform forearm hand rotation motor task for 20 sec first with the right hand then and left hand. During walking tests, subjects were asked to walk for about 4 meters at a self-

selected pace using their typical way of walking. First, they walked for a few steps towards their right, then they turned around and walked back and forth. During the tests, subjects had 4 motion sensors on both lower and upper limbs. They were instructed to letting their arms comfortable during walking. Since the asymmetrically reduced arm swings contribute significantly in neurological gait disorder, the data from arm swings were also collected. Each sensor (Shimmer3) consisted of 3-axial accelerometers and gyroscopes (sampling rate of 102.4 Hz, accelerometer range of +/-16g and gyroscope range of +/-2000 dps). With the clock drift correction the accuracy was remained over long recordings. The sensor data of all time points (x, y, and z-axes of accelerometers and gyroscopes) was saved on the SD cards of the sensors and processed offline. With respect to the position of the subject, x-axis represented left and right, y-axis was back and forth, and z was the vertical movement.

Each test occasion was video recorded and timestamps of the sensor data were synchronized with the time points of the videos. The videos were randomly presented to experts for clinical ratings. The randomization of the videos was to ensure the ratings of the patients were regardless of the time the dose was administered.

Methods

This chapter summarizes the methods that were used in this thesis for data analysis and method developments. Based on the type of data and the research question, the methods were selected. The development of the methods incorporated time-domain, frequency-domains and statistical measures in order to extract information from raw data. Statistical methods were employed to analyze the importance of the extracted quantitative measures which were used in data-driven methods to be mapped to clinical measures. Additionally, the statistical methods were used for evaluating the methods and their clinimetric properties.

Features were extracted from time series of data; that is, the ordered sequence of values of variables over time. Selecting a method for analyzing time series should be according to their statistical components. The methods employed for extracting features from raw data were spatiotemporal which means shape (location), frequency and time domain of the target have been taken into account.

During some studies of the thesis (Paper I, II and III) the quantification and analysis of fine motor movements were investigated using a smartphone. During the second part of this thesis (paper IV and V) the quantification of motor functions was done using motion sensor data of upper and lower body segments. Plus in one study (paper VI) the analysis of hand rotation tests was done using both smartphone and motion sensors data. Features in relation to the data were calculated to produce meaningful quantitative measures representing the severity of the motor symptoms and fluctuations of those symptoms. There were multiple features extracted in each paper. Statistical moments (mean, variation, trend, skewness, minimum, maximum and kurtosis) were applied on different quantities such as location, time, frequency and mathematically calculated measures from the data. We describe the main approaches used for the development of the methods including extraction of the features or measures for analysis.

Approximate entropy

There are chaotic behaviors present in clinical time series which are not detected by statistical moments like mean and variance, or with rank order statistics like the median. In order to quantify the amount of irregularity in time

series of data in this thesis, the approximate entropy (ApEn) method was employed [68]. This method as a statistical and a nonlinear measure reflects the similarity between a chosen window of a given duration and the next set of a window of the same duration. The approximate entropy formula is presented below:

$$ApEn(m,r,N)=\frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\log C_i^m(r)-\frac{1}{N-m}\sum_{i=1}^{N-m}\log C_i^{m+1}(r) \quad (1)$$

Where m is the length of the window being compared, r is filtering level, N is the total number of data samples and $C(r)$ is correlation integral. The two parameters of m and r remain fixed during the calculation.

This method was employed in the paper I for feature extraction. The measure of irregularity for the time series of tapping signal, tapping speed, spiral drawing speed, and spiral drawing radial velocity was calculated.

In paper II this method was developed using the spiral drawing data from the second dataset from PwPD and compared to two other methods which presented the drawing speed and radial velocity.

In paper III this method was verified using the main dataset and further investigated in spiral drawing tests that were performed by PD patients and healthy controls. The amount of irregularity in drawing speed patterns within the drawing speed signals of the spiral drawings was quantified. The clinimetric properties of the extracted temporal irregularity score were investigated.

In paper IV where the leg agility test was analyzed, two features were calculated by application of this method on two time-series which were the magnitude of acceleration and orientation data.

In paper V the amount of irregularity was measured for magnitude acceleration and orientation time series data of tests including alternate hand movements, leg agility, walking, and arm swing. In addition, during alternate hand movement tests the irregularity measure was calculated for acceleration and orientation time series of x , y , z coordinates ($Xacc$, $Yacc$, $Zacc$, $Xgyr$, $Ygyr$, $Zgyr$).

Paper VI included the study of the acceleration time series data during alternate hand rotation tests. Similar irregularity features to paper V for alternate hand movements time series were calculated excluding the orientation data.

Discrete wavelet transform

Wavelet and time series of the clinical domain are complex and nonstationary. Like many real-world problems they contain low-frequency components occurring for long durations as well as high-frequency components occurring for short durations. Multiresolution decomposition methods are able to separate the components of a signal by iteratively breaking the time series into different independent scales [69]. This process is a so-called “mathematical microscope” where the focus on different parts of the signals can be adjusted for

extraction of the sub-components from the time series [70]. A multiresolution analysis approach is the Dynamic Wavelet Transform (DWT). This method resamples the wavelets in a discrete manner and decomposes the signals into smaller components. It has a key advantage over the Fourier transform which is temporal resolution. It means it captures the frequency changes and time location information. DWT adjusts the resolution based on the frequency components. The low-frequency consists of narrow frequency resolution and coarse time resolution. On the contrary, the high-frequency analysis consists of coarse frequency resolutions and narrow time resolution. DWT provides two series of high-pass and low pass filters using the most common set known as Daubechies wavelets [71]. In fact, the time series signal is decomposed into two wavelet coefficients. One for low-frequency components that is the coarse approximation and another is the high-frequency component or details. The first represents the main features of the signal whereas the later represents the short and fast fluctuations. Coefficients for low-frequency components are further decomposed into two levels of high and low-frequency components. From all decompositions, low-frequency components wavelet coefficients are the results of the multiresolution analysis. For the above-mentioned theory the frequency of the original time series can be assumed to be 100 Hz. Then high-frequency bandwidth from the first decomposition would be 50-100 Hz, the second high-frequency from the second decomposition would be 25-50 Hz. The high and low-frequency bandwidths after the final decomposition would be 12.5-25 Hz and 0-12.5 Hz.

In this thesis, DWT was applied on time series data of different motor tests. In the paper I DWT was applied on time series of spiral drawing signals. DWT decomposed the spiral drawing signals in order to extract features related to both the overall severity level of the motor symptoms as well as fast the displacements in relation to time and location.

In paper IV the time series of *Macc* and *Mgyr* from leg agility test data was decomposed into three levels. The first level consisted of low-frequency components of 0-25.6Hz and high-frequency components of 25.6-51.2Hz. After decomposing the low-frequency components of the first level, the second level consisted of low frequency components of 0-12.8Hz and high-frequency components of 12.8-25.6Hz. Finally, the third level signals were further decomposed to the low (0-6.4Hz) and high (6.4-12.8Hz) frequencies. Mean and standard deviation of *Macc* and *Mgyr* were calculated for third level high-frequency components to generate four features.

In paper V when motor symptoms of PD were quantified using the fusion of motion sensor data, the same features as in paper IV were calculated for leg agility test. Application of DWT on time series of *Xacc*, *Yacc*, *Zacc*, *Xgyr*, *Ygyr*, *Zgyr*, *Macc* and *Mgyr* collected from alternate hand movement tests provided the 1st, 2nd and 3rd levels of high-frequency components. Mean and standard deviation of high-frequency components in all three levels resulted in 48 features. In the calculation of features for walking tests, three-level DWT

was applied on *Macc* and *Mgyr* of signals of walking. Mean and standard deviation of third level high-frequency of *Macc* and *Mgyr* resulted in four features. Similarly, DWT was employed on *Macc* and *Mgyr* time series of arm swing data to produce similar features to ones extracted from the walking test.

Principal components analysis

Multivariate analysis often involves the analysis of data with substantially correlated variables. Principal component analysis (PCA) as a type of multivariate analysis method is a statistical procedure that transforms a set of possibly correlated variables into a set of linearly uncorrelated variables. The transformation is defined so that the first principal component contains the largest variance present in data. PCA is used widely in different areas of predictive modeling, dimension reduction and exploratory data analysis. A component produced by this approach has a corresponding eigenvalue and eigenvector. An eigenvalue determines the magnitude of the new feature space whereas the eigenvector determines the direction of it [72].

Fitting models and interpreting results of data analysis in data-driven models can be problematic because of collinearity which refers to the correlation between variables in a dataset. For this reason, using PCA is preferred since the components are uncorrelated to each other. The correlation matrix method was used for producing the principal components (PC).

After the extraction of principal components, the key point is to determine how many components should be retained for further analysis. One common approach is the Kaiser-Guttman criterion for factor retention [73] which is based on the distribution theory of eigenvalues. Using this criteria PCs with eigenvalue higher than 1 are retained for further analysis. There is another rule which is based on the cumulative percentage of total variation. Based on this rule the retained PCs total variation should be above 70% or 90% defined as a suitable cut-off. There are two other methods which look into changes in the behavior of the plot of the variance explained [74] [75].

Kaiser-Guttman approach was employed for the PC retention approach in this thesis. It is highlighted that this method is a promising and powerful retention method when it is also easy to compute and visualize [76].

In paper I, the quantification of dexterity in PD during spiral drawing and tapping tests, PCA was used to reduce the dimension of the feature set. After computing PCs and application of retention criteria, there were 7 PCs retained which accounted for 71% of the variation in data. Responsiveness of PCs to treatment outcomes was calculated and compared to the responsiveness of the clinical ratings.

