Universidade do Minho
Escola de Engenharia

Carlos Rafael Moreira Fernandes

Clinical Decision Support using Machine Learning: Parkinsonism and Fabry Disease

Master's Dissertation

Integrated Master's in Industrial Electronics and Computers Engineering

Work elaborated under supervision of:

Prof. Dr. Estela Bicho (Supervisor)

Dr. Flora Ferreira (Supervisor)

Dr. Miguel Gago (Co-Supervisor)

Dr. Olga Azevedo (Co-Supervisor)

October 2019

0.1 Agradecimentos

A realização desta dissertação não teria sido possível sem a contribuição de algumas pessoas, pelo que serão expressos os meus agradecimentos.

Em primeiro lugar quero exprimir a minha gratidão aos meus orientadores, Professora Doutora Estela Bicho e Doutora Flora Ferreira, pelos conselhos, orientação, disponibilidade e conhecimentos transmitidos ao longo deste trabalho.

Ao Doutor Miguel Gago e Doutora Olga Azevedo por toda a ajuda e auxílio prestado relacionado com a componente médica desta dissertação e por terem tronado possível a elaboração deste trabalho de investigação.

Aos meus pais e avós, por todo o apoio, confiança e por me proporcionarem a oportunidade de estudar neste curso.

À Margarida Costa, por todo o incentivo, compreensão, apoio e inspiração imprescindível.

Aos meus colegas e amigos de curso por todo o companheirismo e todos as experiências, não só no decorrer desta dissertação, mas durante toda a formação académica.

A todos os professores ao longo do meu percurso académico, pela contribuição na minha formação e por todos os valores e princípios transmitidos.

Gostava também de agradecer imenso ao projeto onde este trabalho se integra, NORTE-01-0145-FEDER- 000026 (DeM-Deus Ex Machina) financiados por NORTE2020 e FEDER.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

0.2 Resumo

Parkinsonismo Vascular (PVa), doença de Parkinson Idiopática (DPI), duas doenças associadas com Parkinsonismo, e a doença de Fabry (DF) foram investigadas usando métodos estatísticos e de aprendizagem automática (AP). O diagnóstico destas doenças é atualmente um grande desafio devido à enorme variação fenotípica e à sobreposição de fenótipos. De facto, existe um atraso considerável entre as primeiras manifestações destas doenças e o correto diagnóstico clínico. A investigação de biomarcadores capazes de assistir o atempado e correto diagnóstico de DF, PVa e DPI é extremamente importante. O objetivo deste estudo é avaliar a aptidão de métodos de AP quando utilizados para diagnosticar estas doenças.

Foram utilizados sensores vestíveis para obter dados da marcha de 15 pacientes com DPI, 14 pacientes com PVa, 36 pacientes com DF e 36 controlos. Foram também obtidos dados cardíacos e neurológicos de 95 pacientes de DF. Com base nos dados da marcha várias tarefas de classificação binárias foram realizadas utilizando seis métodos supervisionados de AP: *Support Vector Machines* (SVMs), *Random Forests* (RFs), *Multiple Layer Perceptrons* (MLPs), *Deep Belief Networks* (DBNs), *Long Short-Term Memory* (LSTM) and *Convolutional Neural Networks* (CNNs). Foram aplicados métodos de *clustering* para identificar subgrupos homogéneos de: 1) pacientes com Parkinsonismo (DPI e PVa) usando os dados de marcha e 2) pacientes com DF usando os dados cardíacos.

Todas as tarefas de classificação baseadas nos dados da marcha obtiveram ótimos resultados, destacando-se as SVMs e CNNs. Estes métodos deram provas da sua eficácia e precisão como ferramentas auxiliares capazes de assistir no processo de diagnóstico destas doenças. A análise de subgrupos homogéneos de pacientes que sofrem de Parkinsonismo revelou que PVa apresenta os padrões de marcha mais afetados. A análise de subgrupos homogéneos de pacientes com FD identificou que pacientes com manifestações cardíacas severas são mais suscetíveis a manifestações neurológicas. Os padrões de marcha de pacientes com DF revelaram também um potencial desenvolvimento de Parkinsonismo.

Esta dissertação contribuiu para o desenvolvimento de ferramentas capazes de auxiliar no processo diagnóstico e avaliação clínica de DVa, DPI e FD, estas ferramentas facilitam um prévio e correto diagnóstico destas doenças, possibilitam um melhor tratamento e uma melhor qualidade de vida aos pacientes.

Palavras-Chave: Parkinsonismo, Doença de Fabry, Aprendizagem Automática, Dados Cardíacos, Marcha.

0.3 Abstract

Vascular Parkinsonism (VaP), Idiopathic Parkinson's Disease (IPD), which are two diseases associated with Parkinsonism, and Fabry Disease (FD) were investigated with the support of different statistical and machine learning methods. Diagnosis of these diseases remains a challenge mostly due to phenotipical variability and highly overlapping phenotypes, with considerable delay between onset and clinical diagnosis. Additionally, growing evidence linking Parkinsonism with FD has been reported recently. It is then of extreme importance to explore biomarkers capable of assisting the early and correct diagnosis of FD, VaP, and IPD. The aim of this study is to evaluate the effectiveness of machine learning strategies when applied to the diagnosis of these disorders.

Wearable sensors positioned on both feet were used to acquire gait data from 15 IPD patients, 14 VaP patients, 36 FD patients, and 34 control subjects. Cardiac and neurological evaluations were also collected from 95 FD patients. Based on gait data, various binary comparative classification analysis were performed by applying six supervised machine learning algorithms: Support Vector Machines (SVMs), Random Forests (RFs), Multiple Layer Perceptrons (MLPs), Deep Belief Networks (DBNs), Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs). Clustering methods were applied to identify homogeneous subgroups of: 1) patients with Parkinsonism (IPD and VaP) based on gait data and 2) FD patients based on cardiac data.

All classification analysis based on gait data achieved very good results, especially SVMs and CNNs. These classifiers have proven to be very reliable and accurate assistance tools for the diagnosis of these disorders. The Parkinsonism subgroup analysis revealed that the most impaired gait patterns are mainly displayed by VaP patients. The FD subgroup analysis identified that patients with severe cardiac manifestations also display neurological impairments. The gait patterns of FD patients also revealed interesting results suggesting the potential development of Parkinsonism.

This dissertation contributed to the development of clinical diagnostic and evaluation tools of FD, IPD, and VaP, to facilitate the early and correct diagnosis of these diseases, leading to better treatment and improvement of the quality of life of patients.

Keywords: Parkinsonism, Fabry Disease, Machine Learning, Cardiac Data, Gait.

Contents

List of Figures

List of Tables

0.4 List of Acronyms

α-Gal A - α-Galactosidase A

AD - Alzheimer's disease

AI - Artificial intelligence

AIC - Akaike's information criterion

ALS - Amyotrophic lateral sclerosis

ANN - Artificial neural network

ANOVA - One-way analysis of variance

BMI - Body mass index

CDR - Clinical dementia rating

CI - Confidence interval

CM - Confusion matrix

CMRGlc - Cerebral metabolic rate of glucose consumption

CNN - Convolutional neural network

CNS - Central nervous system

COMT - Catechol-O-methyltransferase

CT - Computerized tomography

DAT-SPECT - Dopamine transporter single photon emission computerized tomography

DBN - Deep belief network

DBS - Deep brain stimulation

DR - Dropout rate

DT - Decision tree

DTI - Diffusion tensor imaging

ECG - Electrocardiogram

echo - Echocardiogram

EFA - Exploratory factor analysis

EHR - Electronic health record

ERT - Enzyme replacement therapy

ET - Extremely randomized trees

FD - Fabry disease

FN - False negative

FP - False positive

FOS - Fabry outcome survey

HD - Huntington's disease

HMM - Hidden Markov model

IG - Information gain

IPD - Idiopathic Parkinson's disease

GA - Genetic algorithm

Gb3 - Globotriaosylceramide

GBA - β-glucocerebrosidase

GD = Gaucher disease

GRU - Gated recurrent units

GSL - Glycosphingolipids

KNN - K-nearest neighbor

PARK2 - Parkinson protein 2

PARK7 - Protein deglycase DJ-1

PC - Principal component

PCA - Principal components analysis

PD - Parkinson's Disease

PDAT - Dementia of Alzheimer type

PET - Positron emission tomography

PINK1 - PTEN-induced putative kinase 1

L-Dopa - Levodopa

LDA - Linear discriminant analysis

LR - Learning rate

LRRK2 - Leucine-rich repeat kinase 2

LSTM - Long short term memory

LVH - Left ventricular hypertrophy

LVM - Left ventricular mass

MAO-B - Monoamine oxidase B

MCI - Mild cognitive impairment

MFS - Midwall fractional shortening

ML - Machine learning

MLP - Multiple layer perceptron

MoCA - Montreal cognitive assessment

MR - Multiple regression

MRI - Magnetic resonance imaging

NB - Naïve bayes

NDD - Neurodegenerative diseases

NMS - Non-motor symptom

PICU - Pediatric intensive care unit

RBFN - Radial basis function network

RBM - Restricted Boltzmann machines

RF - Random forest

RNN - Recurrent neural network

SNCA - α-synuclein

SVM - Support vector machine

SYB - Statistical YearBook of Portugal

TBSS - Tract-based spatial statistics

TIA - Transient ischemic attacks

TN - True negative

TP - True positive

VaP - Vascular Parkinsonism

WCSS - Within-cluster sum of squares

Part I :

Introduction and State of the Art

In this section, the motivation and background behind this work are presented, followed by a description of the goals for this work.

Chapter 1

Introduction

Neurodegenerative diseases are one of the biggest threats to public health, with the growth of life expectancy the number of people suffering from these diseases is rising (Heemels (2016)). According to the 2018 Statistical Yearbook of Portugal (SYB), the number of elderly people in Portugal is increasing and is projected that it will continue on this path in the future. In 2018 the number of people with more than 65 years in Portugal was 2 213 274 representing 21.6% of the overall population.

Parkinsonism is a general term that refers to a group of neurological disorders that cause movement disturbances. The most common disease associated with Parkinsonism is Idiopathic Parkinson's disease (IPD), being this the second most common neurodegenerative disorder and the most common motor disorder, affecting millions of people around the world (Lehosit and Leslie J. (2015)). Differential diagnosis between different variants of Parkinsonism is a difficult task, more specifically between IPD and Vascular Parkinsonism (VaP). These diseases exhibit a similar/overlapping phenotype being commonly confused, thus correct diagnosis is complicated and dependent on doctor experience (Lehosit and Leslie J. (2015)). Vascular Parkinsonism and IPD affect primarily the motor system. People with Parkinsonism are two times more likely to fall than people with other neurological conditions (Allen et al. (2013)). Symptoms worsen with disease progression leading to serious complications resulting in premature death. Neurologists should be prepared to perform a correct diagnosis of these diseases and provide specific treatment so that patients can live their lives as healthy, independent and autonomous as possible.

Another disease that greatly affects the quality of life and may lead to premature death is Fabry Disease (FD). The most severe clinical manifestations of FD associated with life-threatening complications are the

damages to the kidneys, heart, and brain. Cerebrovascular complications caused by cerebral vasculopathy are a major cause of morbidity and premature death in patients with FD (Kolodny et al. (2015)). The diagnosis of FD remains a challenge due to rare occurrence, symptom variability, different ages of onset, and severity of progression. The average delay between onset and correct diagnosis of FD is 13.7 and 16.3 years and premature death occurs on average 20 and 15 years earlier, in male and female patients, respectively (Giugliani et al. (2016)). It is then of extreme importance to explore different biomarkers that can assist in the early diagnose of FD.

Lately, there has been growing evidence showing that gait assessment can be a powerful complementary tool in the diagnosis and management of patients with motor impairments (Fernandes et al. (2018); Kubota et al. (2016); Lord et al. (2013)). Gait features can be measured easily, quickly, and at low cost using noninvasive technology such as wearable sensors. Different machine learning (ML) methods based on gait features have been widely used to discriminate between normal and abnormal gait patterns and among different pathological gait patterns such as those presented in patients with Parkinsonism, Huntington's Disease, and Alzheimer's Disease (see e.g, Xia et al. (2015)). The potential of ML methods has not yet been explored for the evaluation of FD patients.

Although the involvement of motor function is not included in the manifestations typically associated with FD, motor impairments during gait, such as postural instability and slower gait, have been demonstrated in FD (Lohle et al. (2015)). Furthermore, the most common FD manifestations are neurologic and gait has been hypothesized as the final outcome of several neurological functions (Amboni et al. (2013)). Recently a connection/interlink between Parkinsonism and FD has also been drawn. People that suffer from FD seem to be more susceptible to the development of Parkinsonism (Wise et al. (2018)). This connection is extremely important and is currently being investigated alongside diagnostic and evaluation systems that can assist doctors in the diagnosis of these disorders.

This work aims to further enhance the current knowledge about the mentioned disorders and develop intelligent and highly reliable ML systems, based on gait data extracted from wearable sensors, capable of assisting doctors in the diagnosis of IPD, VaP and FD so that these disorders can be correctly diagnosed and proper treatment procedures can be implemented as early as possible. Additionally, data extracted from cardiac examination will also be analysed to assess the connection between cardiac and neurological manifestations of FD patients.

1.1 Objectives

With all of the above in mind, the major goals of this work are the following:

- To further evaluate the effectiveness of machine learning methods when distinguishing control subjects from Parkinsonism patients based on gait patterns.
- To evaluate the effectiveness of machine learning methods when distinguishing IPD and VaP patients based on gait patterns and to identify subgroups of patients with Parkinsonism (IPD and VaP) with similar gait patterns.
- To investigate the connection between cardiac and neurological manifestations of FD patients and evaluate the impact of neurological lesions on the gait patterns of FD patients.
- To evaluate the effectiveness of machine learning methods when distinguishing healthy control subjects from FD patients based on gait characteristics.
- The final goal of this work is to investigate, also with the help of machine learning methods, the connection between Parkinsonism and FD by analysing the gait patterns of both disorders.

1.2 Dissertation structure

In Part I of this work, where this Chapter is included, an introduction to the developed work is presented where the context and motivation for this study, the goals that it envisions and its structure are described. The next Chapters describe the characteristics, risk factors, how the diagnosis is executed and current treatment for IPD, VaP, and FD. In the last Chapter of Part I, the current state of the art of machine learning methods used for the diagnosis and classification of IPD, VaP and other neurological disorders is presented. As mentioned before, the potential of machine learning methods has not yet been applied to FD.

Part II of this work uncovers the methodology and materials. It describes how, where and what data was collected and which data is going to be used for the development of the machine learning methods. The machine learning and statistical methods developed in this work are also summarized in Part II.

In Part III the results of this work are detailed. The first Chapter of Part III describes the data preparation and the evaluation metrics used to evaluate the machine learning methods. The obtained results are presented, compared and discussed. The conclusion and future work close this Part III and the presented work.