In paper IV when quantification of motor symptoms from leg agility test data was performed, PCA was used for the same purpose and 5 PCs were retained at the end which totally explained 84% of the variances in data. The

retained PCs were used in predictive modeling. In addition, to assess the differences in motor test results between the right and left legs of PD patients, t-tests were performed on the first PC of the extracted features.

In paper V PCA was used as an unsupervised method and most amount of information from the feature set was retained. The retained PCs were used in different machine learning methods to map the scores to clinical ratings.

In paper VI PCA was applied on feature sets which were extracted from alternate hand movement test using instruments of smartphone and motion sensor. The eigenvectors comprising the coefficients corresponding to each feature were calculated. In order to estimate the importance of the features, after computing the components, the larger absolute value of the coefficient corresponded to the more importance of the feature.

Linear mixed effect modelling

A mixed model consists of two effects, random and fixed effects. Mixed effect modeling is useful in settings that contain repeated measurements. Mixed effect models fit well with longitudinal studies where repeated observation of the same variables is collected over a short or long period of time. E.g. repeated objective measures that are collected at multiple time points of the day, or the effect of the treatment that is embedded in the progression of the PD disease. Mixed effect models are preferred to repeated models like ANOVA because they take care of missing values. This method is needed in order to derive valid and reliable conclusions from data where the study design is unbalanced. For example, the amount of variability within subjects or within groups need to be taken into account with such models.

Linear mixed effect model (LME) is an instance of the mixed effect model [77]. LME incorporates the variations within and between subjects which is called random effect. The inclusion of random effects are to present the effects of subjects that are not reflected by fixed effects. LME was used in paper II and III to assess the significance of the produced score among four subject groups of healthy, early, intermediate and advanced patients. Restricted maximum likelihood estimation which estimates the precision and inference of the effects based on their asymptotic distribution rather than estimating their maximum likelihood [78].

Cross-correlation

In time series analysis and signal processing cross-correlation is a measure of similarity which reflects the displacement of one signal to other [79]. For this, assuming $f(t)$, and $g(t)$ to be two series, the cross-correlation is defined as below.

$$(f * g)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{f(t)} g(t + \tau) dt \quad (2)$$

Where $\overline{f(t)}$ is conjugate of the signal $f(t)$, τ is the lag (displacement). Using cross-correlation one signal is shifted along the other signal until the best match is found. The formula slides the $g(t)$ series along the x-axis by calculating the integral of the product at each position. The value of $(f * g)$ will be maximized when the two series match. When the positive and negative areas align a large contribution due to the integral of the product. In Paper VI, the cross-correlation method was used to measure the similarity of the time series collected by smartphone Inertial Measure Units (IMU) and wrist motion sensors during hand rotation tests. The measure of cross-correlation between the signals was done at lag of zero which means there was no sliding by either of the signals.

Linear regression

Linear regression (LR) is an approach to model the relationship between a dependent variable and an independent variable. When there is more than one independent variable, the process is called multiple linear regression. With the goal of predicting the outcomes, linear regression is used to form a predictive model to an observed data set of values of the response and independent variables. After developing this model, given the additional values of the independent variable without the accompanying response value, the fitted model can be able to predict the responses. In this thesis, this model was used for prediction purpose.

Linear regression model assumes that the relationship between the dependent and independent variables is linear. The model takes the below form:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (3)$$

Where y is the dependent variable, x is the independent variable, α is the intercept, β is the slope of the line, and ε is the error term.

In paper I, linear regression was used as one of the four models to map the independent variables to the five clinical rating scales of UPDRS #23, UPDRS #25, UPDRS #31, SumUPDRS, TRS and dyskinesia. Independent variables were the combination of the extracted features from raw spiral drawing and tapping data in a smartphone.

In Paper IV LR was used to investigate the linear relationship between the variables and targets. Using LR the scores of four rating scales including TRS, UPDRS #31, SumUPDRS, and Dyskinesia were predicted where features of motion sensor data collected from gait were used as input.

In paper V, the LR modeling was used to predict the TRS ratings by mapping the fusion of the extracted features from upper and lower body segments to the TRS rating scale.

Stepwise regression

Stepwise regression is a statistical method for fitting regression models to predictive variables for finding the best subset in order to enhance prediction accuracy. The procedure takes steps in selecting the variables to be included in the final subset. At each step, a variable is considered to be added or subtracted based on their significant effect in the model fit. There are different approaches to selecting the variables, including forward selection, backward selection, and bidirectional elimination. In this thesis, forward selection has been the main approach for the selection of the features. It starts first with including no variables in the model and then it tests the addition of each variable by using a model fit criterion which is based on regression analysis. The inclusion of the variables should most significantly improve the fit. After a new variable is added, there is a test to check if some variables can be deleted without significantly increasing the residual sum of squares (error). The process is repeated until there is no variable can be added or removed.

This method was used in the paper I, paper IV, and paper V for selection of the most important features in relation to the clinical ratings.

Least absolute shrinkage and selection operator

Least absolute shrinkage and selection operator (Lasso) is a regression analysis method for selecting most important variables (features) in order to improve the accuracy of the prediction and increase the interpretability of the model with regard to the response. It alters the model fitting process to select a subset that enhances the prediction to provide the highest possible accuracy.

It uses a penalty for both fitting and penalization of the coefficients by putting a constraint on the sum of the absolute values of the model parameters. The sum must be below a fixed upper bound value. For this, lasso works with an objective function for regularization (shrinkage). It penalizes the coefficients of the regression variables to shrink some of them to zero. The objective function is as below:

$$\frac{1}{N} \sum_{i=1}^N f(x_i, y_i, a, b) \quad (4)$$

Where y_i is the outcome and x_i is the covariate vector for the i^{th} case. b is the standard norm [25] and the objective of the lasso is to minimize the objective function [25]. This method was used in paper V for selection of the most important variables with regard to TRS.

Decision trees

A decision tree (DT) is a classification modeling of the data which works as a hierarchical tree where the nodes are questions about the associated features with the data items. In the simplest form, each node is a question of a yes or a no, where each answer points to a child node presenting a new question or it's a final node called leaf. Passing over the nodes, children and reaching to the leaves at the end the conditional probability of data belonging to classes are estimated based on the probability distribution of the defined classes [80]. The leaves correspond to the weights of the classes of interest. The calculated weight assigned to each class is associated with the predictions to be made.

In paper I this method was used as one of the four methods to classify the features extracted from spiral drawing and tapping data collected from smartphones to the five clinical rating scales of UPDRS #23, UPDRS #25, UPDRS #31, SumUPDRS, TRS and dyskinesia. The UPDRS scales contained four classes which ranged from 0 (normal) to 4 (severely impaired). TRs contained seven classes of -3 (very off), -2, -1, 0 (on), 1, 2, 3 (very dyskinetic). Dyskinesia presented four classes ranging from 0 (normal) to 4 (very dyskinetic). Decision trees were examined as a method in paper IV in addition to two other machine learning methods to investigate the targets and their relation to the variables extracted from gait sensors data in paper IV. Paper V similarly used this method to classify the fusion of features extracted from the data of multibody segments. There were considered seven classes from the TRS clinical rating scores ranging from -3 to +3.

Multilayer perceptrons

A multilayer perceptron (MLP) is a feed-forward artificial neural network getting a set of inputs and generating a set of output. It consists of layers of nodes for the input layer, hidden layer, and output layer. MLP is a supervised learning method for distinguishing nonlinearly separable data using backpropagation for training. It uses an activation function which maps the weighted inputs to the output of neurons. The activation function can be linear or nonlinear and the two common ones are sigmoid. Node weights are adjusted based on corrections that minimize the error in the output. MLP was used in paper I to use extracted features calculated from different smartphone data as inputs, and

the rating scales as targets which makes the process as a supervised learning procedure.

Support vector machines

Support vector machine (SVM) learning method is a common supervised learning method used for classification and regression associated with nonlinearly related data. It classifies the data by finding a hyperplane that maximizes the distance between classes of data. Support vectors are in fact the vectors that define the margin between classes. An objective function is defined to separate the classes and it is maximized to make a hyperplane with the largest possible margin. Since in this thesis the data is highly nonlinear, SVM handles the classification by using a kernel function to map the data into different spaces where a hyperplane cannot be used for separation. The kernel function transforms the data into a higher dimensional feature space for making it possible to perform the linear separation [81]. In this thesis, high dimensional feature sets are calculated. SVM is used for regression analysis of data collected from smartphone screen (paper I), smartphone IMU (paper VI), and motion sensors during gait test (paper IV).

In paper I, SVM was used to predicting the motor states of PD patients. The combined feature set calculated from spiral drawing and tapping features were used as inputs to SVM and the five rating scales were set as the target labels. 10-fold cross validation method where 9 folds are used for training and 1 fold is used for testing the model were employed. It was performed within a loop of 100-125 times. This was to evaluate the performance of the SVM. SVM automated the assessment of motor dexterity by predicting the rating scores using the data from the smartphone screen. The procedure of predicting the rating scores was similar to what was done in paper I, except the targets which were different in other papers. In addition to the cross-validation method in paper V the method of Leave-1-Out cross validation was examined. In this method, considering the data of 10 PwPD, the parameter optimization is performed on 9 of the 10 PwPD and then the performance of the algorithm is tested on the 10th PwPD. At this step, the 10th pair is the test set and the other nine pairs are the training data. This process is repeated for 10 times, each time leaving out a different pair to be used in the test set. Data were left out of the training set and then after the training procedure, it was used as a testing set.

Results

Paper I

Dexterity in PD was quantified after extracting 37 features from spiral drawing and tapping tests data collected by smartphone and using machine learning methods. Machine learning methods MPL, RT, LR, and SVM methods were employed. SVM yielded higher correlations to clinical ratings compared to other methods. The correlations between smartphone-based scores and mean ratings were 0.52 (UPDRS item #23; finger tapping), 0.47 (UPDRS #25; rapid alternating movements of hands), 0.57 (UPDRS #31; body bradykinesia and hypokinesia), 0.46 (sum of the three UPDRS items), 0.64 (dyskinesia), and TRS (0.59). When assessing the test-retest reliability of the scores it was found that, in general, the clinical scores had better test-retest reliability than the smartphone-based scores. Only the smartphone-based predicted scores on the TRS and dyskinesia scales had good repeatability with intra-class correlation coefficients of 0.51 and 0.84, respectively. Smartphone-based scores differed significantly between patients and healthy controls.

In estimating the effect sizes to evaluate the responsiveness of the smartphone-based scores, clinician-based scores had higher effect sizes indicating better responsiveness in detecting changes in relation to treatment interventions. However, the first PC of the 37 features was able to capture changes throughout the levodopa cycle and had trends similar to the clinical TRS and dyskinesia scales. The responsiveness of the PC scores vs clinical scales is depicted in Figure 3.

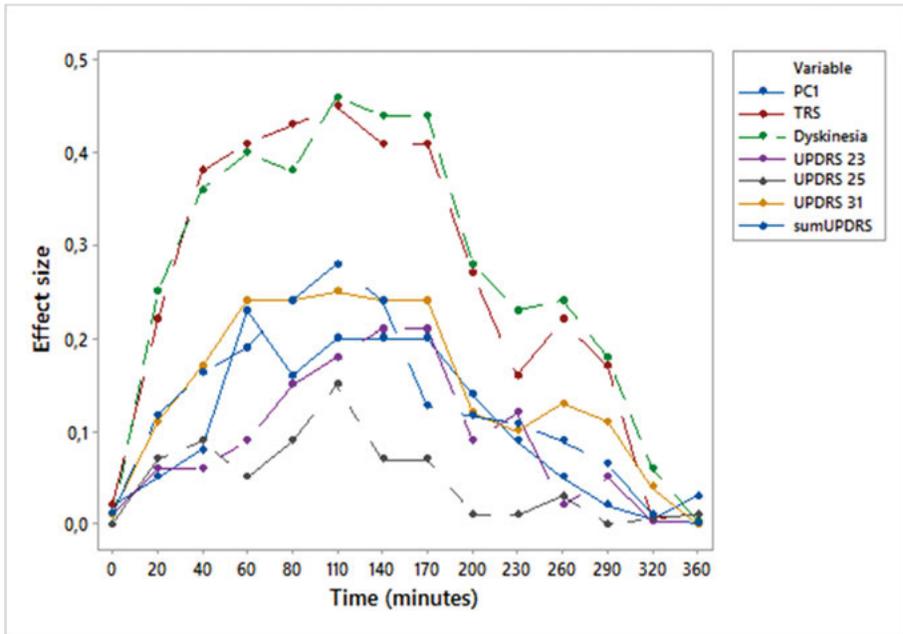


Fig 3. Sensitivity assessment of PCI and mean ratings of the three movement disorder specialists on the three UPDRS items, TRS and dyskinesia across the levodopa test cycle for all patients. The first data point in the X axis represents the change in scores between the first two baseline (without medication) measurements. The second data point represents the change in scores between first baseline and third measurement, and so on. Number of tests per time slot: 0 (n=19), 20 (19), 40 (n=19), 60 (n=19), 80 (n=18), 110 (n=17), 140 (n=17), 170 (n=17), 200 (n=17), 230 (n=17), 260 (n=14), 290 (n=14), 320 (n=11), and 360 (n=11).

Paper II

The responsiveness of an ApEn based score was compared to the results of two methods which the first was based on the drawing speed and the second was based on the radial velocity of spiral drawings. The method based on ApEn generated a temporal irregularity score (TIS). The score from the method based on spiral drawing was called SD-DV and the score from the method which was based on radial velocity was called WSTS (Wavelet spiral test score). Comparing the mean scores of the three methods between four subject groups of healthy, early, intermediate and advanced, the mean TIS score was significantly different between healthy subjects and advanced PwPD ($P < 0.005$). In contrast to TIS and WSTS, mean SD-DV scores were not significantly different across the groups. There were correlations between WSTS and SD-DV scores ($r = 0.69$, $P < 0.001$). TIS was correlated to neither WSTS nor SD-DV scores. Classification of healthy controls and PD patients was done by linear regression analysis after combining WSTS and TIS. The

accuracy was 85% and the weighted AUC was 0.89. Calculated effect sizes for TIS were higher than WSTS and SD-DV which indicated greater responsiveness than the two other scores. In addition, TIS had higher test-retest reliability (0.9) compared to the other two scores.

Paper III

Using the main dataset in this paper the clinimetric properties of the ApEn method was further investigated. A temporal irregularity score (TIS) was produced. When assessing the differences of TIS between tests performed by PwPD at different PD stages and healthy controls, the values were found to be significantly different between healthy subjects and advanced PwPD (p-value = 0.02), similar to the results in paper II (p-value < 0.05). They were not significantly different between the healthy subjects and the whole patient group (p-value = 0.07). There was a significant difference (p-value = 0.00) in age between the healthy controls (mean = 64 years) and the whole patient group (mean = 71 years). The results for differences of mean TIS when considering the patient groups as the main effect and the age as a covariate are shown in Figure 4.

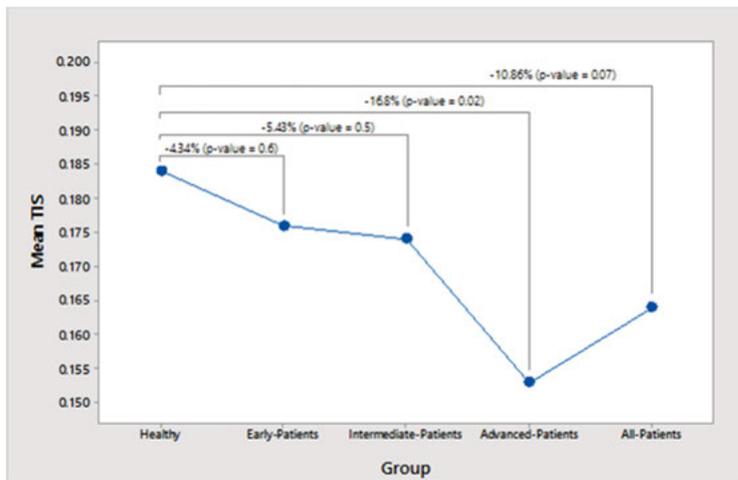


Fig 4. Linear mixed effect (LME) fixed effects coefficients of the temporal Irregularity Score (TIS) (with age as covariate) across the five subject groups. P-values (groups: Early, Intermediate, Advances, and All patients) with respect to healthy subjects were: 0.6, 0.5, 0.02, 0.07. Number of participants: Healthy (n=22), Early (n=7), Intermediate (n=8), Advanced (n=4), All patients (n=19). Number of observations: Healthy (n=176), Early (n=93), Intermediate (n=90), Advanced (n=57), All patients (n=240).

There was no significant difference (p -value = 0.52) in gender distribution between healthy controls and patients. The test-retest reliability of TIS was 0.81 (ICC), which was similar to the results in ST1 ($R = 0.74$) and indicated the consistency of TIS between the three spirals.

When assessing the sensitivity of TIS to the treatment response for patients, the results indicated that the patients' scores can capture some effects of the medication from off to on and wearing off effects (Figure 5). For healthy subjects, the effect sizes were smaller than the effect sizes of patients.

The correlations between TIS and mean clinical rating scores (UPDRS, Dys) were weak: -0.18 for UPDRS item 23 (finger tapping), -0.11 UPDRS item 25 (rapid alternating movements of hands), -0.29 for UPDRS item 26 (leg agility), -0.24 UPDRS item 27 (arising from chair), -0.20 for UPDRS item 29 (gait), -0.10 for UPDRS item 31 (bradykinesia), and -0.31 for Dys (Dyskinesia).

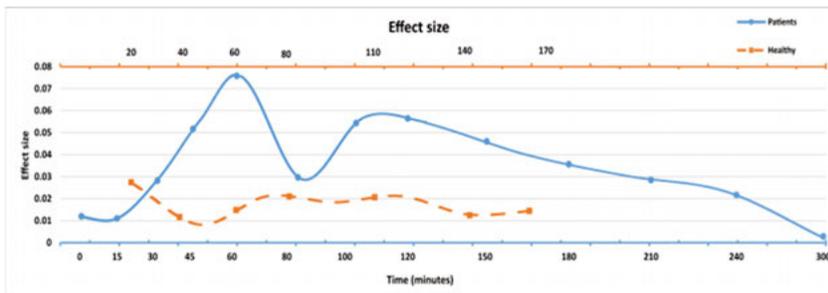


Fig 5. Sensitivity assessment of TIS across the test occasions. The lower x axis represents the minutes after taking the levodopa dose for patients and the upper x axis represents the tests time points for healthy subjects. The first data point in the x axis represents the difference in scores between first (baseline) and second measurements; the second data point represents the difference in scores between first and third measurements, and so on. Number of patients for periods: 0 ($n = 19$), 15 ($n = 19$), 30 ($n = 19$), 45 ($n = 19$), 60 ($n = 18$), 80 ($n = 18$), 100 ($n = 18$), 120 ($n = 18$), 150 ($n = 18$), 180 ($n = 17$), 210 ($n = 15$), 240 ($n = 13$), 300 ($n = 9$). Test 15 at 360 min contained only one patient, which was not enough to be included in this analysis. Number of healthy subjects for periods: 20 ($n = 22$), 40 ($n = 22$), 60 ($n = 22$), 80 ($n = 22$), 110 ($n = 22$), 140 ($n = 22$), 170 ($n = 22$).