1.3 Contribution of this work

The following publications have been made based on the work developed in this dissertation:

- **Fernandes, C.**, Fonseca, L., Ferreira, F., Gago, M., Costa, L., Sousa, N., Ferreira, C., Gama, J., Erlhagen, W. & Bicho, E. (2018, December). "Artificial Neural Networks Classification of Patients with Parkinsonism based on Gait". In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2024-2030).

- Ferreira, F., **Fernandes, C.**, Gago, M., Sousa, N., Erlhagen, W., & Bicho, E. Normalization of foot clearance and spatiotemporal gait data using multiple linear regression models. In Program and Book of Abstracts XXVI Meeting of the Portuguese Association for Classification and Data Analysis (CLAD) (p. 71).

- Ferreira, F., **Fernandes, C.**, Gago, M., Sousa, N., Erlhagen, W. & Bicho, E. (2019, May) "Identification of gait patterns among patients with parkinsonism based on normalized gait obtained using multiple regression models". In Special Issue - Statistics on Health Decision Making: state of the art, Vol 1 No 1 2019. in press.

- **Fernandes, C.**, Ferreira, F., Gago, M., Azevedo, O., Sousa, N., Erlhagen, W. & Bicho, E. (2019, June) "Performance analysis of different kernels on SVM classification using different feature selec-

tion methods on Parkinsonian gait". In VI Workshop on Computational Data Analysis and Numerical Methods, (WCDANM) (pp.17-18).

- **Fernandes, C.**, Ferreira, F., Gago, M., Olga, A., Sousa, N., Erlhagen, W. & Bicho, E. (2019, November). "Gait classification of patients with Fabry's disease based on normalized gait features obtained using multiple regression models". In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

Chapter 2

Neurodegenerative and genetic diseases

2.1 Idiopathic Parkinson's Disease

Idiopathic Parkinson's disease (IPD) is the most common neurodegenerative disorder associated with Parkinsonism, affecting 1% of the world's population over 65 years old. It is most often present in patients after age 60, and age is considered the most consistent risk factor for developing IPD (Lehosit and Leslie J. (2015), Wirdefeldt et al. (2011)).

The first detailed description of IPD was developed in 1817 by James Parkinson, but the conceptualization of this disease is still under development today (Goetz (2011)). One of the main clinical challenges of IPD currently being tackled is the need for diagnostic procedures that allow a definitive diagnosis at the early stages of the disease (Kalia and Lang (2015)).

2.1.1 Characteristics

The main pathological finding currently associated with IPD is the degeneration of dopaminergic neurons in the brain, more specifically in the nigrostriatal pathway, which leads to a dopamine deficiency. Dopamine is a chemical that neurons release to communicate. When dopamine is absent an imbalance with neurotransmitters occurs whose result is an array of motor and non-motor symptoms. Symptoms do

not develop until 50-60% of the dopaminergic neurons are destroyed, and the underlying cause of neural loss is currently unknown (Wirdefeldt et al. (2011), Perera and Thevathasan (2014)).

All cardinal signs of IPD are related to motor function, these include resting tremors, bradykinesia, rigidity and postural instability. Gait disturbance at the early stages of the disease is also present. On gait examination, the patients exhibit slowness of gait, shorten stride length and shuffling gait (Lord et al. (2013)). Non-motor symptoms of IPD include anxiety, depression, hypotension, constipation, paresthesias, cramps, olfactory dysfunction, and seborrhoeic dermatitis. With disease progression the cognitive ability of patients may also decrease (Wirdefeldt et al. (2011)).

Idiopathic Parkinson's disease is a complex disease with heterogeneous manifestations. Various studies have attempted to classify subtypes of IPD, but there is still no consensus on the classification of IPD subtypes. However, the current observations suggest that there are two primary subtypes: tremor dominant (absence of other severe motor symptoms) and akinetic-rigid (slowness of movement, muscle stiffness, and gait disorder) (Marras and Lang (2013), Kalia and Lang (2015)).

With disease progression both motor and non-motor symptoms are accentuated, these can be initially managed with medication (Lehosit and Leslie J. (2015)).

2.1.2 Risk Factors

The greatest risk factor for the development of IPD is age. The occurrence of IPD increases steadily with age reaching a peak after 80 years of age. This tendency has important implications, with the increasing life expectancy of the general population the number of IPD cases are estimated to increase by 50% by 2030 (Dorsey et al. (2007)). This disease has a greater presence in Europe, North America and Australia (Pringsheim et al. (2014)).

Molecular genetic analyses have identified various genes that are associated with Parkinsonism, the most common are: α-synuclein (SNCA), β-glucocerebrosidase (GBA), Parkinson protein 2 (PARK2), PTEN-induced putative kinase 1 (PINK1), protein deglycase DJ-1 (PARK7) and Leucine-rich repeat kinase 2 (LRRK2) (Nuytemans et al. (2010)). The GBA gene seems to be the most correlated with the prevalence of IPD, a multicenter study conducted on data from 5691 IPD patients and 4898 controls concluded that

the odds ratio for any GBA mutation in IPD patients versus controls was of 5.43 (Sidransky et al. (2009)). However, most cases of IPD are sporadic and do not seem to have any association with gene mutations (Klein and Westenberger (2012)).

Recently various studies have associated environmental and behavioral factors to the development of IPD. Exposure to pesticides, consumption of dairy products, history of melanoma, traumatic brain injury and traumatic loss of consciousness are related to an increased risk of IPD. Other factors appear to reduce the risk of IPD, these are the following: smoking, caffeine consumption, higher serum urate concentrations, physical activity, and the use of some common medication like ibuprofen (Ascherio and Schwarzschild (2016), Dick et al. (2007)). Studies have also shown that the risk of developing IPD is greater for men than women (Benito-León et al. (2004), Baldereschi et al. (2000)).

2.1.3 Diagnosis

Currently, there is no specific test or diagnostic procedure that confirms the occurrence of IPD. Biomarkers and specific neuroimaging findings are scarce so the diagnosis of IPD is based on clinical criteria and the presence of parkinsonian motor features (Wirdefeldt et al. (2011)).

Neurologists analyse the medical history, conduct neurological examination, brain imaging, movement examination and evaluate the state of the patient over time. With all of the above in mind, it can be concluded that the diagnosis of IPD is observational, a high level of expertise is then required to perform an early and correct diagnosis.

Different authors have proposed different diagnostic criteria for the diagnosis of IPD. The first set of criteria was proposed by Hoehn and Yahr in 1967 (Hoehn and Yahr (1967)). Hoehn and Yahr proposed that IPD is present when two or more of the following symptoms are displayed: rigidity, postural instability, resting tremor, and hypokinesia. The authors also proposed a scale that summarizes the progression of Parkinsonism symptoms. Table 1 describes the Hoehn and Yahr scale.

Another common diagnostic criteria are proposed by the UK Parkinson's Disease Society Brain Bank. The proposed diagnostic criteria suggest that IPD must be assigned when bradykinesia is present alongside one of the following symptoms: rigidity, resting tremor and postural instability (Hughes et al. (1992)).

Table 1: Hoehn and Yahr scale

Stage	**Hoehn and Yahr Scale**
1	Unilateral involvement only usually with minimal or no functional disability
2	Bilateral or midline involvement without impairment of balance
3	Bilateral disease: mild to moderate disability with impaired postural reflexes; physically independent
4	Severe disability, still able to walk or stand unassisted
5	Confinement to bed or wheelchair unless aided

According to Berg et al., there is an urgent need to formulate new criteria for the diagnosis of IPD. New strategies and approaches are currently under investigation to design new criteria that make use of the knowledge that has been obtained in the last years of research. The focus of the new criteria should be to assist in the early and correct diagnosis of IPD (Kalia and Lang (2015), Berg et al. (2013)).

Idiopathic Parkinson's disease is often confused with other atypical parkinsonian disorders. One of the disorders that is mostly confused with IPD is Vascular Parkinsonism (VaP), particularly at early stages. A definitive diagnosis of IPD requires post-mortem examination (Lehosit and Leslie J. (2015))

2.1.4 Treatment

Currently, there is no way to stop or reverse the progression of IPD, the available therapies are only able to manage the disease symptoms. The treatment plan consists of managing the disease symptoms in order to improve the quality of life (QOL) of patients (Lehosit and Leslie J. (2015)).

In 1961, Birkmayer and Hornykiewicz discovered that patients with IPD suffered from a lack of dopamine in the brain, this led to the development of a therapy that is based on the replacement of the lost dopamine. Since dopamine can't cross the blood-brain barrier the medication that is given to patients is levodopa (L-Dopa). Levodopa is a precursor of dopamine that is capable of crossing the blood-brain barrier, being then converted into dopamine. This medication is still the gold standard for treating IPD (Lehosit and Leslie J. (2015)).

With disease progression, the brain's storage capacities for dopamine and L-Dopa declines. Since L-Dopa has a short plasma half-life the minor storage capacity leads to a pulsating plasma profile, which

raises the following problem, after taking L-Dopa for a few years the medication effect might only last for a few hours which results in a series of motor fluctuations and an excess of movement (dyskinesia) (Oertel (2017)).

To improve the treatment and mitigate the negative L-Dopa outcomes that the patient might experience other medications can be administered before or in combination with L-Dopa. Levodopa can be administered with inhibitors that degrade the enzyme monoamine oxidase B (MAO-B) (this enzyme is responsible for the breakdown of L-Dopa) if the activity of this enzyme is inhibited the levels of L-Dopa in the brain will increase. Another medication that can be used is peripheral inhibitors that degrade the enzyme catechol-O-methyltransferase (COMT), this enzyme also degrades dopamine. A different solution that can be adopted initially, while the motor symptoms are not too severe, is the use of dopamine agonists. These are similar to dopamine and stimulate the dopamine receptors in the brain. The use of these agonists help to delay the intake of L-Dopa (Oertel (2017)).

A more advanced option is to provide a nearly constant supply of L-Dopa to the blood, this can be achieved with an external pump connected to the small intestine (Oertel (2017)).

When medication is no longer capable of improving symptoms and the patient is being extremely impacted by the disease the treatment approach currently available is to perform surgery, more specifically, deep brain stimulation (DBS). These approaches are based on the implementation of electrodes in the brain, more specifically in the globus pallidus internal and subthalamic nucleus, to allow the reduction of IPD symptoms (Perera and Thevathasan (2014), Oertel (2017)). This procedure allows the decrease of medication which reduces the adverse effects that can occur when the patient is subjected to long term medication.

Surgical procedures are usually delayed 10 to 15 years after the diagnosis of IPD. However, a study conducted by Schuepbach et al. (2013) investigated the results of performing deep brain stimulation earlier in the disease course. The results show that surgery treatment was superior to classical medical therapy. This study might impact the current therapy approach of IPD.

2.2 Vascular Parkinsonism

Vascular Parkinsonism (VaP) was first described in 1929 as a syndrome of arteriosclerotic parkinsonism. The symptoms described by Critchley in 1929 were motor complications in elderly hypertensive people (Critchley (1929)). At the time this syndrome was rejected by various neurologists, it was argued that the vascular incidents seen in some IPD patients were random, but with the evolution of radiology white matter lesions were identified in patients with parkinsonism which supported the concept of VaP (Fitzgerald and Jankovic (1989)). These discoveries lead to the conclusion that VaP is a distinct clinical entity that is different from IPD in more than one aspect (Demirkiran et al. (2001), Foltynie et al. (2002)). It is estimated that VaP accounts for 12% of all cases of Parkinsonism (Thanvi et al. (2005)) and, according to a study conducted by Colosimo et al. (2010), VaP is the second most common type of parkinsonism following IPD.

Currently, IPD and VaP are distinct entities being the most critical task and main difficulty the differentiation between these two diseases upon diagnosis (Gupta and Kuruvilla (2011)).

2.2.1 Characteristics

Vascular Parkinsonism results from multiple infarcts in the brain that cause lesions, mostly in the basal ganglia and/or subcortical white matter (Sibon et al. (2004)). This disorder is characterized by lower body parkinsonism with a rapid and abrupt onset and rapid impairment of gait and postural instability (Lehosit and Leslie J. (2015)). The main distinctive feature of VaP is the severity of early postural instability and gait disturbance, VaP gait patterns are characterized by shuffling gait and it may also exhibit freezing of gait, even at early stages (Zijlmans et al. (2004)). Response to levodopa treatment is non existent or poor when compared to IPD (Thanvi et al. (2005)). Cognitive impairment is generally higher in VaP patients than in IPD patients, this might occur because VaP results from inadequate blood supply to the brain leading to greater white and gray matter lesions. Vascular risk factors, like hypertension and history of stroke, are commonly present in VaP patients.

The occurrence of cerebral infarcts is most frequent in VaP patients than in normal subjects or IPD patients (Winikates and Jankovic (1999)). A study that investigated the clinical history of strokes of 69 VaP patients and 277 IPD patients concluded that stroke events were present in 43.5% of VaP patients compared to only 2.9% in IPD patients and multiple strokes where only present in VaP patients (Winikates and Jankovic (1999)). Findings of neuroimaging of VaP patients include diffuse subcortical white matter or gray matter lesions, particularly in the basal ganglia area. However, it is important to note that these cerebrovascular lesions are also commonly seen on neuroimaging of elderly individuals, therefore these appearances are not limited to cases of VaP (Lehosit and Leslie J. (2015)).

Non-motor VaP symptoms include gastrointestinal complications, urinary complications, apathy, respiratory disorders, and skin disorders. The three most common non-motor symptoms are fatigue, attention/memory impairment and psychiatric disorders (Colosimo et al. (2010)).

2.2.2 Risk Factors

As discussed, the underlying cause of VaP is vascular disease so the risk factors of VaP are similar to those of vascular disease. The most common vascular risks are diabetes mellitus and hypertension (Thanvi et al. (2005)). Hypertension was initially described in 1929 by Critchley but it was only recently concluded as an important risk factor of VaP. The connection between parkinsonism and other vascular risk factors such as smoking, high cholesterol, obesity, and cardiac disease has been studied recently. Studies concluded that cognitive and gait impairments are associated with the previously enumerated vascular risk factors (Pilotto et al. (2016), Malek et al. (2016)).

Various studies have provide evidences that VaP incidents increase with age, VaP patients usually display an older age onset than IPD patients (Demirkiran et al. (2001), Winikates and Jankovic (1999)). Vascular risk factors and a history of strokes are also more prevalent in VaP than in IPD patients (Demirkiran et al. (2001); Rektor et al. (2018); Winikates and Jankovic (1999)).

The presence of anticardiolipin antibodies and antiphospholipid antibodies has been associated with VaP (Huang et al. (2002)). These antibodies can affect the regulation of blood clotting, increasing the risk of blood cloth which increases the risk of stroke.

2.2.3 Diagnosis

Currently, there is no specific test/exam that can be performed to diagnose VaP. The neurologist, similar to the IPD procedure, needs to conduct a series of neurological examinations and analyse the patients' medical history to perform a diagnose. The evolution of radiology examination has allowed the conduction of Computerized Tomography (CT), Magnetic Resonance Imaging (MRI) and Dopamine Transporter Single Photon Emission Computed Tomography (DAT-SPECT). These exams support the diagnosis of VaP (Gupta and Kuruvilla (2011)). However, the diagnosis is still difficult since the found vascular lesions may also be present in radiological examinations of IPD patients (Lehosit and Leslie J. (2015)).

In 2004, Zijlmans et al. were the first to propose clinical criteria for the diagnosis of VaP. The criteria were based on the following three steps:

1. The patient had to display bradykinesia and at least one of the following: rest tremor, muscular rigidity or postural instability.

2. The patient needed to display cerebrovascular disease, this needs to be supported by brain imaging or the presence of symptoms that are consistent with stroke.

3. The relationship between parkinsonism and cerebrovascular disease aligned with one of the following: (1) an acute or delayed progressive onset; (2) an insidious onset of parkinsonism with extensive white matter lesions, and the presence of early shuffling gait or early cognitive dysfunction.

However, these criteria have not yet been validated on a large number of patients (Gupta and Kuruvilla (2011)).

As stated, the main problem when performing the diagnosis of VaP and IPD is the differentiation between these two disorders. In the last Section of this Chapter (Section 2.2.5) a closer analysis of the methods used to differentiate these diseases is presented.

2.2.4 Treatment

Since VaP is related to cerebrovascular complications vascular risk factors like hypertension, diabetes mellitus and smoking should be controlled. The treatment should also be directed at the prevention of stroke, this is an effort to stall the disease progression (Thanvi et al. (2005)).

The only treatment currently available that improves the parkinsonism symptoms is L-DOPA. However, the response to L-Dopa is usually poor, this may happen because lesions in VaP are mostly located in the basal ganglia and not in the nigrostriatal pathway (Gupta and Kuruvilla (2011)). A study showed that when the VaP patient has complications in the nigrostriatal pathway (one of the four dopaminergic pathways) the response to L-Dopa is usually positive (Zijlmans et al. (2004)). In 2017, Miguel-Puga et al. performed a systematic review of the response rate of L-Dopa. This study concluded that the L-Dopa response rate in VaP patients was of 30.4% (95% confidence interval of 23% to 38.8%). It was also concluded that the odds for a good response to levodopa in VaP increases by 15.5% when the lesions are in the nigrostriatal pathway. Despite the low response rate it is recommended that every VaP patient must be administered with L-Dopa for at least 3 months to conclude the response level (Miguel-Puga et al. (2017), Gupta and Kuruvilla (2011)). The use of dopamine agonists should not be an option because VaP patients are more susceptible to their side effects. Deep brain stimulation is also not an option for the treatment of VaP (Gupta and Kuruvilla (2011)).

A study has also been conducted to assess the potential of drainage of cerebrospinal fluid (CSF) via lumbar puncture (LP) (Ondo et al. (2002)). This study conducted the procedure on 40 patients and concluded that 15 VaP patients had a significant gait improvement after the procedure, 12 had no effect and 13 had mild improvement. This is a promising treatment approach that needs further investigation.

The assistance of VaP patients through behavior therapy is also very important because patients' gait might be limited by the fear of falling (Thanvi et al. (2005)).

2.2.5 Differential diagnosis

According to Gupta and Kuruvilla, the most important task when diagnosing VaP and IPD is the differentiation between these disorders, due to prognostic and therapeutic implications (Gupta and Kuruvilla (2011)). The differential diagnosis between IPD and VaP can be extremely difficult because these diseases, especially at early stages, present highly variable and overlapping phenotypes (Ali and Morris (2015)). Additionally, there is no specific procedure or diagnostic exam that can be performed to detect these diseases. Diagnosis is based on medical history, neurological examination, brain imaging, movement examination and proper accompaniment of the patient. Since the diagnosis is observational a high level of expertise is necessary to perform an early and correct diagnosis. An early and correct diagnosis of these diseases is extremely important because of the following reasons: treatment approaches differ according to the disease; disease progression also varies which may lead to an unexpected prognostic; neuroprotective interventions depend on early diagnosis to significantly lead to better long term outcomes (Lehosit and Leslie J. (2015)).

The most evident traits of VaP that differentiate it from IPD are the early postural instability and falls, gait instability with shuffling steps, a variable stride length, absence of festination (involuntary quickening of gait), presence of pyramidal signs (loss of reflexes), and early cognitive impact. When the patient presents traits that point to VaP these can be supported by neuroimaging that shows white matter lesions and/or strategic subcortical infarcts. It is possible that IPD and VaP symptoms can overlap but the primary pathological findings in VaP should be different than the diagnostic criteria of IPD. For example, VaP patients rarely show a loss of neurons in the substantia nigra, on the contrary, IPD patients show a significant loss of neurons in this region. (Lehosit and Leslie J. (2015), Thanvi et al. (2005), Gupta and Kuruvilla (2011), Zijlmans et al. (2004)).

Lately, there has been an increased interest in gait assessment for the diagnosis of IPD. Growing evidence shows that gait assessment can be a powerful biomarker in the diagnosis and management of patients with Parkinsonism related disorders (Lord et al. (2013), Kubota et al. (2016)). Contrarily to neuroimaging, gait patterns can be measured easily, quickly, and at low cost using non-invasive technology like wearable sensors. Gait disturbance is the first symptom in 90% of the cases of VaP patients (Fitzgerald

and Jankovic (1989)). Studies have also shown that stride length in VaP gait is usually short, the velocity is slower when comparing to IPD, and VaP patients have a tendency to present a characteristic shuffling gait (Winikates and Jankovic (1999), Bäzner et al. (2000)).

Recently, various studies have used gait assessment to differentiate between IPD and controls, these studies achieved promising results. The methods used by these studies were mainly the following, Support Vector Machines (SVM), Random Forest (RF) and Artificial Neural Networks (ANN) (Md. Tahir and Manap (2012), Nancy Jane et al. (2016), Rovini et al. (2018), Alam et al. (2017)). In late 2018, the first effort to differentiate between IPD and VaP using gait assessment and Artificial Neural Networks (ANN) has been made with the first results of this dissertation (Fernandes et al. (2018)). This initial study has shown that gait has the potential to be a powerful biomarker when differentiating between IPD and VaP.

Different brain exams like Positron emission tomography (PET) and DAT-SPECT have been proven as great complements along side MRI for the diagnosis of IPD and VaP. Findings of these scans can support the diagnosis since lesions are much more frequent in VaP patients (Gupta and Kuruvilla (2011), Heiss and Würker (1999)).

In 2001 Korten et al. investigated the prevalence of stroke in IPD, it was concluded that the history of strokes in IPD is lower than in controls. This study hypothesized that the probability of stroke decreases with the level of dopamine in the brain (Korten et al. (2001)). However, another study has shown that the risk of stroke is higher in IPD patients than controls (Jellinger (2003)).

In the following Table a summary of three studies that analysed neuroimaging of VaP and IPD patients is presented.

Table 2: Brain imaging studies comparing the presence of lesions in VaP and IPD patients

Study	**Group Size**	**Normal vs. Abnormal Findings**	**Location Affected Zones**
Rampello et al.	32 IPD Patients, 45 VaP Patients	75% of IPD patients present a normal scan while 0% of VaP patients present a normal scan.	88% of VaP patients and 7% of IPD patients show white matter lesions. 38% of VaP patients and 7% of IPD patients present lesions of the basal ganglia.
Demirkiran et al.	50 IPD Patients, 16 VaP Patients	70% of IPD patients present a normal scan while 0% of VaP patients present a normal scan.	87.5% of VaP patients and 16% of IPD patients show white matter lesions. 37.5% of VaP patients and 8% of IPD patients present lesions of the basal ganglia.
Winikates and Jankovic	175 IPD Patients, 69 VaP Patients	57.1% of IPD patients present a normal scan while 0% of VaP patients present a normal scan	95.7% of VaP patients and 22.3% of IPD patients present ischaemic lesions. 75.4% of VaP patients and 16% of IPD patients show white matter lesions. 44.9% of VaP patients and 4.6% of IPD patients present lesions of the basal ganglia.

2.3 Fabry Disease

2.3.1 Characteristics

Fabry disease is an X-linked lysosomal storage disorder caused by deficient or absent activity of the enzyme α-galactosidase A (α-Gal A) due to a variety of gene mutations, particularly the GLA gene. Reduced or absent activity of this enzyme leads to the progressive accumulation of glycosphingolipids, primarily globotriaosylceramide (Gb3), in cells and organs throughout the body, leading to malfunction of the different body systems and ultimately to premature death around 50 years old due to renal, cardiac or cerebrovascular complications (Giugliani et al. (2016)).

Since this disease is an X-linked disorder (the GLA gene is located on the X chromosome) previously it was thought that FD only impacted males, but the Fabry Outcome Survey (FOS) has confirmed that females are not only carriers of FD but also similarly affected (Mehta et al. (2004)). The classic FD phenotype is estimated to be present in 1 out of 37 104 male newborns according to a study conducted with Italian newborns (Spada et al. (2006)), a study conducted with Japanese newborns concluded that this phenotype is present in 1 out of 7057 newborns (Inoue et al. (2013)). Additionally, 1 out of 4600 Italian male newborns and 1 out of 3024 Japanese newborns presented low α-Gal A activity and tested positive for "atypical" FD mutations that might lead to the development of FD.

In 2009 Mehta et al., conducted a study to understand the main death causes of FD patients. This study concluded that after 2001 cardiac disease was the main cause of premature death of FD patients both in patients who had and had not received enzyme replacement therapy (ERT), the most common cardiac symptom that these patients displayed was left ventricular hypertrophy (LVH). These patients also displayed other symptoms alongside cardiac manifestations, being the most common proteinuria, stroke and transient ischemic attacks (TIA) (also known as mini-strokes). Before 2001 the main death cause of FD patients was renal failure. These results indicated that the importance of renal disease is decreasing, this may be because ERT (became only available after 2001) was able to stabilize renal complications. Contrarily the importance of cardiac disease has been increasing, most mortality cases of FD patients seem

to result from the progression of cardiac manifestations. Recently, in 2018, a study conducted by Brito et al. based on data extracted from the Portuguese Registry of Hypertrophic Cardiomyopathy (Cardim et al. (2018)) focused on bringing awareness of FD to cardiologists (Brito et al. (2018)). Cardiac involvement in FD is frequent and sometimes the only severe manifestation of this disease, so cardiologists need to be aware of the possibility of FD because the early initiation of ERT may improve and can even change the course of the disease. According to Linhart et al. (2007), the mean age onset of cardiac manifestations in FD patients is 31.5 and 39.9 years old, in males and females, respectively. This study was conducted on 714 FD patients where the most common cardiac manifestation was LVH, other manifestations include myocardial fibrosis, palpitations, arrhythmia, and dyspnea. There were no patients with exclusively cardiac manifestations and the frequency of cardiac symptoms was similar in both men and women, however, men displayed an earlier age of onset. Left ventricular hypertrophy was correlated with age, gender, and renal function. The presence of LVH was also associated with a higher frequency of other cardiac manifestations, 89% of patients with LVH displayed other cardiac symptoms. Linhart et al. hypothesised that these other cardiac symptoms may have led to the develop of LVH (Linhart et al. (2007)).

The most frequent symptoms of FD are neurologic, being the most common neuropathic pain and acroparesthesia, the most devastating neurologic consequence of FD is stroke, which occurs more frequently in FD patients than in healthy subjects, and is normally seen alongside cardiovascular manifestations like LVH and hypertension (Giugliani et al. (2016), Rolfs et al. (2005)). Various studies have also shown that glycosphingolipid storage also occurs in neurons of brain regions affected in neurodegenerative diseases such as the substantia nigra (de Veber et al. (1992), Kaye et al. (1988)). More than 50% of FD patients show an abnormal MRI scan, all patients with an abnormal scan show white matter lesions and 32% of patients with abnormal scan show subcortical gray matter lesions. A direct correlation between age and white matter lesions was also found, the number of lesions and severity increased with age (Ginsberg et al. (2006)). Unfortunately, ERT does not seem to reduce the risk of stroke in FD patients (Giugliani et al. (2016)). A study conducted on 167 subjects (110 FD patients and 57 age-matched controls) by Lohle et al. (2015) concluded that FD patients displayed a slower gait, transfer speed, poorer fine manual dexterity, and lower hand speed. These motor symptoms were related to cerebrovascular disease and were significantly correlated with disease severity and were similarly displayed by male and female patients (Lohle et al. (2015)).

2.3.2 Risk Factors

As mentioned before, FD is an X-linked lysosomal storage disorder, this means that the GLA gene whose mutations cause FD is located on the X chromosome.

The human being has 23 pairs of chromosomes, the 23rd pair is composed of two copies of the X chromosome or one X and one Y chromosomes depending on the gender of an individual. Males have one X chromosome and one Y chromosome while females have two X chromosomes. Fabry disease can be inherited, a male FD patient is always going to pass his affected X chromosome to his daughter. If a male FD patient has a son the X chromosome is not passed, the Y chromosome is passed instead, which is not a problem. A female FD patient has a 50% chance of passing her affected X chromosome to her son or daughter (Foundation (2019)).

Mutations of the GLA gene can also occur naturally leading to FD. Since female patients have two X chromosomes the disease onset might be delayed when compared to male patients. However, various studies have concluded that females are also severely impacted by this disease even though symptoms onset is usually delayed (Wilcox et al. (2008)).

2.3.3 Diagnosis

Due to rare occurrence and symptom variability, this disease is frequently misdiagnosed. According to Giugliani et al. (2016) the delay from onset to a correct diagnosis is 13.7 and 16.3 years, in male and female patients, respectively (Giugliani et al. (2016)). The early and correct diagnose of this disease is imperative, various studies have shown the importance of starting enzyme replacement therapy (ERT) as early as possible since serious complications are unlikely to develop when ERT is initiated before organ malfunction (Mehta et al. (2009, 2004); Shah and Elliott (2005)).

Since the most common manifestation of FD is neurologic a study conducted by Rolfs et al. (2005) on 721 subjects tried to determine the frequency of unrecognized FD in a cohort of cryptogenic stroke (stroke of unknown origin) patients. The results show that 28 (4%) of the 721 cases of stroke were undiagnosed FD patients. This study concludes that FD is very common in subjects who have suffered unexplained

strokes and that FD must be considered as the underlying cause in all cases of unexplained strokes in young subjects (between 18 and 55 years old) (Rolfs et al. (2005)).

2.3.4 Treatment

Until 2001, the treatment of FD was limited to the management of symptoms, medication that would manage or alleviate the various manifestations was the only treatment available. In 2001 enzyme replacement therapy (ERT) became available, agalsidase alpha and agalsidase beta are the two ERT procedures currently performed. The administration of these enzymes results in the reduction of glycosphingolipids in various organs. This procedure is currently the gold standard and it has demonstrated safety and good results in various studies. The treatment is able to achieve the best results with early initiation, this is before organs are critically affected. One major downside of this treatment is the cost, annually this treatment costs around 140 000 euros (Thadhani and Pastores, (2002)).