Paper IV

Data-driven methods were developed and their clinimetric properties were examined for detection and quantification of PD motor states using motion sensors data of leg agility tests. 24 features were extracted and their relevance to clinical ratings was investigated. Most of the features were selected as relevant

predictors for all four scales. Six features were not selected at all by the regression models. Those features were:

Feature #9 and #10 (Mean of the third level high-frequency components after applying DWT on magnitude of acceleration and orientation)

Feature #15 (Number of peaks in magnitude acceleration)

Feature #16 (Standard deviation of peaks of magnitude of acceleration)

Feature #17 (Slope of the regression line calculated for peaks magnitude of acceleration over time)

There were four features which were selected in the four regression models, including:

Feature #2 (Mean magnitude of orientation)

Feature #6 (Skewness of magnitude of orientation)

Feature #8 (Maximum magnitude of orientation)

Feature #22 (Mean of displacement)

Where features #2, #6, and #8 were based on gyroscope signals and #22 was based on accelerometer signals. When investigating the distribution of the features in the individual models it was noticed that both gyroscope- and accelerometer-related features were equally important as predictors of the 4 clinical scales. From the PCA, 5 PCs were retained accounting for 84% of the variance in the data.

After employing machine learning methods (SVM, LR, and DT) on the selected features by stepwise regression and PCA, the absolute correlation coefficients between scores produced by machine learning methods and mean clinical ratings ranged from 0.29 to 0.83 (see Table 3). The best combination (feature selection plus machine learning method) was stepwise regression and SVM.

Table 3. Absolute correlation coefficients (RMSE) between mean rating of three raters and machine learning-based scores using stepwise regression and PCA.

	Stepwise			PCA		
	SVM	LR	DT	SVM	LR	DT
TRS	0.81 (0.77)	0.74 (0.86)	0.64 (1.02)	0.60 (1.05)	0.55 (1.07)	0.57 (1.01)
UPDRS #31	0.83 (0.53)	0.77 (0.59)	0.73 (0.65)	0.61 (0.74)	0.54 (0.76)	0.56 (0.77)
SUMUPDRS	0.78(1.65)	0.76(1.59)	0.74(1.74)	0.66(1.90)	0.61(1.93)	0.63(1.99)
Dyskinesia	0.67 (0.50)	0.56 (0.58)	0.35 (0.7)	0.41 (0.65)	0.38 (0.64)	0.29 (0.71)

All the coefficients had a P-value < 0.001. TRS is overall mobility, UPDRS #31 is body bradykinesia, SUMUPDRS is the sum of the UPDRS #26 (leg agility), UPDRS #27 (arising from a chair) and UPDRS #29 (gait) and Dyskinesia is the severity of dyskinesia (involuntary movements).

The validity results of applying PCA as a feature selection method were lower than applying stepwise regression. These results were consistent for the three machine learning methods. Therefore, it was decided to use the scores produced from stepwise regression and SVM in subsequent analysis.

Responsiveness of the SVM-based scores on the four scales was reasonable in relation to the clinical scores (Figure 6). The biggest gap between the effect sizes could be seen when assessing the responsiveness of dyskinesia.

In the sample set, 89% of patients had asymmetrical motor symptoms. There were 9 patients who were affected mostly on their right side and 8 patients on the left side. In the two groups of patients, the PC1s between right and left legs were not different (P-value = 0.40 for right side affected PD patients; P-value = 0.48 for left side affected PD patients). These results indicate that the PD asymmetry did not have any effect on the performance of the leg agility tests, as measured by the PC1.

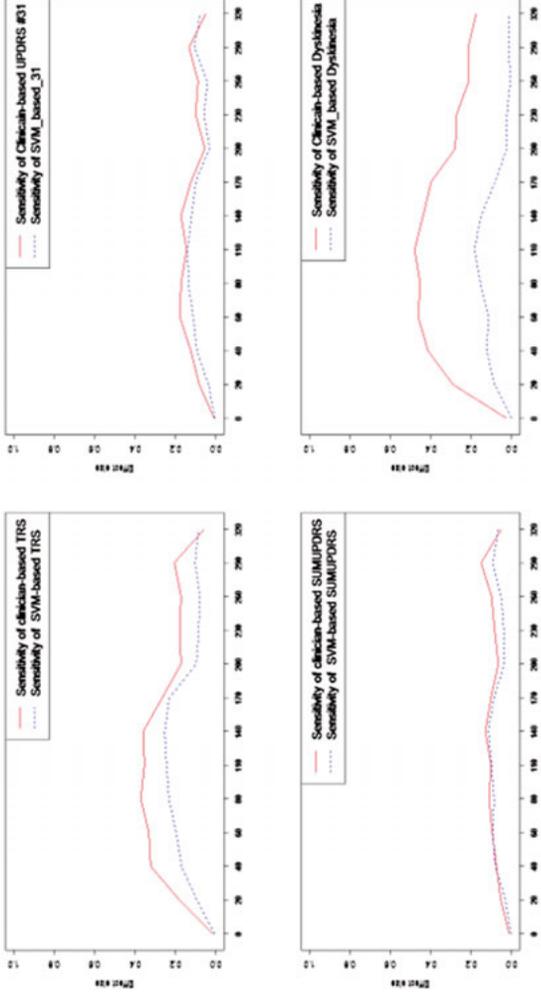


Fig 6. Responsiveness to treatment analysis of the mean clinical ratings (solid lines) and SVM-based scores (dashed lines) for each of the four scales across the test occasions. The horizontal axis represents the minutes after taking the levodopa dose. The vertical axis represents the effect sizes representing the changes in scores between baseline and later tests e.g. changes between baseline test and first test, baseline test and second test and so on. Number of tests per time slot: 0 ($n=19$), 20 ($n=19$), 40 ($n=19$), 60 ($n=19$), 80 ($n=19$), 110 ($n=19$), 140 ($n=19$), 170 ($n=19$), 200 ($n=19$), 230 ($n=18$), 260 ($n=15$), 290 ($n=13$), 320 ($n=11$). Since for test #15 at 350 min there was only observation performed by a patient it was decided to not include that observation in the calculation of effect sizes.

Paper V

Multisensor data fusion methods were developed and evaluated for quantification of PD motor symptoms. Treatment response index from multimodal motion sensors (TRIMMS) scores was constructed based on the scores derived from the LR method which were applied on features extracted from all tests. The results including correlation coefficients and root mean square errors from developed methods are shown in Table 4.

Among the results of single tests, the features which were selected by Lasso from Arm swing (AS) and were used as inputs to SVM algorithm, contained the most valid ($R = 0.89$, $RMSE = 0.50$) information to the mean TRS. This result is similar to when the AS features were selected by stepwise and were used in SVM ($R = 0.89$, $RMSE = 0.51$). These results are presented in bold in Table 4.

With respect to the performance of each machine learning methods using 10-fold CV, the SVM and LR performed generally with higher validity than DT ($R = 0.74-0.93$ for SVM, $0.49-0.87$ for DT and $0.51-0.95$ for LR). It's visible that these two methods have almost similar results on some features sets, e.g. walking feature set selected by stepwise; combined feature set selected by Lasso. However, when the features were selected by PCA, the validity of machine learning methods to mean TRS were lower than when they were selected by the two other methods.

In fact, the LR algorithm resulted in the highest validity ($R = 0.95$, $RMSE = 0.34$, $p\text{-value} = 0.00$ with 10-fold; and $R = 0.93$, $RMSE = 0.34$, $p\text{-value} = 0.00$ with L1O) to mean TRS where the features from combined set were selected by stepwise method.

Table 4. Validity results of machine learning methods, (10-fold cross validation correlation coefficient; RMSE) [Leave-one-out correlation coefficient; RMSE] after applying feature selection.

	Stepwise			Lasso			PCA		
	SVM (R,RMSE) [R,RMSE]	DT (R,RMSE) [R,RMSE]	LR (R,RMSE) [R,RMSE]	SVM (R,RMSE) [R,RMSE]	DT (R,RMSE) [R,RMSE]	LR (R,RMSE) [R,RMSE]	SVM (R,RMSE) [R,RMSE]	DT (R,RMSE) [R,RMSE]	LR (R,RMSE) [R,RMSE]
LA [†]	(0.76;0.73) [0.40;1.08]	(0.59;0.92) [0.15;1.35]	(0.69;0.81) [0.39;1.11]	(0.79;0.68) [0.37;1.07]	(0.65;0.90) [0.12;1.39]	(0.70;0.79) [0.39;1.11]	(0.78;0.72) [0.33;1.12]	(0.49; 1.0) [0.23;1.24]	(0.62;0.87) [0.40;1.05]
HR ^{††}	(0.81;0.66) [0.59;0.88]	(0.69;0.82) [0.37;1.17]	(0.83;0.63) [0.72;0.77]	(0.82;0.64) [0.72;0.75]	(0.77;0.71) [0.62;0.89]	(0.76;0.73) [0.69;0.80]	(0.77;0.70) [0.36;1.05]	(0.66;0.84) [0.30;1.23]	(0.70;0.78) [0.50;1.30]
W [‡]	(0.77;0.71) [0.51;0.95]	(0.70;0.80) [0.22;1.28]	(0.76;0.73) [0.61;0.93]	(0.79;0.68) [0.37;1.07]	(0.65;0.90) [0.12;1.39]	(0.70;0.79) [0.39;1.11]	(0.74;0.75) [0.48;0.99]	(0.55;0.95) [0.39;1.13]	(0.51;0.98) [0.18;1.33]
AS ^{**}	(0.89;0.51) [0.79;0.68]	(0.81;0.64) [0.70;0.80]	(0.83;0.62) [0.71;0.85]	(0.89;0.50) [0.77;0.72]	(0.78;0.68) [0.68;0.82]	(0.82;0.63) [0.66;1.02]	(0.85;0.58) [0.71;0.79]	(0.73;0.79) [0.66;0.87]	(0.77;0.71) [0.58;1.16]
All ^{§§}	(0.90;0.50) [0.65;0.85]	(0.77;0.72) [0.48;1.05]	(0.95;0.34) [0.93;0.40]	(0.93;0.44) [0.78;0.73]	(0.87;0.57) [0.39;1.14]	(0.91;0.46) [0.79;0.75]	(0.89;0.55) [0.54;0.94]	(0.79;0.70) [0.61;0.93]	(0.89;0.53) [0.65;0.89]

[†] Leg agility features. $N = 24$. ^{††} Alternate hand rotation features. $N = 88$. [‡] Walking features. $N = 37$. ^{‡‡} Arm swings while walking features. $N = 29$. ^{§§} All is combination of features from leg agility, alternate hand movements, walking and arm swings while walking. $N = 178$.