A study conducted in 2009 by Mehta et al. analysed a 5-year treatment based on ERT on 181 FD patients. The results of patients with cardiac hypertrophy were positive, reduction of the left ventricular mass (LVM) index and increment of the midwall fractional shortening (MFS) were observed. The LVM index and MFS of patients without hypertrophy remained stable in more than 90% of the patients. Other symptoms of FD were present in greater proportion in patients with LVH, this means that LVH is an important symptom of FD and is associated with other clinical events of FD disease. Renal function also improved in more than 50% of patients. This study also concluded that the treatment of patients with advanced cardiac and renal involvement was not optimal and that these conditions were not reversed by ERT. The treatment of pre-existing cerebrovascular disease also had little success. The study concludes that an early diagnosis is crucial for successfully treat FD (Mehta et al. (2009)).

Another study analysed the outcomes of a 10-year treatment of 45 adults suffering from FD with heart manifestations. This study showed that ERT was able to improve heart failure, stabilize or improve angina scores, there was no development of LVH in patients without LVH at the time of treatment initiation and there was improvement of LVH in patients with LVH at the time of treatment initiation (Kampmann et al. (2015)). Another study of 33 Spanish patients with FD also showed that ERT was able to stabilize LVH

(Rivera Gallego et al. (2006)). These studies show that ERT has long term positive benefits and is capable of improving or stabilizing cardiac manifestations. However, there is a real need to develop new kinds of treatment for FD, particularly treatments that are capable of stabilizing neurological manifestations (Mehta et al., (2009)).

2.4 Fabry Disease and Parkinsonism

As stated in Section 2.1, Parkinsonism has been previously associated with GBA mutations. These mutations cause Gaucher disease (GD), which is a disease that also results from the diminished activity of enzymatic activity, more specifically the glucocerebrosidase enzyme (Neumann et al. (2009)). Additionally, recent studies have also hypothesized a link between Parkinsonism and FD (Lohle et al. (2015); Wise et al. (2018); Wu et al. (2008)).

Fabry disease, as stated before, results due to GLA mutations which lead to diminished or absent activity of the enzyme α-Gal A. In 2008, a study conducted by Wu et al. showed that the activity of α-Gal A enzyme is lower in IPD patients than in controls, this study hypothesised that the decreased activity of this enzyme may impair the functions of the autophagic-lysosomal system which may lead to the aggregation process of α-synuclein, being this protein linked genetically to IPD (Wu et al. (2008)).

As also mentioned before, in 2015 Lohle et al. examined 110 FD patients and 57 match age healthy subjects, this study found that subjects with GLA mutations display symptoms of bradykinesia, slower gait, slower hand speed, and postural instability. However, these FD patients did not reveal significant cognitive deficiencies, hyposmia, REM sleep behavior or parkinsonian motor features that commonly precede neurodegenerative disorders like IPD. The study hypothesises that the storage of glycosphingolipids in the brain and the disruption of the autophagy-lysosomal pathway, which is responsible for digesting long-lived proteins, leads to a distinct phenotype with mild motor impairment and nonmotor symptoms such as depression and pain. The study also states that the contribution of FD to the development of neurodegenerative disorders should not be ruled out and that further research should be conducted (Lohle et al. (2015)). There have also been case reports describing FD patients that display parkinsonism symptoms

with a positive response to L-Dopa treatment (Buechner et al. (2006), Borsini et al. (2002)).

The most recent study connecting FD and Parkinsonism was conducted by Wise et al. (2018), this study pointed out that individuals with GLA mutations might be more susceptible to the development of Parkinsonism, it also points out that since strokes are common in FD patients these incidents may lead to the development of VaP, but further research is also required. The life expectancy of FD patients has lately been increasing due to ERT, so Wise et al. also hypothesizes that a link between Parkinsonism and FD was not uncovered in the past because FD patients were not able to live long enough to develop Parkinsonism (Wise et al. (2018)).

Chapter 3

Theoretical background

The big data era is upon us, every second almost 10 hours of video are uploaded to youtube (Hale (2019)), there are more than 1 billion websites (Deyan (2019)), and so on. This flood of data calls for methods that differ from traditional computing, the problem with traditional computing is that it suffers from a major limitation - human intelligence and ability to reason. Traditional computing requires the need to explain to the computer how a certain task is performed, to achieve this, a programmer needs to know how to perform the task so that it can explicitly code a series of steps that allow the computer to perform the specific task. The need to explain step by step to a computer how a task is performed is a serious limitation. To unlock their full potential computers must be able to accomplish tasks that humans can't explain explicitly. An example of such a task is recognizing pictures of different animals. It is extremely hard or even impossible to develop an algorithm that explicitly explains to a computer how to differentiate between a leopard, a tiger, a cheetah, and a jaguar. To solve this limitation of traditional computing a new field was developed, this field is Machine Learning (ML).

3.1 Machine Learning

According to Murphy (2012), machine learning provides a set of methods that are able to automatically detect patterns in data, these uncovered patterns are used to predict future data or to perform other kinds of

decision making (Murphy (2012)). Going back to the example of differentiating animals, the ML approach to solving this problem is to provide the computer with example pictures of the different animals and an ML method that detects patterns in data, the ML method might be Support Vector Machine (SVM), Convolutional Neural Network (CNN), among others. This technique of learning with labeled data is called supervised learning, there are two other main types of learning techniques, these are unsupervised learning and reinforcement learning. These learning techniques, still according to Murphy (2012), can be defined as follows.

Supervised learning is the task of learning from labeled training data, the goal is to learn the mapping from the inputs x to the outputs y, that is $y = f(x)$, so that predictions can be made on future data. When the output y is categorical the task is known as classification when y is real-valued the task is known as regression.

Unsupervised learning is, contrarily to supervised learning, the task of learning from unlabelled training data. The goal is to find patterns/correlations in data to obtain more knowledge. The most common unsupervised learning tasks are clustering and dimensionality reduction.

Lastly, **reinforcement learning** is the task of learning what actions to take based on reward and penalty signals. The system must learn by itself the strategy that gets the most amount of rewards. For example, this type of learning is implemented by robots to learn how to walk. Another recent and extremely good example of the power of reinforcement learning is the AlphaGo project by DeepMind's (Silver et al. (2017)).

Supervised learning and unsupervised learning are the two learning techniques that will be implemented in this work, as stated these techniques address different tasks, the most common can be categorized as classification, regression, clustering and dimensionality reduction.

In **classification** problems the output y is divided into multiple classes (two or more), such that $y \in \{0, ..., C\}$, where C is the number of classes. A classification problem can be binary ($C = 2$) or multiclass ($C > 2$). A simple example of a classification problem is email spam filtering, the output classes are spam ($y = 1$) or not spam ($y = 0$), the input is a representation of the email text and the goal is to classify the email into spam or not spam depending on the input.

A **regression** problem is also a supervised learning task, the output y of these problems is continuous, that is $y \in \mathbb{R}$. The goal of regression is to learn from data and then predict a real value for each input.

A simple example is predicting the stock price of the following day given the current market conditions.

Clustering is an unsupervised learning task, clustering techniques detect patterns in the input features with the goal of dividing the inputs into different subsets or groups, each input is given a specific group by the clustering method. The input samples within a cluster (group) are similar to one another and dissimilar to the samples of the other clusters. The concept of similarity is defined in terms of a distance function, for example, the euclidean distance. A simple example of a clustering problem is using social media data to identify communities within a city.

Dimensionality reduction is another common unsupervised learning task that tries to combine the input data and project it to a lower-dimensional subspace while capturing the structure of the data. Reducing the number of features may turn an intractable problem into a tractable one because, most of the time, there are input features that are redundant or irrelevant (just noise) to the problem being tackled. The goal is to select the features that describe most of the variability. Dimensionality reduction is often used to reduce the representation of data before supervised methods to improve development costs and improve convergence. Note that dimensionality reduction can also be considered a supervised learning task, this happens when the output is also used to select the most important features.

Clustering and dimensionality reduction techniques are often used for data mining purposes. Data mining, according to Han et al. (2011), focus on discovering relevant properties and patterns in data. Currently, enormous amounts of data are collected every day which creates a need for powerful and versatile tools that automatically uncover valuable information from the tremendous amount of data that is being collected. Data mining is a recent field that tries to provide tools to effectively transform vast amounts of data into organized knowledge (Han et al. (2011)).

Machine learning and data mining are often confused and used interchangeably, but these are two different fields that are used together to achieve better results. Data mining focus on extracting valuable insights from data. Machine learning main focus is the development of algorithms that are able to learn and make predictions without explicitly being told what steps to take. So machine learning uses data mining to provide meaningful data to learning algorithms and data mining uses machine learning methods like clustering to detect patterns in data. Data mining ultimately is a recent field whose dedication is retrieving knowledge from data (Han et al. (2011); Murphy (2012)).

3.2 Dimensionality reduction

Dimensionality reduction refers to a group of methods that can be applied to a dataset to reduce the number of input features while maintaining the relevant properties of the data. Whether it is a clustering, classification or regression task the input data must contain critical information about the problem we are trying to solve. However, the complexity and the high number of features of a dataset impact the computational cost of the learning algorithm and its performance. The ideal is to only use the features that are relevant to the problem at hand, but most of the time it is not possible to know a priori which input features are the most important, so having a separated preprocessing step (dimensionality reduction) before clustering, classification or regression tasks is of extreme importance for the following reasons (Alpaydin (2010), Mohri et al. (2012)):

- The complexity of learning algorithms is dependent on the number of input features and the sample size. The reduction of the data dimensionality improves memory consumption and computation cost which facilitates subsequent data operations.
- If an input feature is irrelevant it must be removed, this will improve the performance of learning algorithms and reduce the problem complexity.
- Low dimensionality usually leads to more robust and simpler models. These simpler models are less sparse because they are not modeling outliers and noise that may lead to over-fitting.
- The underlying process of a problem is easier to understand if explained with fewer features, this allows for greater knowledge extraction.
- Data visualization can be achieved if the data dimensionality can be successfully reduced to two or three dimensions, this improves exploratory analysis.

The ultimate goal of dimensionality reduction, as the name suggests, is to reduce the data dimensionality without losing relevant information. There are two main procedures for achieving dimensionality reduction, these are feature selection and feature extraction.

3.2.1 Feature selection

Feature selection seeks to reduce the data dimensionality by selecting k of the d features that are most important to the problem being tackled, the features that were not selected are removed and not used in subsequent operations. The goal is to find a subset of the original features that best explains the underlying pattern of the data. This technique is most useful when the data is composed of features that are redundant and irrelevant, these features can then be eliminated achieving the end goal of dimensionality reduction. It is important to note the difference between irrelevant and redundant features, a redundant feature is one that can be relevant but is strongly correlated with another relevant feature, an irrelevant feature is just random noise that does not contain any information about the data pattern (Alpaydin (2010), Zareapoor and K. R (2015)).

3.2.2 Feature extraction

Feature extraction does not seek the subset of features that best explain the data, this approach instead focuses on creating a new set of features that are a function/combination of the original features, the result is a more compact feature space. The original features are all used to create the new feature space (Alpaydin (2010)).

Both feature selection and feature extraction methods share the same goals of reducing the dimensionality of the feature space. Feature extraction techniques are most useful when individual features are not very discriminative of the data, instead, valuable information and patterns might be recognized when the features are combined (Alpaydin (2010)).

Feature extraction methods are mainly unsupervised, the most common and widely used are Principal Components Analysis (PCA) and autoencoders. An autoencoder is an artificial neural network architecture that is divided into two stages, an encoding stage and a decoding stage. The encoding stage converts the input data into a lower space while preserving critical information. The decoding stage takes the output of the encoding stage and converts it into a representation very similar to the original data. An autoencoder is basically trained to generate the input data from a lower-dimensional space (it tries to copy its input

to its output), this is called a generative model. In this process the autoencoder learns to preserve as much information as possible, the output of the encoding stage can then be used as a lower-dimensional representation of the original data. Principal Components Analysis (PCA) is the most common approach to dimensionality reduction. To reduce the number of features PCA learns a lower dimension representation of the input data, more specifically it learns an orthogonal, linear transformation of the original features whose result is a new set of features, smaller than the original, that capture the relevant structure of the data (Murphy (2012), Goodfellow et al. (2016)).

A detailed explanation of the dimensionality reduction methods used in this work is presented in Chapter 5.

3.3 Machine Learning applied to medicine

To better understand the current context of this work and the current research in the medical field related to IPD, VaP, FD and other neurodegenerative disorders several studies and research papers, that make use of the machine learning algorithms applied in this dissertation for the solution of several medical tasks, will now be referred and analysed.

Machine learning applied to the medical space is a big field, so it is not possible to conduct analyses over all the studies conducted in this space. A selection of the studies and papers that are most relevant to the current dissertation are presented next. The analyses of the researches are organized by the used machine learning algorithm.

3.3.1 Multiple Layer Perceptron

In 2010 a study conducted by **Joshi et al. (2010)** strived to distinguish between IPD and Alzheimer's disease based on the most influencing risk factors of these diseases. Data from 487 patients of which 60% suffered from AD and 40% from IPD was collected. The data was preprocessed and feature selection was

performed. The methods used for feature selection were chi-square, gini impurity, relief attribute selection, among others. The identified risk factors were age, diabetes mellitus, heart disease, hypertension, smoking, low-density lipoprotein, alcohol, body mass index (BMI) and genetic factors. After feature selection, the risk factors with the strongest association were age, genes, diabetes, smoking, and stroke. The selected data was then fed to various machine learning classifiers. The classifiers used were Random Forest (RF), Radial Basis Function Network (RBFN) and Multiple layer Perceptron (MLP). The learning algorithms that obtained the best classification results were the MLPs and the RFs, both achieved an accuracy of 99.25%.

A study conducted by **Manap et al. (2011)** attached to the participant's body 37 reflective markers that were traced using 6 infra-red cameras to collect gait, kinematic, and kinetic features from 12 IPD patients and 20 healthy subjects. The subjects were asked to walk at a self-selected speed on two embedded force plates while the infra-red cameras were collecting information. The walking task was conducted three times, and the best trial was selected.

The values of the ground reaction force were normalized according to the participants' weight. Feature selection was performed using statistical methods (independent t-test and Pearson correlation coefficient test), the significant level was $p < 0.05$. The implemented machine learning algorithm was an MLP, the number of epochs was 500, and the activation function of the hidden layer and the output layer were the hyperbolic tangent and the sigmoid functions, respectively.

The performance was estimated using a 4-fold cross-validation method. The task was to classify the individuals as IPD patients or healthy subjects. The features selected using the statistical methods were step length, speed, the maximum extension of hip angle, and vertical minimum mid stance force. With these 4 features as input, the MLP was able to obtain an accuracy rate of 95.63%. When using only all gait features the accuracy was 81.25%, using only all the kinetic features the accuracy was 81.25%, and using only all the kinematic features the accuracy was 84.38%.

In 2012 a study conducted by **Khemphila and Boonjing (2012)** used a Multiple Layer Perceptron to classify between IPD patients and healthy subjects based on voice features. The original dataset was composed of 22 features, these were reduced using gini impurity to 16 features. The number of participants was 31, out of which 23 were IPD patients. The best accuracy achieved by the MLP classifier was of 83.33%.

In 2016, **Perumal and Sankar (2016)** conducted a study whose goal was to analyse the features

that were most important in the early detection of IPD. The study extracted kinetic and spatial-temporal data from 93 IPD patients and 73 healthy control subjects and found the features that are most relevant for the differentiation between these two groups. Tremor features were also extracted and processed using machine learning techniques to assess the characteristics of IPD tremors. However, the tremor features were not used for classification.

To reduce the influence of bodyweight the data was normalized to the percentage of the subject bodyweight. Feature selection was conducted by performing a one-way analysis of variance (ANOVA) test, a p value less than 0.05 was considered significant. The most relevant features were step distance, stance and swing phase, heel and normalized heel forces. Various learning algorithms were used to classify between IPD and healthy subjects, including Linear Discriminant Analysis (LDA), SVMs and MLPs. The performance of the classifiers were evaluated using 5-fold cross validation, the learning algorithm that achieved the highest performance was LDA with an accuracy of 86.9%. Unfortunately, the performance of the SVMs and MLPs was not stated in the paper.

More recently, in 2018 a study conducted by **Wroge et al. (2018)** sought to use biomarkers derived from the human voice to differentiate between IPD patients and healthy subjects. The data was collected by Sage Bionetworks using an iPhone application. The smartphone collected the voice activity of the participants articulating the /aa/ phoneme for 10 seconds. Data from patients that were on medication was discarded. The voice data was preprocessed using a Python library called PyAudioAnalysis. Feature extraction was performed using an open-source tool denoted OpenSmile. Afterward, methods from AVEC (Audio-Visual Emotion recognition Challenge) were used to create the AVEC dataset and other methods were used to extract the GeMaps (Geneva Minimalistic Acoustic Parameter Set) features creating the GeMaps dataset.

Different machine learning algorithms were used to classify the participants. The algorithm with the best performance on the AVEC features dataset was the MLP with an accuracy of 86%, on the GeMaps features dataset the best learning algorithm was Gradient Boosted Decision Trees with an accuracy of 82%.

3.3.2 Support Vector Machines

A study conducted by **Md. Tahir and Manap (2012)** collected spatial-temporal, kinetic and kinematic data from 12 IPD patients and 20 healthy controls using 37 markers attached to the subjects' body, the markers were traced by infra-red cameras. Three data normalization techniques were applied to the collected data, these were ground reaction force data normalization, intragroup data normalization, and intergroup data normalization. Support vector machines were developed for classification with three types of kernel functions: linear, radial basis and polynomial. A Multiple Layer Perceptron was also implemented, the network had 11 hidden neurons and a single output neuron. The MLP was trained using backpropagation. The learning rate of the MLP was 0.3, the activation function of the hidden layer was hyperbolic tangent and the activation function of the output layer was sigmoid. The classification task consisted in the differentiation between IPD patients and control subjects, the performance of the classifiers was evaluated using a 4-fold cross-validation method. The MLP classifier was able to achieve an accuracy of 95.8%, however, the SVM was the superior classifier achieving an accuracy of 98.2%. The performance was higher when the input data was normalized based on the intragroup normalization method and the most important features were spatial-temporal.

Another study conducted in 2012 by **Orrù et al. (2012)** provided an overview of a vast number of studies that used SVMs to identify neuroimaging biomarkers of neurological and psychiatric disorders. This study analysed a total of 50 published studies that utilized a range of neuroimaging techniques including structural MRI, functional MRI, diffusion tensor imaging (DTI) and positron emission tomography (PET). The majority of the analysed studies focused on the classification of probable dementia of Alzheimer type (PDAT), the following Table shows a summary of the results of the various studies. All studies related to Parkinsonism are present in Table 3.

The study concluded that the published studies thus far indicate that the application of SVM and other learning algorithms to neuroimaging data may allow the prediction of patient condition as well as treatment response at the individual level. These algorithms could support clinicians at the early stages of the different disorders.

Table 3: Studies investigating the diagnostic potential of the SVM algorithm based on neuroimaging data.

Author	Comparison	Sample Size	Technique	Accuracy (%)
Duchesne et al. (2008)	PDAT vs. HC	PDAT = 75 HC = 75	Structural MRI	92
Kloppel et al. (2008)	PDAT vs. HC	PDAT = 20 HC = 20	Structural MRI	95
Kloppel et al. (2008)	PDAT vs. FTLD	PDAT = 18 HC = 19	Structural MRI	89.2
Plant et al. (2010)	PDAT vs. HC	PDAT = 32 HC = 18	Structural MRI	90
Plant et al. (2010)	MCI vs. HC	MCI = 24 HC = 18	Structural MRI	97.62
Chen et al. (2011)	PDAT vs. HC	PDAT = 21 HC = 20	Resting-state functional MRI	87
Chen et al. (2011)	MCI vs. HC	MCI = 15 HC = 20	Resting-state functional MRI	95
Zhang et al. (2011)	PDAT vs. HC	PDAT = 51 HC = 52	Structural MRI and PET	93.2
Zhang et al. (2011)	MCI vs. HC	MCI = 99 HC = 52	Structural MRI and PET	76.4
Ecker et al. (2010)	HC vs. ASD	ASD = 20 HC = 20	Structural MRI	85
Duchesne et al. (2009)	IPD vs. PPS	IPD = 16 PPS = 16	Structural MRI	90.6
Focke et al. (2011)	IPD vs. PSP	IPD = 21 PSP = 10	Structural MRI (GM)	87.1
Focke et al. (2011)	IPD vs. PSP	IPD = 21 PSP = 10	Structural MRI (WM)	96.77
Focke et al. (2011)	IPD vs. MSA	IPD = 21 MSA = 11	Structural MRI (GM)	71.87
Focke et al. (2011)	IPD vs. MSA	IPD = 21 MSA = 11	Structural MRI (WM)	Bad
Focke et al. (2011)	IPD vs. HC	IPD = 21 HC = 22	Structural MRI (GM or WM)	Bad

ASD - autistic spectrum disorders; FTLD - frontotemporal lobar degeneration; HC - healthy controls; MCI - mild cognitive impairment; PDAT - probable dementia of Alzheimer's type; IPD - Idiopathic Parkinson's Disease; PPS - Parkinson Plus Syndromes; PSP - Progressive Supranuclear Palsy; n.s. - nonsignificant; GW - gray matter; WM - white matter.

Also in 2012, a study conducted by **Haller et al. (2012)** used brain MRI imaging to differentiate between IPD patients and patients with parkinsonism characteristics. The MRI technique used was diffusion tensor imaging. The study included 17 IPD patients and 23 other parkinsonism patients (3 suffered from VaP), the groups were age and gender matched. The data was preprocessed with a tract-based spatial statistics (TBSS) technique using the FSL software library. The learning algorithm used was SVMs and the performance was measured using 10-fold cross-validation. The best SVM classifier used a radial basis kernel function and achieved an accuracy of 97% $\pm$ 7.54%.

In 2015, **Xia et al. (2015)** investigated the potential of features extracted from gait rhythm signals for the classification of patients with neurodegenerative diseases (NDD) and normal subjects. The gait dataset was composed of 19 subjects with Huntington's disease (HD), 13 subjects with Amyotrophic lateral sclerosis (ALS), 15 subjects with IPD and 16 healthy controls. The subjects were asked to walk at their normal pace along a straight 77 meters long hallway for 300 seconds. The gait signals were measured with ultrathinforce-sensitive switches placed inside each subjects' shoes. The resulting gait signals were processed by several feature extraction methods resulting in 9 groups of gait features. Four different machine learning classifiers including SVMs, MLPs, RFs, and K-Nearest Neighbors (KNNs) were implemented using the different set of extracted features, the best classifier for the classification of healthy subjects vs. NDD was the SVM classifier achieving a performance of 96.83%, the performance of the classifiers was evaluated using leave-one-out cross-validation. The SVM was also the superior classifier when classifying IPD vs. healthy subjects achieving an accuracy of 100%, and when classifying HD vs. healthy subjects achieving an accuracy of 100%. The results of this study have made meaningful contributions to the diagnosis of NDD based on gait features extracted using non-invasive technology.

In 2014 a study conducted by **Shahbakhi et al. (2014)** proposed the differentiation between IPD patients and healthy subjects based on voice analysis. The voice signals were collected at the University of Oxford in an IAC sound-treated booth using a head-mounted microphone. The participants were 23 IPD patients and 8 healthy people, these were asked to pronounce the letter "a" for 3 seconds. A total of 195 vocal samples were collected for classification and all samples were normalized in amplitude. After data preprocessing the number of extracted features was 22, these were reduced using a feature selection method called genetic algorithm (GA). The learning algorithm used for classification was SVM with a radial basis kernel function. The dataset was divided into a training set (75%) and a test set (25%), the best results were an accuracy of 94.50% using the following 4 features, Fhi (Hz), Fho (Hz), jitter (RAP) and shimmer (APQ5).

3.3.3 Recurrent Neural Networks

In 2009, **Dutta et al. (2009)** developed a Recurrent Neural Network (RNN) to perform automatic identification between healthy gait patterns and neurological disorders gait patterns. The collected data was composed of 16 healthy subjects, 15 IPD patients, 20 Huntington's disease (HD) patients, and 13 Amyotrophic lateral sclerosis (ALS) patients. The subjects were asked to walk at their normal speed a 77m long hallway. To measure the gait data force-sensitive insoles were placed in the shoes of the subjects. The study considered the gait time series and extracted features from the stance, swing and double support measurements using cross-correlation based feature extraction, a total of 15 features were computed using this technique. The model proposed by this study was configured to be a four-class classification model, the four classes corresponded to healthy subjects, IPD patients, HD patients, and ALS patients. The model worked in two separated steps, the first step was the classification of the subject as healthy or neurological disorder patient, if the subject was classified as a neurological patient it would be further evaluated in the second step where the differentiation between the three neurological disorders was made. The overall accuracy of the proposed model was of 90.2%. The accuracy of the first and second steps was 90.6% and 89.8% respectively.

In 2016, **Choi et al. (2016)** developed a predictive model that uses data from the electronic health record (EHR) to predict the medical condition and medication dosage of subjects. The developed algorithm was a temporal model that used RNNs to analyse the time-stamped EHR data. The primary goal of the study was to use the EHR data to predict the medical conditional and the recommended medication, this study had the secondary goal of predicting the time to the patient's next visit to their doctor. The data used in this work was collected from the Sutter Health Palo Alto Medical Foundation and it consisted of 263706 patients with an average of 54.61 visits per person. The output label was composed of 1183 ICD-9 (International Classification of Disease Ninth Revision) codes that represent a medical condition and 595 GPI medication codes, in total the output vector had a dimension of 1778. The type of RNNs used was gated recurrent units (GRUs) and the data was divided into 85% for training and 15% for testing, different numbers of hidden neurons were implemented, these varied from 100 to 2000. The performance was measured using the Top-k recall. When predicting only the medical condition the performance of the

RNN was 79.58% recall@30, when predicting only the medication it was 85.53% recall@30. When the model was used to predict the medical condition, the medication and the visit time simultaneously the performance was of 72.48% recall@30.

Another study conducted in 2017 by **Che et al. (2017)** proposed a deep learning model that learns patient similarity from multi-modal longitudinal medical records using an RNN. The goal of this study was to learn the similarities between patient records using medical records collected through the years, the networks tried to match temporal patterns of different patient sequences. The result is a similarity metric that can be used to perform personalized predictions. To evaluate the performance of the proposed architecture the authors collected data from 683 IPD patients and 217 control subjects. The data was preprocessed, and missing values were filled using the last occurrence carry forward strategy. The number of features was 319 from 7 domains categories: 1) motor symptoms/complications, 2) cognitive functioning, 3) autonomic symptoms, 4) psychotic symptoms, 5) night-time sleep problems and daytime sleepiness, 6) depressive symptoms, and finally, 7) Anxiety and Depression scale. Several learning algorithms were evaluated while predicting using all the data and only using the top K similar subjects based on the RNN similarity metric. The metrics used to evaluate performance were precision, recall, and F1-score, the learning algorithms that were trained using the top K similar subjects outperformed the traditional methods, the best result was achieved by the RNN with an SVM classifier, with an F1-Score of 0.77, the F1-Score of the SVM without the RNN step was of 0.75.

Also in 2017, **Lipton et al. (2017)** conducted a study that focused on using data from electronic health records. The data was collected from caregivers in the pediatric intensive care unit (PICU), the observations included sensor data, vital signs, lab test results and subjective assessments. The number of collected features was 13, these include body temperature, heart rate, diastolic and systolic blood pressure, among others. The problem was formulated as multilabel classification, the output vector is composed of 128 ICD-9 codes, these indicate conditions such as acute respiratory distress, congestive heart failure, seizures, and renal failure. The total number of PICU episodes is 10 401, these were divided into a training set (80%), a validation set (10%) and a test set (10%). The validation set was used to tune the hyperparameters of the RNN, the type of RNN used was long short-term memory (LSTM). The final model was composed of 2 hidden layers with either 64 hidden neurons per layer with no dropout rate or 128 hidden neurons per layer with a dropout rate of 0.5. The performance metrics used were micro AUC

and macro AUC. The LSTM models were compared to other learning algorithms like MLP and Logistic Regression. LSTM was able to outperform the other learning algorithms achieving a Micro AUC of 0.8560 and a Macro AUC of 0.8075.

3.3.4 Deep Belief Networks

In 2014, **Liu et al. (2014)** proposed a deep learning architecture to perform an early diagnosis of Alzheimer's disease (AD). The early diagnosis of AD plays a significant role in the progression of the disease and allows the implementation of measures that manage the disease impact. This study proposed a deep learning architecture based on Deep Belief Networks (DBNs) to classify between AD patients and healthy subjects. The network was also able to predict the risk of Mild Cognitive Impairment (MCI) subjects evolving to AD patients, the instances of MCI can be classified as MCI non-converters (ncMCI) or MCI converters (cMCI). The collected data were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI). The data was composed of MRI images from 65 AD patients, 67 cMCI patients, 102 ncMCI patients, and 77 healthy subjects. The grey matter volumes and cerebral metabolic rate of glucose consumption (CMRGlc) patterns were extracted from the MRI data, all data were normalized before the classification task. The performance metrics used were accuracy, sensitivity, and specificity. An SVM model was developed to serve as a baseline. Both classifiers were evaluated using 10-fold cross-validation. The classifiers were used for three different tasks, these were the classification of AD vs. Controls, MCI vs. Controls and Controls vs. cMCI, ncMCI and AD. The results show that the DBN model is superior to the SVM model at every task. On the first task, the DBN model was able to achieve an accuracy of 87.76%, a sensitivity of 88.57% and a specificity of 87.22%. On the second task, it was able to achieve an accuracy of 76.92%, a sensitivity of 74.29% and a specificity of 78.13%. On the third task, the performance was an accuracy of 47.42%, a sensitivity of 65.71% and a specificity of 83.75%.