In Figure 7 the actual TRS scores and the TRIMMS scores which were based on the LR method across the time points are shown.

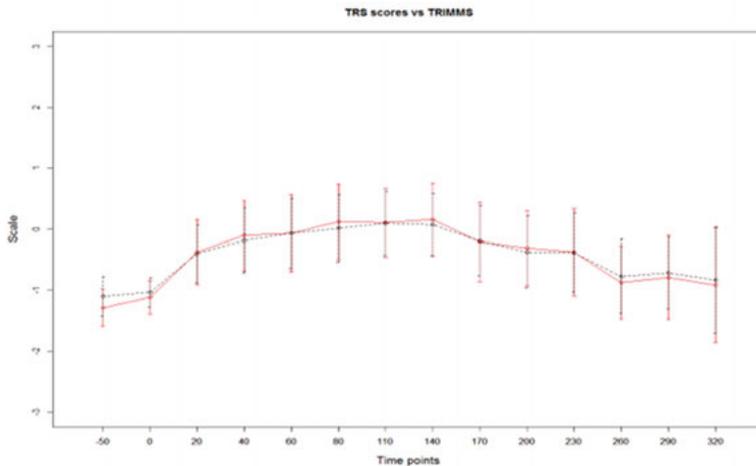


Fig 7. Actual TRS scores vs TRIMMS scores generated by the LR method across the test occasions. The straight line is for mean clinical TRS ratings and the dashed line is for TRIMMS scores. The X-axis presents the time points of tests. The time point in which the first test was taken within 50 min before the dose intake is denoted as -50. The Y-axis presents the range of TRS. Number of tests per time points: -50 (n=14), 20 (n=15), 40 (n=16), 60 (n=15), 80 (n=16), 110 (n=16), 140 (n=17), 170 (n=16), 200 (n=15), 230 (n=15), 260 (n=13), 290 (n=11), 320 (n=8), 350 (n=2). Test 15 at 350 min contained only one patient and it was not enough to be included.

Responsiveness to levodopa for mean TRS scores and TRIMMS scores are shown in Figure 8. TRIMMS scores showed good responsiveness to levodopa even though they were slightly less sensitive than TRS scores at some time points, e.g. 40, 80 and 110 minutes.

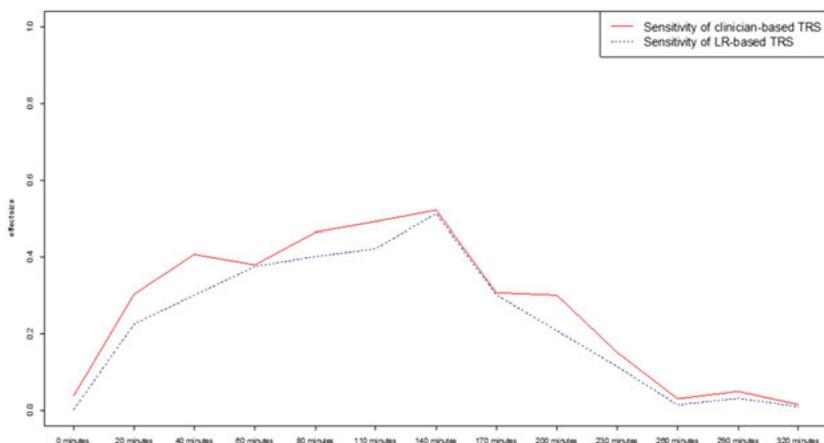


Fig 8. Responsiveness to treatment analysis across the test occasions presented by the straight line for mean clinical TRS ratings and dashed line for TRIMMS scores. The X-axis presents the minutes after taking the dose. The Y-axis presents the effect sizes which represents the amount of change in scores between baseline and later tests e.g. effect size at 0 minutes indicate the amount of change between first test and second test and 20 minutes indicates the amount of change between first test and third test. Number of tests per time points: 0 (n=15), 20 (n=15), 40 (n=16), 60 (n=15), 80 (n=16), 110 (n=16), 140 (n=17), 170 (n=16), 200 (n=15), 230 (n=15), 260 (n=13), 290 (n=11), 320 (n=8), 350 (n=2). Test 15 at 350 min contained only one patient, which was not enough to be included in effect size calculation. TRIMMS scores show good responsiveness to levodopa compared to the responsiveness of the clinician-based TRS scores. At some times (40, 80, 110 minutes) the effect sizes of the TRIMMS scores are less than the effect sizes of the TRS scores. The responsiveness trend of the TRIMMS scores is similar to the responsiveness trend of the TRS scores.

The correlation of the TRIMMS to other clinical ratings was further investigated. Correlation coefficients of the TRIMMS scores derived by LR to other clinical ratings are shown in Table 5. LR-based scores had the highest correlation coefficients to bradykinesia, dyskinesia, and gait. Their lowest correlation coefficient was to UPDRS item #25 (arising from a chair).

Table 5. Correlation coefficients and RMSE of TRIMMS to UPDRS items and dyskinesia.

	UPDRS #23	UPDRS #25	UPDRS #26	UPDRS #27	UPDRS #29	UPDRS #31	Dyskinesia
TRS	-0.36	-0.54	-0.40	-0.61	-0.80	-0.85	0.87
LR-based scores	-0.34	-0.51	-0.39	-0.60	-0.79	-0.84	0.85

UPDRS #23 is finger tapping, UPDRS #25 Rapid alternating movements of hands, UPDRS #26 is leg agility, UPDRS #27 is arising from chair, UPDRS #29 is gait, UPDRS #31 bradykinesia. All correlation coefficients were significant ($p < 0.001$). The TRIMMS scores are highly correlated first to bradykinesia (UPDRS #31), then to dyskinesia and to gait (UPDRS #29).

Paper VI

The mean cross correlation coefficient of the similarity measure was 0.84 indicating IMU sensor captures similar acceleration quantities to what is captured by wrist motion sensor during hand rotation tests.

From each of the IMU, wrist motion sensor, and a previous study (St1) feature sets, 11 PCs were retained as part of the data that contained the most amount of information from datasets and each of which had eigenvalues higher than 1. The accumulated variances of the PCs were 83%, 80%, and 82% for IMU, wrist motion sensor and St1, respectively. The most 10 important features extracted from PCs of the three feature sets are summarized in Table 6. The common features are shown in bold. More than 50% of the features from the data in this study were in common with the features calculated from acceleration data of St1. These features were the mean, SD of the Macc and Discrete Wavelet Transform (DWT) of Macc, as well as the SD of 2nd DWT on X-axis. In addition, SD of 2nd DWT on z-axis was similar between smartphone and St1.

Table 6. First 10 most important features, for IMU, wrist motion sensor, and St1.

	IMU	MOTION SENSOR	ST1
Mean M_{acc}	2	2	8
Mean Acc on X-axis		9	
Mean DWT on M_{acc}	3	3	7
SD of M_{acc}	5	8	2
SD Acc on X-axis		5	
SD Acc on Y-axis	4		
SD of Z-axis	9	7	
SD of DWT on M_{acc}	1	1	3
SD DWT on X-axis		4	
SD DWT on Z-axis	8	10	
SD 1st level DWT on M_{acc}			9
SD 1st level DWT on X-axis	10		6
SD 2rd level DWT on M_{acc}			1
SD 2rd level DWT on X-axis	6	6	5
SD 2rd level DWT on Y-axis			10
SD 2rd level DWT on Z-axis	7		4

Macc is magnitude of acceleration. SD is standard deviation. Acc is acceleration. DWT is discrete wavelet transform method.

Machine learning methods were used to map the sensor data to the states defined in TRS. Using the feature sets from IMU and wrist motion sensors as individual testing sets, the performance of the method for estimating the state of the healthy subject was explored. The best prediction performance was expected to be 0. Using the feature set from wrist motion sensor data as testing set in SVM, the absolute mean value of the predicted TRS scores was 0.05. With the same approach but using feature set from IMU as a testing set, the absolute mean of the predicted TRS scores was 0.01.

Discussion and conclusions

This thesis described the methods for objective assessment of PD motor states and dexterity during different standardized tests where the data were collected by different sensor systems. The main interest of the thesis was to investigate using what sensors and from which body segments the PD motor symptoms could be detected best in regards to the rating scales. For this purpose, multiple tests from multiple body parts of PwPD were quantified using different data-driven methods.