In 2015, **Zhang et al. (2015)** tackled the problem of real-time activity recognition, they proposed a DBN model that analyses raw input data from accelerometers of a smartphone and classifies them as one of the following seven activities: walking, running, standing, sitting, lying, walking upstairs, or walking downstairs. The developed DBN model was firstly trained offline and then implemented on a Samsung

Galaxy Note 3 to perform real-time activity recognition. The authors used a Samsung smartphone to collect the accelerometers data while the subjects were performing different activities, the phone was positioned in the right pant pocket of the subjects. After collection, the data was split into frames, the frame length was 1s. The data was then split into training data (80%) and test data (20%). The final result was 42000 frames for training and 10100 for testing. Other learning methods such as SVM, KNN, RF, and MLP were also developed to serve as a baseline. The performance metric used was the error rate, the DBN classifier performed the best achieving an error rate of 0.014, the second-best classifier was the KNN achieving an error rate of 0.024, the MLP was able to achieve an error rate of 0.041.

In 2016, **Alsheikh et al. (2016)** considered the problem of using data from triaxial accelerometers to recognize human activity. Currently, accelerometers are present in almost all mobile phones and wearable devices so there is a big opportunity to work with this type of data. This work made use of deep learning architecture, more specifically a DBN to detect human activity using triaxial accelerometer data. The developed DBN was trained using the accelerometer data in two stages, the first stage was unsupervised learning where the weights of the network were initialized, the second stage was supervised fine-tuning. Three public datasets were used in this study, these were the WISDM Actitracker dataset (data from 29 subjects), the Daphnet freezing of gait dataset (data from 10 subjects) and the Skoda checkpoint dataset (data from 1 subject). The developed DBN model was able to outperform former learning approaches used to tackle these datasets. On the WISDM dataset, it beat the performance of an ensemble learning method by 3.93% achieving an overall accuracy of 98.23%. On the Daphnet dataset it was able to achieve an accuracy of 91.5% beating the former best learning algorithm (KNN) and on the Skoda dataset it was able to achieve an accuracy of 89.38%, the former best learning algorithm was a Hidden Markov Model (HMM) with an accuracy of 86%.

In 2016, a study conducted by **Costa et al. (2016)** at the University of Minho investigated if machine learning classifiers were able to diagnose Alzheimer's disease (AD) based on postural control kinematics and the Montreal Cognitive Assessment (MoCA). Noninvasive diagnostic tools that aid in the diagnosis of AD are highly requested. Inertial measurement units are inexpensive, portable and noninvasive, these devices are able to collect kinematic data that can be used in the diagnosis of AD. Four machine learning classifiers were used in this study to assess the discriminative power of postural control kinematics, these were: SVMs, MLPs, RBFNs, and DBNs. The collected data included 36 AD patients and 36 healthy

controls, because of significant differences between the anthropometric data the kinematic variables were adjusted according to age, education, height, and weight. Eighteen kinematic features were collected, the MoCA variable was also collected. The data were divided into a training set (50%), a validation set (10%), and a test set (40%). The features were ranked using Mann-Whitney U-test and incremental error analysis was performed to determine the number of features that produce the best results. The classifiers were evaluated based on 3-fold cross-validation. The best classifier was the MLP with an accuracy of 86% when trained only on the kinematic variables. Including the MoCA feature resulted in an accuracy of 96.6% was achieved. The second-best classifier was the DBN with an accuracy of 78% when trained on the kinematic features and an accuracy of 96.5% when the MoCa feature is included.

3.3.5 Convolutional Neural Networks

In 2016, **Sathyanarayana et al. (2016)** assessed the feasibility of predicting sleep quality given the physical activity records of an individual during awake time. Physical activity and sleep are highly interrelated health behaviors. The awake time physical activity has an impact on sleep quality and vice versa. This study collected data from 92 adolescents over 1 full week. Each participant was provided with an ActiGraph GT3X+ device, placed on their nondominant wrist, the device was water-resistance and must be worn at all times for the full week. The sleep quality was accessed by determining the ratio of total minutes asleep to total minutes in bed if this ratio is greater than 85% the sleep quality is good, otherwise is bad. Various neural network architectures were used to classify the sleep quality given the physical activity data, these include MLPs, CNNs, simple RNNs, and LSTMs. The awake time data was then used to predict sleep quality. To evaluate the model's performance several metrics such as accuracy, precision, recall, F1-score and area under the curve (AUC) were calculated. The best model differed according to the performance measure, LSTMs were able to achieve the highest AUC (97.14%) and recall (100%). However, CNN achieved the highest F1-score (94.44%) and accuracy (92.86%). The best precision measure (93.94%) was achieved by the MLP architecture. The simple RNN was the worst model, this is not surprising because this architecture suffers from the vanishing gradient problem.

In 2017, **Rajpurkar et al. (2017)** developed a 34-layer CNN that maps a sequence of electrocar-

diogram (ECG) samples to rhythm classes to detect irregular heart rhythms also known as arrhythmias. The collected dataset was composed of 34121 ECG records from 29163 patients, the ECG records were recorded using a non-invasive and continuous monitoring device called Zio Patch. Each ECG record was 30 seconds long and the number of rhythm classes was 14. Patients with normal rhythms were included in the 29163 patients so that the number of occurrences of each rhythm class is balanced. The dataset was split into a training set (90%) and a validation set (10%), the split was performed so that there is no patient overlap between the training and validation set. A test set was then collected from 328 unique patients resulting in 336 records, these records were classified by a committee of three board-certified cardiologists. Then, each record in the test set was classified by 6 individual cardiologists not participating in the committee group. The performance of these 6 cardiologists was averaged and compared against the CNN model. The CNN was able to outperform the individual cardiologists, the obtained results were the following (CNN vs. Cardiologists), precision: 0.800 vs. 0.723, better recall: 0.784 vs. 0.724 and better F1 score: 0.776 vs. 0.719.

In 2018, **Pereira et al. (2018)** collected time-series data for IPD identification. The data was feed into a CNN that detected features from the time-series and classified the individuals as healthy or IPD patients. The data were collected from six different handwritten activities, these include drawing circles on a paper, drawing circles in the air, drawing spirals, drawing meanders, left-wrist movements, and right-wrist movements. The individuals performed the activities with a smartpen that contains six sensors, these are a microphone, finger grip, axial pressure of ink refill, tilt and accelerations in the X direction, tilt and accelerations in the Y direction and tilt and accelerations in the Z direction. The control group was composed of 20 healthy individuals and the patients' group was composed of 14 individuals. The activities were performed more than once by the individuals, resulting in a total of 1512 time-series. This data was divided into a training set (50%) and a test set (50%), to deliver consistent experiments the partition considered the individuals, that is every time-series of an individual must be either on the training set or in the test set so there's no overlap. To access the performance of the network each activity was processed by the CNN independently and the final result was obtained by majority voting. The best performance obtained by the CNN over the time-series was an accuracy of 93.49%.

In 2018, **Oh et al. (2018)** developed a detection system for IPD using CNNs and electroencephalogram (EEG) signals. As discussed IPD arises when dopaminergic neurons start to die, so EEG signals

might reveal features that are able to differentiate IPD patients from healthy individuals. The data was composed of 20 IPD patients and 20 healthy subjects. The EEG records lasted for 5 minutes, these were collected using an emotive EPOC neuroheadset with 14 channels. The collected data was then composed of 14 features, each feature represents a channel. The developed CNN was designed in Python using Keras, the model was composed of 13 layers and the optimization method was Adam. The data was split into a training set (80%) and a test set (20%). The training data was used to fine-tune the models using the performance results of 10-fold cross-validation as a guideline. The best model was able to achieve an accuracy of 88.25%, a sensitivity of 84.71% and a specificity of 91.77% on the test set.

3.3.6 K-Means and Hierarchical Clustering

In 2008, **Post et al. (2008)** investigated clinical heterogeneity of newly diagnosed IPD patients using cluster analysis to describe the subsets in terms of different symptoms. The dataset was composed of 133 patients collected from hospitals in the Netherlands, the exclusion criteria were age greater than 85 years old. This work used the K-Means clustering method, and Euclidean distance was used to compute the within-cluster variance. The data were standardized before the cluster analysis. The different clusters were compared in terms of impairment, disability, perceived quality of life and use of dopaminergic therapy. The number of optimal clusters was chosen to be two and three, the differences between the values of the different clusters were analysed using Mann-Whitney U-test for the two-cluster case and Kruskal-Wallis for the three-cluster case. When the number of clusters was chosen to be two the patients were divided into a group of younger patients and a group of older patients. When the number of clusters was three, the results show an intermediate group with respect to age. In both cases the older the onset group the higher the progression rate and the level of motor impairment. In the three cluster case, the intermediate group showed higher anxiety and depressive symptoms. The results show that increased age at onset was correlated with higher disease severity and lower levels of perceived quality of life.

In 2012, **Liepelt-Scarfone et al. (2012)** used cluster analysis to investigate the heterogeneity of IPD. Clustering was used to detect different cognitive profiles and stages to predict patients with a potential risk of developing dementia. The dataset was composed of 121 IPD patients, of these 121 patients, 24 had

dementia. Exploratory Factor Analysis (EFA) was performed on the collected data to identify cognitive domains. The results of EFA were six Factors representing different cognitive domains, Factor 1 represented the frontal lobe function, Factor 2 represented word-memory and recall, Factor 3 represented attention, Factor 4 represented logical memory, Factor 5 represented praxis and visual perception and lastly Factor 6 represented fluency and naming ability. Hierarchical clustering was performed twice, (1) on the whole data (121 IPD patients) and (2) on the IPD patients that didn't suffer from dementia (n = 97). Both analyses revealed two clusters regarding the cognitive domains defined by the EFA. The first hierarchical cluster analysis assigned all 24 IPD patients with dementia to cluster 2, this cluster represented poorer neuropsychological performances. The second hierarchical cluster analysis replicated the grouping in 92.8% of all IPD patients without dementia, all the IPD patients with no dementia that were included in cluster 2 in the first analysis were not regrouped to cluster 1. The cluster analysis resulted in a clear-cut division of the six cognitive domain Factors. The IPD patients with lower value of Factor 3, Factor 4 and Factor 5 were assigned to cluster 1. Cluster 2 was composed with patients that showed lower values of Factor 1, Factor 2 and Factor 6. The second cluster analysis results only differed for Factor 1.

In 2017, **Mu et al. (2017)** also investigated the clinical heterogeneity of IPD. This work used cluster analysis to search for IPD subtypes from a large cohort of patients across all stages of the disease using a combination of motor and non-motor features. The dataset was composed of 951 IPD patients. All the features were standardized before the cluster analyses and the chosen clustering method was K-Means. The first clustering analysis was performed using nine non-motor symptoms (NMS) domains, the four cardinal motor signs (tremor, bradykinesia, rigidity, postural instability) and motor complications, this was called the "domains clustering". The second clustering analysis was performed only using non-motor symptoms (NMS) and it was called "symptoms clustering". The number of clusters K chosen for the "domains clustering" was 4, cluster D1 included 428 patients mildly affected in all domains, cluster D2 included 180 patients severely affected in non-motor domains but mildly affected in motor domains, cluster D3 included 232 patients severely affected in motor domains but mildly affected in non-motor domains and cluster D4 included 64 patients severely affected in all domains, these patients presented the greatest symptoms severity. The optimal number of clusters for the "symptoms clustering" was 6. Cluster S1 was the largest cluster with 456 patients, these patients were mildly affected in all NMS, cluster S2 included 201 patients and the symptom severity was greater than cluster S1 but could not be classified as a

mild/moderate cluster. Clusters S3-S6 displayed increased motor and overall symptom severity, although each cluster represented a unique subset of NMS. This work provided valuable insight into the importance of NMS when trying to recognize subgroups of IPD.

In 2018, **Uribe et al. (2018)** investigated the presence of cortical brain atrophy in early stages of IPD. T1-weighted images acquired on 3-tesla Siemens MRI scanners of 77 newly diagnosed and still untreated IPD patients were collected. All imaging and non-imaging data were acquired before L-Dopa treatment. An agglomerative hierarchical clustering method using Euclidean distance was used, the number of optimal clusters was found to be 2. Cluster 1 included 33 patients characterized by cortical thinning in bilateral orbitofrontal, anterior cingulate and lateral and medial anterior temporal gyri, cluster 2 included 44 patients and showed cortical thinning in bilateral occipital gyrus, superior parietal gyrus and left postcentral gyrus. The patients included in cluster 2 also showed impairment in different cognitive domains. The results of this work show that IPD neuroimaging data is able to represent two subgroups of cortical thinning, one with mainly anterior atrophy and other with mainly posterior atrophy and impaired cognitive performance.

Part II :

Methodology and materials

Chapter 4

Materials and methodology for data collection

4.1 Study Population

The study population was recruited from *Hospital Senhora da Oliveira* Neurology and Cardiology departments in the district of Guimarães, Portugal. Local hospital ethics committee approved the protocol of the study, submitted by ICVS/UM and Center Algoritmi/UM. Written consent was obtained from all subjects or their guardians.

4.1.1 Parkinsonism

This study includes 14 patients with VaP (6 females and 8 males, mean age: 81.07$\pm$4.29), 15 with IPD (4 females and 11 males, mean age: 76.6$\pm$4.29) and 15 age-matched healthy subjects (6 females and 9 males, mean age: 75.4$\pm$5.73). For all patients, the exclusion criteria were: the presence of resting tremor, moderate/severe dementia (clinical dementia rating (CDR) $>$ 2), musculoskeletal disease and overt clinical progression since diagnosis (Hoehn-Yahr $>$ 3).

Gait and anthropometric data were collected for this study. The following Table 4 summarizes the

anthropometric data of the study group related to Parkinsonism.

Table 4: Anthropometric data of healthy controls and parkinsonism study groups.

	Healthy Controls	IPD Patients	VaP Patients
Age (years)	75.4 ± 5.73	76.6 ± 4.29	81.07 ± 4.29
Weight (kg)	68.79 ± 7.25	73.24 ± 12.53	65.47 ± 10.39
Height (m)	1.63 ± 0.068	1.67 ± 0.082	1.61 ± 0.088

Characteristics are displayed as mean $\pm$ standard deviation.

4.1.2 Fabry Disease

This study includes 95 Fabry Disease patients (64 females and 41 males, mean age 51.63 ± 16.55) For all patients the exclusion criteria were: age inferior to 18 years old and the presence of neurological diseases not related to Parkinsonism. For patients that performed gait analysis, the exclusion criteria also were: presence of musculoskeletal disease, depression, moderate-severe dementia (CDR > 2) and rheumatological disorders.

All 95 patients performed cardiac and neurological examination, information about central nervous system (CNS) lesions was collected for all 95 patients. The number of patients that suffer from CNS lesions is 50 (mean age 60.34 $\pm$ 14.07) and the number of patients that do not suffer from CNS lesions is 45 (mean age 41.96 $\pm$ 13.51).

Of the 95 collected patients, only 20 (11 females and 9 males, mean age 58.25 ± 15.43) performed gait analysis. Thirteen of the 20 patients that performed gait analysis suffer from CNS lesions, furthermore, 3 of these 13 FD patients also suffer from IPD. The mean age of the patients that suffer from FD and IPD simultaneously is 66.0 ± 6.56.