In a review study which included the viewpoints of a panel of movement disorder specialists the importance of objective measures and how they should be organized was discussed. They reflected the importance of continuously monitoring the motor movements and measuring the complications of PD. The responsiveness of the measurements to the treatment throughout the daily life of PD patients was highlighted as well. Long term monitoring of PD can be useful in order to capture a complete picture of PD from the PD onset. Since about 50% of the dopaminergic neurons are lost before the motor symptoms show up, detecting early PD is crucial [82]. Capturing the abnormalities from the early stages of the disease can help the methods to be developed according to how the motor symptoms and the disease progress. The type of the data collection we had in this thesis was a short term which means the data from motor functions were collected during a day or in a short time. The developed methods in the short term have been promising in detecting the states of the motor symptoms. In studies included in this thesis, it is shown that the short term sensor measurements were able to capture the impairments from different body segments of PwPD and it might be possible to expand the methods using long term collected data.

An aspect that was reflected by PD experts was “what should be measured in PD?” which was part of our interest in this thesis as well. As it is also reflected in our review study [33] bradykinesia and tremor are the most measured PD symptoms. Differentiating dyskinesia from tremor, and motor fluctuations from non-motor fluctuations are difficult for PD, and early fluctuations need an accurate measurement of the movements where micro-level data need to be collected. However, we discovered some body segments might reflect more symptom related information. Hand movements during walking showed promising results compared to the other body segments that were investigated. These results are shown in paper V where the convergent validity

of the methods using different features sets was calculated from multiple body segments and was shown.

Moreover in the studies covered by this thesis different instruments were employed for collecting the data. Depending on the symptom and the type of data to be collected various sensors can be employed. Our research showed that the clinimetric properties of the methods in paper IV and V with motion sensors were better than the results in paper I where the combination of spiral and tapping tests were analyzed. This comparison can be done since more or less the data of same PD patients and healthy controls were analyzed. However, as we have highlighted in paper VI, using smartphones for collecting symptom data has many advantages over the motion sensors. We aimed to investigate the feasibility of quantifying PD motor symptoms using acceleration data of smartphones and the results were promising. In overall, the recent MEMS technology has provided devices that are portable, light, unobtrusive, inexpensive and accurate in measurements. Therefore, wearable sensors can enable measurements in uncontrolled environments as well as in clinics. The dominant body positions for the placement of sensors in paper VI for objective motor symptom assessment have been discussed. In a study [83] it was found that mounting sensors on the back of the hands suits best and reflects more information about symptoms. They highlighted that adding more sensors will not necessarily increase the accuracy or efficiency of the methods for objective assessment. This has to be investigated when using smartphones for collection and processing the PD symptom data.

When focusing on the results provided by the machine learning methods that were developed for quantification of the states in this thesis, SVM provided promising results compared to other learning methods. This method was fast in implementation and able to separate the nonlinearly related data of motor symptoms and clinical ratings. There was an exception like in paper V where the LR provided the best convergent validity. Hence examination of different learning methods is suggested.

The main part of the data in this thesis is collected during a one-day levodopa challenge which reflects the full cycle for possible PD states including off, on/dyskinetic, off. This full cycle allowed sensor measurements to collect all relevant information needed for quantification and individualization of doses which was another main goal of the study design. Individualization of doses in PD requires adequate factors to be provided to the dosing algorithms. In many PD patients, the generic prescription might not be efficient since the motor symptoms of this disease are subjective and one dose does not fit all. Personalized medicine can incorporate many other factors such as subject's clinical, genetic, and environmental information [84].

The smartphone-based system presented in Paper I suggests that quantification of dexterity in PD using smartphone data of spiral drawing and alternate tapping tests is feasible. This system is useful for measuring treatment-related changes in patients with PD.

The approximate entropy method presented in Paper II indicates that temporal irregularity score that was calculated for the drawing speed during spiral tests is useful for long term diagnostics. The method can reasonably discriminate well between spiral drawings drawn by patients in different stages of PD and healthy subjects. As compared to the previous measures, this method quantifies a different aspect of upper limb motor severity. The results from investigating the properties of this method in paper III approves the results in paper II and shows that this method is able to capture some effects of levodopa medication that were presented in patients' performances.

The findings from the data-driven methodology presented in paper IV demonstrate good clinimetric properties for quantification the motor states in PD using motion sensors data during leg agility tests. The proposed methodology could form the basis for developing systems for follow up of the effects of treatment and individualizing treatments in PD.

The multimodal sensor fusion platform presented in paper V indicate that using a machine learning method that fused information from standardized motor tasks leads to highly valid, reliable and sensitive objective measurement of PD motor symptoms. The platform provides high accuracy for discriminating the healthy controls from PD patients as well.

The results presented in paper VI suggest that smartphone IMUs provide sufficient information for quantification of the motor states during hand rotation tests.

Limitations

This thesis exhibited a considerable amount of inter-rater variability in rating scores of the clinical ratings. This is the case in studies including subjective ratings and cannot be avoided. This problem might be solved by training the raters for scoring the symptoms or to include more experts for ratings. Another problem was that the scales like TRS contained few scores for extreme cases (e.g. score -3 and +3) which caused the methods to have not enough extreme observations to learn about those situations and perform better when quantifying such cases. Regarding this issue in paper IV, a large proportion of assigned scores by TRS were between -2 and +1. The methods tended to underestimate the scores and to concentrate their predictions around the means of the population. This made it difficult for the machine learning methods to predict cases outside of this range. Comparing this to the previous work on quantifying PD motor states using motion sensor data during walking tests [16], the results of leg agility data analysis showed it is less suitable for capturing dyskinesias since the correlations and RMSE in this study were lower. This was also reflected in the lower responsiveness of the methods to treatment as compared to the responsiveness of the clinical TRS. In paper V the performance of the machine learning with the results in the two studies of leg agility

(paper IV) and alternate hand rotation [19] was slightly different. This was because only the data of subjects who could perform all tests were included in the dataset for analysis e.g. there were two advanced patients who did not perform the walking tests. In the study with alternate hand rotation, due to the lack of observations at tails, the PC scores were calibrated based on the values of the PCs that corresponded to the tails of the distribution of mean TRS values. Whereas in paper V the PC scores were not calibrated.

Paper II included our previous study which developed and evaluated the clinimetric properties of an ApEn score for measuring temporal irregularity score (TIS) in PD using digital spiral analysis. In that study, the second dataset was used for analysis. But the correlations between TIS and clinical ratings, the sensitivity of TIS to a single dose of levodopa over a course of test trials, and the reproducibility of the previously reported TIS properties were not possible to be studied. Therefore in paper III, the results of our previous study were verified and further investigated using the calculated upper limb temporal irregularity of PD patients with same methodology (ApEn), using the main dataset including new patients group, using another screen resolution, and during shorter term measurements.

In paper VI, quantification of hand movements using the data from one healthy control was not enough. The entire study needs to be done using data from a complete dataset which includes both healthy controls and the PwPD. The dataset of PwPD can preferably be from all disease stages and all symptom states to include a wide picture of the motor functions.

At the end, one of the issues with wearable sensors is that whether they provide assurance on safety, efficacy, and privacy. This is a potential risk in wearable technologies. The privacy of the patients related to the legal obligations regarding privacy and the ownership of the data need to be ensured.

Future studies

A goal of future work of this thesis is to develop a continuous or real-time objective monitoring of PD to be included in individualizing of the doses for PD patients. For this, the methods need to be evaluated in different and large datasets.

As in many other studies, a limitation in this study was the adopted small dataset which limits the generalizability of the results. In order for the results to be applicable in clinics the longitudinal and large-sized validations are needed [85]. However, quantification of the short-term data was promising. The design of the study included the levodopa cycle where it made the evaluation of medication effects possible.

The data collection, development and application of the methods in future work requires a large number of PwPD to be examined. The data to be collected on a daily basis, free-living and during the long term. We explored that

the results from paper V were based on the data that was collected from multi-body segments. Therefore future study design may include the tests to be performed by both upper and lower extremities for achieving high accuracy in quantification of the motor symptoms. This is because the symptoms show up with different degrees and at different body parts. Although sustainability and acceptance, of the body worn sensors in life of PwPD needs to be considered.

Analysis of the arms swings while walking at larger scale would be of interest, since analysis of the respective dataset provided interesting information with regard to the motor symptoms. For this, smartwatches or wrist motion sensors can be employed for collecting the data. They are available in market and performing the test is easy. Plus smartwatches are an accepted accessories in daily life of many patients.

The long term collection of the data allows monitoring the progression of the disease. In addition, it is important that PwPD receive feedbacks regarding the responds they provide in questionnaires and the objective motor tests that they perform.

There are only a few systems that have the standards. Only two of them have regulatory FDA approval. Kinesia and Global kinetics [86], [87]. Practical guidance on objective measurement and the optimum use of devices is lacking [44]. Future work should incorporate the developed objective measures into clinical guidelines. Optimization of objective measures needs to be done by cutoff values that separate the controlled from uncontrolled symptoms. The cutoffs can be subsequently tested and redefined by consensus.