The following Table 5 summarizes the anthropometric data of the 20 FD patients that performed gait analysis.

Table 5: Anthropometric data of Fabry disease patients that performed gait analysis.

	With CNS Lesions	Without CNS Lesions
Age (years)	65.62 ± 11.59	44.57 ± 11.95
Weight (kg)	63.72 ± 10.06	67.24 ± 10.96
Height (m)	1.59 ± 0.076	1.64 ± 0.052

Characteristics are displayed as mean $\pm$ standard deviation.

4.1.3 Healthy subjects and Fabry disease

At the later stages of this work, a study group composed of 36 FD patients (24 females and 12 males) and 34 healthy subjects (21 females and 13 males) became available for gait analysis. For all FD patients, the exclusion criteria were: the presence of resting tremor, moderate-severe dementia (CDR > 2), depression, less than eighteen years of age, extensive intracranial lesions or neurodegenerative disorders, musculoskeletal disease and rheumatological disorders. The information about the subjects' age, weight and height were also collected. These anthropometric characteristics are summarized in Table 6.

Table 6: Anthropometric data of healthy controls and Fabry disease patients study groups for gait analysis.

	Healthy Controls	FD Patients
Age (years)	53.32 ± 23.45	49.42 ± 18.00
Weight (kg)	68.39 ± 9.42	65.67 ± 9.17
Height (m)	1.68 ± 0.092	1.62 ± 0.084

Characteristics are displayed as mean $\pm$ standard deviation.

4.2 Gait acquisition system

To perform the gait assessment two Physilog® sensors (Gait Up®, Switzerland) were used. The Physilog ® sensor is equipped with a high-quality 3D accelerometer, a 3D gyroscope, a barometric pressure sensor and the capability of recording the data on an SD card. This equipment has been used to acquire gait for various studies on Neurological and Chronic disorders (Up (2019)).

The sensors positioned on both feet were used to measure different gait variables of each stride (also known as gait cycle). A gait cycle is composed of two phases, a stance phase, and a swing phase. The gait cycle begins when the reference foot contacts the floor and ends when that same foot contacts the floor again. The participants were asked to walk 60-meter continuous course (30 meters corridor with one turn) in a self-selected walking speed while the sensors were doing the data acquisition. A total of 17 gait variables were collected.

4.2.1 Collected gait features

The gait variables measured by the Physilog sensors are the following: **speed** (velocity of each gait cycle), **cycle duration** (duration of one gait cycle), **cadence** (number of gait cycles in a minute), **stride length** (distance between successive initial ground contact of the same foot), **stance** time (the time during which the foot is on the ground), **swing** time (the time during which the foot is in the air), **loading** (percent of stance between the heel strike and the foot being fully on the ground), **foot flat** (percent of stance where the foot is fully at on the ground), **pushing** (percent of stance between the foot being fully on the ground and the toe leaving the ground), **double support** (percent of the gait cycle where both feet touch the ground), **peak swing** (maximum angular velocity during swing), **strike angle** (angle between the foot and the ground when the heel hits the ground), **lift-off angle** (angle between the foot and the ground when the toes are leaving the ground), **maximum heel clearance** (maximum height above the ground reached by the heel), **maximum toe clearance 1** (maximum height above the ground reached by the toes after heel max clearance), **minimum toe clearance** (minimum height of the toes during swing phase) and **maximum toe clearance 2** (maximum height above the ground reached by the toes just before heel strike) (GaitUp Switzerland (2018)).

Two datasets were created with the measured gait variables, the All Strides dataset which includes the value of each gait feature for every stride (step) of all subject's (all the time-series), and the Mean Strides dataset that contains the arithmetic mean and the coefficient of variation (CV) of each gait variable for all subject's strides (mean and CV of the time-series). The All Strides dataset is composed of the 17 described gait variables and the Mean Strides dataset is composed of 34 gait variables (the 17 described

gait variable and the CV of each gait variable).

4.3 Cardiac exams

Cardiac examination was only performed for FD patients. Electrocardiogram (ECG) and Echocardiogram (Echo) were performed to evaluate the cardiac manifestations of FD disease.

Echocardiogram is an exam that uses Doppler ultrasonography to create pictures of the heart, these pictures are more accurate and detailed than the standard x-ray images. The collected images are then analysed and various Echocardiogram features are acquired (Association (2019a)).

Electrocardiogram is another cardiac exam that records the electrical activity of the heart using conductors placed on the skin. The device used to perform the Electrocardiogram was the Holter monitor. This exam has, on average, a duration of 24 hours and outputs a series of heart events that summarize the heart functioning. (Association (2019c), Association (2019b)).

4.3.1 Collected cardiac features

Echocardiogram features

For this work 22 Echocardiogram features were collected, these are the following: **MV E/A Ratio** (mitral valve ratio between early diastole (E Wave) and atrial contraction (A Wave)); **MV A Vel** (mitral valve A Wave blood flow velocity); **MV Dec T** (mitral valve deceleration time); **MV E Vel** (mitral valve E Wave blood flow velocity); **E' Lateral** (E Wave using tissue doppler imaging at lateral mitral annulus position); **E' Septal** (E Wave using tissue doppler imaging at septal mitral annulus position); **E/E' Lateral** (ratio between E Wave and E' Wave measured at lateral mitral annulus); **E/E' Medial** (ratio between E Wave and E' Wave measured at medial mitral annulus); **LVPWd** (left ventricular posterior wall thickness); **E/E' Septal** (ratio between E Wave and E' Wave measured at septal mitral annulus); **IVSd** (interventricular septum thickness at end-diastole); **LVIDd** (left ventricular internal dimension at end-diastole); **LADiam/SC**

(left atrial diameter measured at subcostal position); **AoDiam** (aorta diameter); **S' Lateral** (peak systolic velocity using tissue doppler imaging at lateral mitral annulus position); **LVdMassInd ASE** (left ventricular mass at end-diastole indexed to body surface area according to the American Society of Echocardiography (ASE)); **LADiam** (left atrial diameter); **S' Septal** (peak systolic velocity using tissue doppler imaging at septal mitral annulus position); **A' Septal** (A Wave using tissue doppler imaging at septal mitral annulus position); **A' Lateral** (A Wave using tissue doppler imaging at lateral mitral annulus position); **LVdMass ASE** (left ventricular mass at end-diastole according to the ASE); **LVIDd/SC** (left ventricular internal dimension at end-diastole and at subcostal position).

Electrocardiogram features

For this work 27 Electrocardiogram features were collected, these are the following: **HR Max** (maximum heart rate); **Unique APCs** (atrial premature complexes); **Total SVE Beats** (number of all supraventricular ectopic beats); **QT Mean** (mean of the QT intervals); **Unique PVC** (unique premature ventricular contractions); **Total VE Beats** (number of al ventricular ectopic beats); **QTc $>$ 450** (QT interval percentage that is grater than 450ms corrected for heart rate extremes); **QT Min** (minimum value of the QT interval); **QT Max** (maximum value of the QT interval); **QTc Mean** (mean of the QT intervals corrected for heart rate extremes); **Total Heart Beats** (total number of heart beats); **HR Mean** (mean heart rate); **Unique VEs** (unique ventricular ectopic events); **ASDNN 5** (average standard deviation of all 5-min normal R-R intervals); **QTc Max** (maximum QT interval corrected for extreme heart rates); **MaxSTCa3** (maximum ST interval for 3 coronary arteries); **MaxSTCa2** (maximum ST interval for 2 coronary arteries); **MaxSTCa1** (maximum ST interval for 1 coronary arteries); **QTc Min** (minimum value for the QT interval corrected for extreme heart rates); **MinSTCa1** (minimum ST interval for 1 coronary arteries); **RMSSD** (root mean square differences of successive R-R (heartbeat) intervals); **SDNN** (Standard deviation of the of normal R-R intervals); **SDANN 5** (Standard deviation of sequential 5-minute of normal R-R interval means); **HR Min** (minimum heart rate); **MinSTCa3** (minimum ST interval for 3 coronary arteries); **MinSTCa2** (minimum ST interval for 2 coronary arteries); **Longest R-R** (maximum difference between two R-R peaks).

4.4 Gait normalization

The importance of foot clearance and spatial-temporal gait measurements in distinguishing dysfunctions of older adults and disorders that impact gait function has been increasing (Md. Tahir and Manap (2012), Dadashi et al. (2013), Daliri (2012)). However, differences in subjects' biometric properties including age, height, weight, sex as well as walking speed may lead to disperse gait characteristics which limit the recognition of patterns in foot clearance and spatial-temporal gait measurements. The correlations between physical properties and gait variables reduce the ability to detect and correctly classify between-group differences in gait features (Wahid et al. (2015)).

In 2015, Wahid et al. proposed various methods such as dimensionless equations, detrending method and multiple regression (MR) approaches to tackle the effect of between-subject physical differences (Wahid et al. (2015)). Wahid et al. compared these three approaches and concluded that multiple regression was the superior method. Multiple regression (MR) was able to reduce the correlations between subject-specific physical properties and gait features. Additionally, Wahid et al. also showed that the MR normalization was able to improve the accuracy of machine learning methods when differentiating between parkinsonian and healthy controls gait patterns. In this work, the MR approach proposed by Wahid et al. in 2015 and most recently in 2016 was employed (Wahid et al. (2016, 2015)). Spatial-temporal gait variables including cycle duration, cadence, stance, swing, loading, foot flat, pushing, double support, speed, stride length, peak swing, and foot clearance gait variables including strike angle, lift-off angle, maximum heel clearance, maximum toe clearance 1, minimum toe clearance and maximum toe clearance 2 were normalized using a MR approach as follows:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \tag{4.1}$$

where $\hat{y}_i$ represents the prediction for the dependent gait variable for the ith observation, x_{ij} represents the jth physical property including age, height, weight, gender, speed and stride length, β_0 represents the intercept term, β_j represents the coefficient for the jth physical property and ε_i represents the residual error.

The coefficients of the MR model are estimated using the control subjects dataset. The best fitted MR models are then used to normalize each gait variable by dividing the value of the original dependent gait variable y_i by $\hat{y}_i$ according to:

$$y_i^n = \frac{y_i}{\hat{y}_i} \tag{4.2}$$

where y_i^n represents the normalized gait feature for the ith observation.

Similar to the approaches implemented by Mikos et al. and Wahid et al., age, height, weight, sex, and self-selected walking speed were used as independent variables when implementing the MR models (Mikos et al. (2018); Wahid et al. (2016, 2015)). Additionally, this work also included the subjects' stride length as an independent variable, as it was shown to significantly affect foot clearance gait variables in recent studies (Alcock et al. (2018), Ferreira et al. (2019)).

The multicollinearity between the independent features was quantified using variance inflation factors (VIF), if VIF $>$ 5 the features are regarded as correlated and are omitted from the regression model (Belsley (1991)). For each gait feature, the best MR model was selected based on the adjusted R^2 and Akaike's information criterion (AIC) values. After normalization, the Spearman's rank-order correlation coefficient (ρ) was used to assess the correlations between the independent features and the gait features. Since gender is a categorical feature the coefficient of correlation between gender and all the gait features was assessed using point biserial coefficient of correlation (Lev (1949)).

According to Burnham and Anderson (2002) the AIC is given by the following:

$$AIC = 2k - 2\ln(L) \tag{4.3}$$

where k represents the number of parameters and L denotes the maximized value of the likelihood functions. The model with the lowest AIC score is preferred, scores can be negative or positive (the absolute value of the score is not important).

The Spearman's rank order correlation coefficient was computed according to Rees (1989) as follows:

$$\rho = \frac{cov(rank_x, rank_y)}{\sigma_{rank_x}, \sigma_{rank_y}} \tag{4.4}$$

where x and y are the raw values, the raw values are converted to ordered ranked values resulting in $rank_x$ and $rank_y$, $cov(rank_x, rank_y)$ represents the covariance of the ranked values, and σ_{rank_x} represents the standard deviation of the x ranked values.

4.5 Feature scaling

The measurement unit data features can affect the data analysis, for example changing the speed feature from m/s to km/h may lead to very different analysis results. To avoid the impact of feature range and measurement units the data features should be scaled before being used as input of machine learning algorithms. With very few exceptions, machine learning algorithms do not perform well when the input features have very different ranges, because features with high magnitude will dominate features with low magnitude. Feature scaling aims to give equal importance to all features, this practice is extremely important for classification algorithms like neural networks or distance measurement algorithms like clustering (Geron (2015); Han et al. (2011)). The ultimate goal of feature scaling is that data should be homogeneous and composed of small values, that is all the features should have values of roughly the same magnitude (Geron (2015); Han et al. (2011)).

The two most common feature scaling methods are min-max scaling and standardization. Min-max scaling rescales the values so that they end up ranging from 0 to 1. Standardization makes all the features have a mean of 0 and a standard deviation of 1. Both these methods have advantages and disadvantages, some algorithms work better if the data features are in a range from 0 to 1, standardization does not provide a fixed range but is less affected by outliers (Geron (2015)). In this work the feature scaling method used is standardization. This method is described in the following equation:

$$\hat{x}_i = \frac{x_i - mean(x_i)}{std(x_i)} \tag{4.5}$$

where $\hat{x}_i$ is the scaled value of the input feature vector x_i.

Chapter 5

The implemented statistical and Machine Learning methods

5.1 Principal Components Analysis

Principal Components Analysis (PCA), also know as Karhunen-Loeve or K-L method, is one of the most widely used statistical procedures across a multitude of scientific fields. This method dates back to 1901 when it was first described by Karl Pearson (Pearson (1901)). The PCA algorithm was further developed and formalized by Harold Hotelling in 1933 (Hotelling (1933)).

This statistical procedure seeks to reduce the data dimensionality by projecting the original set of features into the directions of highest variance. The goal is to extract the important information from the input features and express it as a set of new orthogonal features that are called principal components (PCs). The extracted PCs are a set of orthogonal features sorted in decreasing order of significance that can be used to represent the input data. The first principal component is the one that captures the most variance, the second principal component must be orthogonal to the first and must capture the second largest possible variance, and so on. A percentage of the total variance is set as a threshold to select the number of principal components (Abdi and Williams (2010)).

Principal components analysis is often able to reveal relationships between the input features that

were not previously suspected, resulting in a better interpretation of the data and the relationships between features (Han et al. (2011)).

According to Han et al. (2011), the fundamental PCA procedure is the following:

- The input data must be scaled before the PCA analysis. Feature scaling ensures that all the features fall into the same range, this also ensures that attributes with large domains will not dominate attributes with smaller domains.
- The next step is the computation of k orthogonal vectors that form a basis for the scaled input. These vectors point in a direction perpendicular to one another (are orthogonal) and are referred to as principal components.
- The principal components are sorted in order of decreasing significance. The first principal component explains the most variance among the data, the second principal component explains the next highest variance, and so on.
- Since the principal components are sorted in decreasing order of significance data dimensionality can be reduced by removing the principal components that explain the least variance. The most significant principal components should be able to capture the essential information of the input data.