References

- [1] D. Salat, and E. Tolosa, "Levodopa in the treatment of Parkinson's disease: current status and new developments," *J Parkinsons Dis*, vol. 3, no. 3, pp. 255-69, Jan 1, 2013.
- [2] C. G. Goetz, B. C. Tilley, S. R. Shaftman, W. Poewe, C. Sampaio, G. T. Stebbins, *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Mov Disord*, vol. 23, no. 15, pp. 2129-70, Nov 15, 2008.
- [3] R. A. Hauser, F. Deckers, and P. Leher, "Parkinson's disease home diary: further validation and implications for clinical trials," *Mov Disord*, vol. 19, no. 12, pp. 1409-13, Dec, 2004.
- [4] A. E. Spottke, M. Reuter, O. Machat, B. Bornschein, S. von Campenhausen, K. Berger, *et al.*, "Cost of illness and its predictors for Parkinson's disease in Germany," *Pharmacoeconomics*, vol. 23, no. 8, pp. 817-836, 2005.
- [5] B. Post, M. P. Merkus, R. M. de Bie, R. J. de Haan, and J. D. Speelman, "Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?," *Mov Disord*, vol. 20, no. 12, pp. 1577-84, Dec, 2005.
- [6] D. Isacson, K. Bingefors, I. S. Kristiansen, and D. Nyholm, "Fluctuating functions related to quality of life in advanced Parkinson disease: effects of duodenal levodopa infusion," *Acta Neurol Scand*, vol. 118, no. 6, pp. 379-86, Dec, 2008.
- [7] A. J. Espay, P. Bonato, F. B. Nahab, W. Maetzler, J. M. Dean, J. Klucken, *et al.*, "Technology in Parkinson's disease: Challenges and opportunities," *Mov Disord*, vol. 31, no. 9, pp. 1272-82, Sep, 2016.
- [8] J. A. Stamford, P. N. Schmidt, and K. E. Friedl, "What Engineering Technology Could Do for Quality of Life in Parkinson's Disease: A Review of Current Needs and Opportunities," *Ieee Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1862-1872, Nov, 2015.
- [9] M. Mevludin, "Mobile systems for monitoring Parkinson's disease," School of Technology and Business Studies, Computer Engineering, Örebro University, 2014.
- [10] P. Mazzoni, B. Shabbott, and J. C. Cortes, "Motor Control Abnormalities in Parkinson's Disease," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 6, Jun, 2012.
- [11] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, *et al.*, "Body-Sensor-Network-Based Kinematic Characterization and Comparative Outlook of UPDRS Scoring in Leg Agility, Sit-to-Stand, and Gait Tasks in Parkinson's Disease," *IEEE Biomed Health Inform*, vol. 19, no. 6, pp. 1777-93, Nov, 2015.

- [12] M. Giuberti, G. Ferrari, L. Contin, V. Cimolin, N. Cau, M. Galli, *et al.*, "On the Characterization of Leg Agility in Patients with Parkinson's Disease," in *Conf Proc IEEE Body Sensor Networks*, 2013.
- [13] M. H. Li, T. A. Mestre, S. H. Fox, and B. Taati, "Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation," *Neuroeng Rehabil*, vol. 15, no. 1, pp. 97, Nov 6, 2018.
- [14] A. Delval, L. Defebvre, and C. Tard, "Freezing during tapping tasks in patients with advanced Parkinson's disease and freezing of gait," *PLoS One*, vol. 12, no. 9, pp. e0181973, 2017.
- [15] L. Palmerini, S. Mellone, G. Avanzolini, F. Valzania, and L. Chiari, "Quantification of Motor Impairment in Parkinson's Disease Using an Instrumented Timed Up and Go Test," *IEEE Trans Neural Syst Rehabil Eng*, vol. 21, no. 4, pp. 664-673, Jul, 2013.
- [16] I. Thomas, F. Bergquist, R. Constantinescu, D. Nyholm, M. Senek, and M. Memedi, "Using measurements from wearable sensors for automatic scoring of Parkinson's disease motor states: results from 7 patients," *Conf Proc IEEE Eng Med Biol Soc*, pp. 131-134, 2017.
- [17] M. M. Hoehn, and M. D. Yahr, "Parkinsonism: onset, progression, and mortality (Reprinted from *Neurology*, vol 17, pg 427-442, 1967)," *Neurology*, vol. 50, no. 2, pp. B1-B16, Feb, 1998.
- [18] A. Merola, A. Sturchio, S. Hacker, S. Serna, J. A. Vizcarra, L. Marsili, *et al.*, "Technology-based assessment of motor and nonmotor phenomena in Parkinson disease," *Expert Rev Neurother*, vol. 18, no. 11, pp. 825-845, Nov, 2018.
- [19] I. Thomas, J. Westin, M. Alam, F. Bergquist, D. Nyholm, M. Senek, *et al.*, "A Treatment-Response Index From Wearable Sensors for Quantifying Parkinson's Disease Motor States," *IEEE Biomed Health Inform*, vol. 22, no. 5, pp. 1341-1349, Sep, 2018.
- [20] J. M. Fisher, N. Y. Hammerla, L. Rochester, P. Andras, and R. W. Walker, "Body-Worn Sensors in Parkinson's Disease: Evaluating Their Acceptability to Patients," *Telemedicine and E-Health*, vol. 22, no. 1, pp. 63-69, Jan 1, 2016.
- [21] C. Hansen, A. Sanchez-Ferro, and W. Maetzler, "How Mobile Health Technology and Electronic Health Records Will Change Care of Patients with Parkinson's Disease," *Journal of Parkinson's disease*, vol. 8, no. s1, pp. S41-S45, 2018.
- [22] B. G. Tabachnick, and L. S. Fidell, *Using multivariate statistics*: Allyn & Bacon/Pearson Education, 2007.
- [23] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, "Multivariate data analysis. 1998," *Upper Saddle River*, 1998.
- [24] N. R. Draper, and H. Smith, *Applied regression analysis*: John Wiley & Sons, 2014.
- [25] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267-288, 1996.
- [26] S. McQuitty, "The Purposes of Multivariate Data Analysis Methods: an Applied Commentary," *Journal of African Business*, vol. 19, no. 1, pp. 124-142, 2018.
- [27] A. Getinet, "Parkinson's Disease: A Concise Overview of Etiology, Epidemiology, Diagnosis, Comorbidity and Management.," pp. 4-6, 2016.

- [28] C. G. Goetz, "The History of Parkinson's Disease: Early Clinical Descriptions and Neurological Therapies," *Cold Spring Harbor Perspectives in Medicine*, vol. 1, no. 1, Sep, 2011.
- [29] L. M. L. de Lau, and M. M. B. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurology*, vol. 5, no. 6, pp. 525-535, Jun, 2006.
- [30] P. Andlin-Sobocki, B. Jonsson, H. U. Wittchen, and J. Olesen, "Cost of disorders of the brain in Europe," *Eur J Neurol*, vol. 12 Suppl 1, pp. 1-27, Jun, 2005.
- [31] A. Gustavsson, M. Svensson, F. Jacobi, C. Allgulander, J. Alonso, E. Beghi, *et al.*, "Cost of disorders of the brain in Europe 2010," *Eur Neuropsychopharmacol*, vol. 21, no. 10, pp. 718-79, Oct, 2011.
- [32] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 79, no. 4, pp. 368-376, Apr, 2008.
- [33] S. Aghanavesi, and J. Westin, "A Review of Parkinson's Disease Cardinal and Dyskinetic Motor Symptoms Assessment Methods Using Sensor Systems," *Internet of Things Technologies for Healthcare, Healthiot 2016*, vol. 187, pp. 52-57, 2016.
- [34] G. Linazasoro, and A. Gorospe, "[Treatment of complicated Parkinson disease with a solution of levodopa-carbidopa and ascorbic acid]," *Neurologia*, vol. 10, no. 6, pp. 220-3, Jun-Jul, 1995.
- [35] D. Nyholm, A. I. Nilsson Remahl, N. Dizdar, R. Constantinescu, B. Holmberg, R. Jansson, *et al.*, "Duodenal levodopa infusion monotherapy vs oral polypharmacy in advanced Parkinson disease," *Neurology*, vol. 64, no. 2, pp. 216-23, Jan 25, 2005.
- [36] A. L. Benabid, S. Chabardes, J. Mitrofanis, and P. Pollak, "Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease," *The Lancet Neurology*, vol. 8, no. 1, pp. 67-81, 2009.
- [37] B. R. Thanvi, and T. C. Lo, "Long term motor complications of levodopa: clinical features, mechanisms, and management strategies," *Postgrad Med J*, vol. 80, no. 946, pp. 452-8, Aug, 2004.
- [38] R. Djaldetti, I. Ziv, and E. Melamed, "Impaired absorption of oral levodopa: a major cause for response fluctuations in Parkinson's disease," *Isr J Med Sci*, vol. 32, no. 12, pp. 1224-7, Dec, 1996.
- [39] S. Fahn, and R. L. Elton, "Unified Parkinson's disease rating scale," *Rec Dev in Parkinson's Dis*, vol. 2, pp. 153-163, 1987.
- [40] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, *et al.*, "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations," *Mov Disord*, vol. 19, no. 9, pp. 1020-8, Sep, 2004.
- [41] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, and N. Hyman, "The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score," *Age Ageing*, vol. 26, no. 5, pp. 353-7, Sep, 1997.
- [42] R. Bhidayasiri, and P. Martinez-Martin, "Clinical Assessments in Parkinson's Disease: Scales and Monitoring," *Int Rev Neurobiol*, vol. 132, pp. 129-182, 2017.

- [43] K. Yang, W. X. Xiong, F. T. Liu, Y. M. Sun, S. Luo, Z. T. Ding, *et al.*, "Objective and quantitative assessment of motor function in Parkinson's disease-from the perspective of practical applications," *Ann Transl Med*, vol. 4, no. 5, pp. 90, Mar, 2016.
- [44] P. Odin, K. R. Chaudhuri, J. Volkmann, A. Antonini, A. Storch, E. Dietrichs, *et al.*, "Viewpoint and practical recommendations from a movement disorder specialist panel on objective measurement in the clinical management of Parkinson's disease," *Npj Parkinsons Disease*, vol. 4, May 10, 2018.
- [45] S. P. Arneric, V. D. Kern, and D. T. Stephenson, "Regulatory-accepted drug development tools are needed to accelerate innovative CNS disease treatments," *Biochem Pharmacol*, vol. 151, pp. 291-306, May, 2018.
- [46] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K. M. Biglan, E. R. Dorsey, *et al.*, "Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study," *Parkinsonism Relat Disord*, vol. 21, no. 6, pp. 650-3, Jun, 2015.
- [47] M. Bachlin, M. Plotnik, D. Roggen, N. Giladi, J. M. Hausdorff, and G. Troster, "A wearable system to assist walking of Parkinson s disease patients," *Methods Inf Med*, vol. 49, no. 1, pp. 88-95, 2010.
- [48] C. Moreau, S. Cantiniaux, A. Delval, L. Defebvre, and J. P. Azulay, "Gait disorders in Parkinson's disease: and pathophysiological approaches," *Rev Neurol (Paris)*, vol. 166, no. 2, pp. 158-67, Feb, 2010.
- [49] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *J Neurosci Methods*, vol. 167, no. 2, pp. 340-8, Jan 30, 2008.
- [50] R. Cano-de-la-Cuerda, L. Vela-Desojo, J. C. Miangolarra-Page, and Y. Macias-Macias, "Isokinetic dynamometry as a technologic assessment tool for trunk rigidity in Parkinson's disease patients," *NeuroRehabilitation*, vol. 35, no. 3, pp. 493-501, 2014.
- [51] X. Liu, C. B. Carroll, S. Y. Wang, J. Zajicek, and P. G. Bain, "Quantifying drug-induced dyskinesias in the arms using digitised spiral-drawing tasks," *J Neurosci Methods*, vol. 144, no. 1, pp. 47-52, May 15, 2005.
- [52] A. Sama, C. Perez-Lopez, J. Romagosa, D. Rodriguez-Martin, A. Catala, J. Cabestany, *et al.*, "Dyskinesia and motor state detection in Parkinson's disease patients with a single movement sensor," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2012, pp. 1194-7, 2012.
- [53] G. Lopane, S. Mellone, L. Chiari, P. Cortelli, G. Calandra-Buonaura, and M. Contin, "Dyskinesia detection and monitoring by a single sensor in patients with Parkinson's disease," *Mov Disord*, vol. 30, no. 9, pp. 1267-71, Aug, 2015.
- [54] M. Mancini, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and L. Chiari, "Trunk accelerometry reveals postural instability in untreated Parkinson's disease," *Parkinsonism Relat Disord*, vol. 17, no. 7, pp. 557-62, Aug, 2011.
- [55] L. F. Yeung, K. C. Cheng, C. H. Fong, W. C. Lee, and K. Y. Tong, "Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway," *Gait Posture*, vol. 40, no. 4, pp. 532-8, Sep, 2014.
- [56] A. L. Adkin, B. R. Bloem, and J. H. Allum, "Trunk sway measurements during stance and gait tasks in Parkinson's disease," *Gait Posture*, vol. 22, no. 3, pp. 240-9, Nov, 2005.

- [57] C. Ahlrichs, A. Sama, M. Lawo, J. Cabestany, D. Rodriguez-Martin, C. Perez-Lopez, *et al.*, "Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients," *Med Biol Eng Comput*, vol. 54, no. 1, pp. 223-33, Jan, 2016.
- [58] B. K. Scanlon, B. E. Levin, D. A. Nation, H. L. Katzen, A. Guevara-Salcedo, C. Singer, *et al.*, "An accelerometry-based study of lower and upper limb tremor in Parkinson's disease," *J Clin Neurosci*, vol. 20, no. 6, pp. 827-30, Jun, 2013.
- [59] T. O. Mera, D. A. Heldman, A. J. Espay, M. Payne, and J. P. Giuffrida, "Feasibility of home-based automated Parkinson's disease motor assessment," *J Neurosci Methods*, vol. 203, no. 1, pp. 152-6, Jan 15, 2012.
- [60] S. Mellone, L. Palmerini, A. Cappello, and L. Chiari, "Hilbert-Huang-based tremor removal to assess postural properties from accelerometers," *IEEE Trans Biomed Eng*, vol. 58, no. 6, pp. 1752-61, Jun, 2011.
- [61] A. Salarian, H. Russmann, C. Wider, P. R. Burkhard, F. J. Vingerhoets, and K. Aminian, "Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system," *IEEE Trans Biomed Eng*, vol. 54, no. 2, pp. 313-22, Feb, 2007.
- [62] J. Westin, M. Dougherty, D. Nyholm, and T. Groth, "A home environment test battery for status assessment in patients with advanced Parkinson's disease," *Comput Methods Programs Biomed*, vol. 98, no. 1, pp. 27-35, Apr, 2010.
- [63] J. Westin, S. Ghiamati, M. Memedi, D. Nyholm, A. Johansson, M. Dougherty, *et al.*, "A new computer method for assessing drawing impairment in Parkinson's disease," *J Neurosci Methods*, vol. 190, no. 1, pp. 143-8, Jun 30, 2010.
- [64] R. J. Dunnewold, J. I. Hoff, H. C. van Pelt, P. Q. Fredrikze, E. A. Wagemans, and B. J. van Hilten, "Ambulatory quantitative assessment of body position, bradykinesia, and hypokinesia in Parkinson's disease," *J Clin Neurophysiol*, vol. 15, no. 3, pp. 235-42, May, 1998.
- [65] R. I. Griffiths, K. Kotschet, S. Arfon, Z. M. Xu, W. Johnson, J. Drago, *et al.*, "Automated assessment of bradykinesia and dyskinesia in Parkinson's disease," *J Parkinsons Dis*, vol. 2, no. 1, pp. 47-55, 2012.
- [66] M. Senek, S. M. Aquilonius, H. Askmark, F. Bergquist, R. Constantinescu, A. Ericsson, *et al.*, "Levodopa/carbidopa microtablets in Parkinson's disease: a study of pharmacokinetics and blinded motor assessment," *Clin Pharmacol*, vol. 73, no. 5, pp. 563-571, May, 2017.
- [67] S. E. Pahlagen, N. Dizdar, T. Hauge, B. Holmberg, R. Jansson, J. Linder, *et al.*, "Interim analysis of long-term intraduodenal levodopa infusion in advanced Parkinson disease," *Acta Neurol Scand*, vol. 126, no. 6, pp. e29-33, Dec, 2012.
- [68] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *J Clin Monit*, vol. 7, no. 4, pp. 335-45, Oct, 1991.
- [69] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to wavelets and wavelet transforms: a primer*.
- [70] B. Burke, "The Mathematical Microscope: Waves, Wavelets, and Beyond," *A Positron Named Priscilla: Scientific Discovery at the Frontier*, Washington, DC.: The National Academies Press., 1994.

- [71] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909-996, Oct, 1988.
- [72] D. Groth, S. Hartmann, S. Klie, and J. Selbig, "Principal components analysis," *Methods Mol Biol*, vol. 930, pp. 527-47, 2013.
- [73] H. F. Kaiser, "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141-151, 1960.
- [74] R. B. Cattell, "The Scree Test For The Number Of Factors," *Multivariate Behavioral Research*, vol. 1, no. 2, pp. 245-276, 1966/04/01, 1966.
- [75] S. A. Farmer, "An Investigation into the Results of Principal Component Analysis of Data Derived from Random Numbers," *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 20, no. 4, pp. 63-72, 1971.
- [76] J. Braeken, and M. van Assen, "An empirical Kaiser criterion," *Psychol Methods*, vol. 22, no. 3, pp. 450-466, Sep, 2017.
- [77] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014.
- [78] M. G. Kenward, and J. H. Roger, "Small sample inference for fixed effects from restricted maximum likelihood," *Biometrics*, vol. 53, no. 3, pp. 983-97, Sep, 1997.
- [79] C. H. Knapp, and G. C. Carter, "Generalized Correlation Method for Estimation of Time-Delay," *Ieee Transactions on Acoustics Speech and Signal Processing*, vol. 24, no. 4, pp. 320-327, 1976.
- [80] C. Kingsford, and S. L. Salzberg, "What are decision trees?," *Nature Biotechnology*, vol. 26, no. 9, pp. 1011-1013, Sep, 2008.
- [81] J. M. Moguerza, and A. Munoz, "Support vector machines with applications," *Statistical Science*, vol. 21, no. 3, pp. 322-336, Aug, 2006.
- [82] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, and F. Seitelberger, "Brain dopamine and the syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations," *J Neurol Sci*, vol. 20, no. 4, pp. 415-55, Dec, 1973.
- [83] L. Lonini, A. Dai, N. Shawen, T. Simuni, C. Poon, L. Shimanovich, *et al.*, "Wearable sensors for Parkinson's disease: which data are worth collecting for training symptom detection models," *Npj Digital Medicine*, vol. 1, Nov 23, 2018.
- [84] N. Titova, and K. R. Chaudhuri, "Personalized Medicine in Parkinson's Disease: Time to Be Precise," *Movement Disorders*, vol. 32, no. 8, pp. 1147-1154, Aug, 2017.
- [85] N. H. Ghassemi, F. Marxreiter, C. F. Pasluosta, P. Kugler, J. Schlachetzki, A. Schramm, *et al.*, "Combined accelerometer and EMG analysis to differentiate essential tremor from Parkinson's disease," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2016, pp. 672-675, Aug, 2016.
- [86] "Kinesia," <https://glneurotech.com/kinesia/products/kinesia-360/>.
- [87] "Global kinetics," <https://www.globalkineticscorporation.com.au/>.