According to Abdi and Williams (2010), PCA can be mathematically defined as follows, assuming that the input data is comprised of I samples and J features, the data can then be represented by the $I \times J$ input matrix X. As stated the input data will be scaled before the analysis, so the columns of X will have zero mean and unit norm. After feature scaling, the matrix $X^{\mathsf{T}}X$ is a correlation matrix, the eigenvalues and eigenvectors of this matrix may be used to describe the data. However, the eigendecomposition of the correlation matrix is a slow procedure, a better option used in practice is to perform singular value decomposition.

The matrix X has the following singular value decomposition:

$$X = P\Delta Q^{\mathsf{T}} \tag{5.1}$$

where P is the matrix of left singular vectors, Q is the matrix of right singular vectors and Δ is the diagonal matrix of singular values. It is important to note that Δ is proportional to the eigenvalues of $X^{\mathsf{T}}X$, more specifically Δ^2 is equal to $\boldsymbol{\lambda}$ (eigenvalues of the correlation matrix). The computation of the sum of squared samples of a feature is performed to calculate the inertia of a feature as follows:

$$\gamma_j^2 = \sum_{n=i}^{I} x_{i,j}^2 \tag{5.2}$$

The PCs are found using singular value decomposition, these are ordered by the amount of inertia explained. The values of the PCs are called factor scores. The matrix of factor scores F is obtained as follows:

$$F = P\Delta \tag{5.3}$$

The matrix Q might be interpreted as a projection matrix, the multiplication of X and Q results in the values of the projections of the data samples onto the PCs. The matrix Q is responsible for the coefficients of the linear combinations used to compute the values of the PCs (the factor scores). This can be proven by combining equation 5.1 and equation 5.3 as follows:

$$F = P\Delta = P\Delta Q^{\mathsf{T}}Q = XQ \tag{5.4}$$

5.2 Mann-Whitney U Test

Nonparametric statistical procedures assume no knowledge about the distribution of the population and are extremely used in data analysis (Devore (2012)). The Mann–Whitney U test, also known as Mann–Whitney–Wilcoxon (MWW) and Wilcoxon rank-sum test, is a nonparametric procedure that tests the null hypothesis that two continuous distributions have the same median, essentially the null hypothesis H_0 assumes that $\widetilde{u_1} = \widetilde{u_2}$. According to Devore (2012), the steps to perform this statistical test are as follows: **(1)** select the random samples from two populations, let n_1 be the number of observations of the smaller

sample, and n_2 the number of observations of the larger sample, if the number of observations is of equal size n_1 and n_2 may be randomly assigned; **(2)** arrange the $n_1 + n_2$ observations in ascending order and set a rank of $1, 2, ..., n_1 + n_2$ for each observation, in case of a tie the rank of the observations is set to be the mean of the ranks that the observations would have if they were dissimilar (e.g. if the third and fourth observations have the same value the rank that should be assigned to both observations is 3.5); **(3)** sum the ranks corresponding to the n_1 and n_2 observations, these sums are denoted by w_1 and w_2, respectively. The total $w_1 + w_2$ only depends on the number of observations as follows:

$$w_1 + w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} \tag{5.5}$$

So, after computing w_1 it may be easier to compute w_2 using the following formula:

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1 \tag{5.6}$$

The H_0 may be rejected in favour of the alternative H_1 that $\widetilde{u_1} \neq \widetilde{u_2}$ if w_1 or w_2 are sufficient small. This decision, in practice, is usually based on the minimum value between U_1 and U_2.

$$U_1 = w_1 - \frac{n_1(n_1 + 1)}{2} \tag{5.7}$$

$$U_2 = w_2 - \frac{n_2(n_2 + 1)}{2} \tag{5.8}$$

The null hypothesis will then be rejected if the smaller value between U_1 and U_2, called the U value, is less than or equal to the desired critical value given by the Table of critical values of the Mann-Whitney U Test.

When the number of observations is large (both samples exceed 8 observations) the value of U_1 (or U_2) can be obtained by approximating the normal distribution with mean and variance given by:

$$u_{U_1} = \frac{n_1 n_2}{2} \tag{5.9}$$

$$\sigma_{U_1}^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \tag{5.10}$$

The Z-Score can then be used to infer the critical region and test H_0.

$$Z = \frac{U_1 - u_{U_1}}{\sigma_{U_1}} \tag{5.11}$$

5.3 Lasso

Lasso is a regression method that performs feature selection and regularization by shirking the coefficients towards zero. The shrinkage penalty makes a trade-off between bias and variance, the lasso method usually results in higher bias to obtain lower variance (Hastie et al. (2009); James et al. (2013)).

According to Hastie et al. (2009), lasso coefficients are estimated by minimizing the residual sum of squares and a shrinkage penalty as follows:

$$\hat{\beta}^{lasso} = \underset{\beta}{\text{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{5.12}$$

where $\lambda \geqslant 0$ is a tuning parameter, as the value of λ increases the greater the impact of the shrinkage penalty, causing the coefficients to approach zero. This penalty has the effect of forcing some coefficients to be exactly zero, performing in this way feature selection. If the penalty is set to zero the Lasso becomes a simple linear regression model. In classification problems instead of using linear regression other approach is used to model the relationship between the input data and the specific outputs. The approach used for classification problems is logistic regression.

In logistic regression a logistic function is used to model the probability that a specific input x results in a given output $y \in \{0, 1\}$. According to James et al. (2013), $p(x)$ can be given as follows:

$$p(x) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}} \tag{5.13}$$

The coefficients β_0 and $\beta = \{\beta_1, ..., \beta_n\}$ are estimated based on the training observations. The

general approach to find the coefficients that best fit the observations is to maximize the log-likelihood. For simplification assuming that $\beta = \{\beta_0, \beta_1, ..., \beta_n\}$ where n is the number of features and assuming that the vector of inputs x_i includes a constant term 1 to accommodate the intercept term β_0 the log-likelihood can be written as follows:

$$\ell(\beta) = \sum_{i=1}^{N} \left\{ y_i \beta^\top x_i - \log(1 + e^{\beta^\top x_i}) \right\} \tag{5.14}$$

where N is the number of observations.

According to Hastie and Qian (2014), lasso can be used with logistic regression by minimizing the negative of the log-likelihood and the shrinkage penalty as follows:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ -\left[\sum_{i=1}^{N} \left\{ y_i \beta^\top x_i - \log(1 + e^{\beta^\top x_i}) \right\} \right] + \lambda |\beta| \right\} \tag{5.15}$$

5.4 Step-Wise Selection

Feature selection can also be performed with the assistance of a statistical method or learning algorithm and a metric that evaluates the significance of each feature for the classification task being tackled. The usual choice is to use logistic regression as the learning algorithm and the coefficient of determination (denoted R^2 or the adjusted R^2) as the evaluation metric. However, other approaches can be used, for example, the significance of features achieved by statistical methods. (James et al. (2013)).

According to James et al. (2013) there are two possible approaches, forward step-wise selection or backward step-wise selection. In forward selection, we start with no features and at each step the feature that increases the chosen significance metric the most is added, when including features no longer increases the performance the process is stopped. In backward selection we start by feeding the learning algorithm with all of the features, then at each step the least significant feature is excluded. The process is stopped when removing a feature no longer increases the performance of the algorithm.

In this work, the adopted metric was the significance value achieved by the Mann-Whitney U-test. The

approach worked as follows, the significance value of all features was assessed using the Mann-Whitney U-test, then a backward step-wise selection was performed where at each step the performance of the selected features was assessed and the feature that presented the least significance was removed, this procedure was performed until there is only one feature left (the one that achieved the highest significance value). The subset of features that achieved the best performance is selected.

5.5 Information Gain

Information gain (IG) measures the difference in the impurity of a dataset before and after it is split by a feature X_m. The Gini index or entropy can be used to measure the amount of impurity in data. These two impurity measuring methods are often used in the training process of Decision Trees (DTs) algorithms. Decision Trees are able to use these methods to retrieve the importance of features. Information gain is then a key concept behind DT algorithms (James et al. (2013), Louppe et al. (2013)).

To understand how IG is used to develop DTs algorithms a brief overview of the DTs structure and learning process is presented next.

5.5.1 Decision Trees

Decision trees are input/output models that are represented by a tree structure, these models can be used to perform classification tasks and feature selection. Let the input set of vectors $(x_1, ..., x_p)$, where p is the number of features, be an input array denoted by X, a decision tree takes this input data and their corresponding output vector $y \in [0, 1, ..., C]$, where C is the number of classes, and learns a set of splitting rules that can later be applied to perform classification. Each node t in the decision tree represents a subspace of the input data X, the root node is represented by all the input data X. Starting at the root, nodes are labeled according to a binary split test that divides their input data into two subsets (e.g. $s_t = (x_m < c)$, where x_m represents the value of the m feature and c is the threshold value).

Dividing a node results in two children nodes t_L and t_R. A node is called a terminal node or leaf if it does not have any children nodes, these nodes are then labeled with the value of the most commonly occurred class in the respective data subset. The predicted output $\hat{y}$ is then the label of the leaf node reached by an input observation when propagated through the tree (James et al. (2013), Louppe et al. (2013)).

To build a decision tree, according to Louppe et al. (2013), one takes a sample of size N from the input data and implements a recursive procedure that identifies at each node t the split test s_t that maximizes the decrease of some impurity measure $i(t)$ (Gini index or entropy). The difference of impurity at each node can be computed by:

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \tag{5.16}$$

where $p_L = N_{t_L}/N_t$, $p_R = N_{t_R}/N_t$ (N_t is the number of input samples of node t). The construction of the decision tree terminates when all features are locally constant or all nodes are leafs (also called pure nodes).

The Gini index and the entropy of a set of samples with C classes are respectively given by:

$$I_G(t) = 1 - \sum_{c=1}^{C} N_{tc}^2 \tag{5.17}$$

$$I_E(t) = -\sum_{c=1}^{C} N_{tc}\, log(N_{tc}) \tag{5.18}$$

where N_{tc} is the fraction of samples from the t node labelled with the cth class.

Unfortunately, DTs suffer from a high variance problem that negatively impacts the accuracy and robustness of the algorithm. However, this problem can be solved by aggregating many DTs, this is called bagging and it tackles the high variance problem (James et al. (2013)).

Bagging and Random Forests

Bagging is an ensemble method developed by Breiman (1996) that introduces random perturbations into the learning procedure of decision trees. It generates multiple training sets from the original data by picking observations at random with replacement, the different training sets are used to build a vast

number of decision trees, the resulting predictions are then combined to reach a final prediction (generally in classification the class with most predictions is chosen as the final prediction) (James et al. (2013)).

In 2001, Breiman presents an extension of the bagging technique known as Random Forests (RFs) (Breiman (2001)). This method combines bagging and a randomized selection of the input features to build a large number of de-correlated trees. Basically, the key idea behind RFs is to further increase the variance reduction achieved with bagging by reducing the correlation between the built trees. Random Forest algorithms at each node selects a random subset of the input features and finds which of these features best splits the data. According to Hastie et al. (2009) the algorithm is as follows:

1. For b = 1 to B (number of trees to build):
   (a) Draw a random training set Z of size N from the input data (Bagging).
   (b) Build, using Z, a random forest tree T_b by recursively repeating the following steps for each node of the tree:
      i. Select m features at random from the p features.
      ii. Pick the best feature/split point among the m features.
      iii. Split the node into two children nodes.
2. Output the random forest of trees $\{T_b\}_1^B$

A prediction is then made based on the following, let $\hat{C}_b(x)$ be the class predicted by the bth random forest tree, the final prediction of the RF algorithm is then $\hat{C}_{RF}(x) = majority\ vote\ \{\hat{C}_b(x)\}_1^B$.

Random Forests can be used to infer feature importance. To compute feature importance the information gain that each feature provides can be added up and then averaged over all B trees of the RF. So, the importance of a feature X_m can be calculated by adding up the weighted impurity decrease $p(t)\Delta i(s_t, t)$ (where $p(t)$ is the proportion N_t/N of samples reaching node t) at all the nodes t where X_m is used, averaged over all B trees of the random forest (Breiman (2001), Louppe et al. (2013)).

5.6 Support Vector Machines

Support vector machines (SVMs) are one of the most used machine learning algorithms and have shown promising results when applied to the classification of neurological disorders using various type of data (Orrù et al. (2012), Md. Tahir and Manap (2012), Haller et al. (2012), Wahid et al. (2015)).

This algorithm was first developed by Vladimir Vapnik in 1963 but only presented in 1992 by Boser, Guyon, and Vapnik (Boser et al. (1992)). The main idea behind SVMs is the search of a hyperplane that best separates the samples from different classes.

An SVM uses a subset of the training samples to find the hyperplane with the maximum margin between the different classes, this optimal hyperplane is determined by applying the method of Lagrange multipliers. The training samples that help to determine the optimal hyperplane are called the support vectors (Vapnik (2000), Mohri et al. (2012), Han et al. (2011)).

Let's considerer the following dataset D:

$$D = \{x_k, y_k\}_{k=1}^{k=m} \tag{5.19}$$

where $x \in \mathbb{R}^m$ is the kth vector of dimension m and y_k is the corresponding binary label, $y_k \in \{-1, 1\}$.

Assuming that the samples drawn from the dataset can be linearly separable the problem can be formulated as finding the optimal hyperplane that separates the training samples with the maximum margin. According to Mohri et al. (2012), the general equation of a hyperplane in $\mathbb{R}^m$ is:

$$w \cdot x + b = 0 \tag{5.20}$$

where $w \in \mathbb{R}^N$ is a non-zero vector normal to the hyperplane and $b \in \mathbb{R}$ is a scalar often refereed to as a bias.

Any sample x_k with a positive label point y_k that lies above the hyperplane satisfies:

$$w \cdot x_k + b > 0 \quad \text{for} \quad y_k = +1 \tag{5.21}$$

And, any sample x_k with a negative label point y_k that lies below the hyperplane satisfies:

$$w \cdot x_k + b < 0 \quad \text{for} \quad y_k = -1 \tag{5.22}$$

However, this is not an optimal solution, samples should not only be to the right or to the left of the hyperplane. The desired solution is one that finds the hyperplane with the largest margin leading to a better generalization. The distance from the hyperplane to the closest samples on either side is called the margin. The goal is to maximize the margin, this goal can be achieved by finding a vector w and a bias b such that:

$$H_1 : \ w \cdot x_k + b \geq 1 \quad \text{for} \quad y_k = +1 \tag{5.23}$$

$$H_2 : \ w \cdot x_k + b \leq -1 \quad \text{for} \quad y_k = -1 \tag{5.24}$$

Any sample that falls on hyperplane H_1 belongs to the positive label $+1$ and any sample that falls on or below H_2 belongs to the negative label -1. Equation 5.23 and equation 5.24 can be combined and rewritten as:

$$y_k(w \cdot x_k + b) \geq 1 \tag{5.25}$$

The training samples that fall on the hyperplanes H_1 or H_2 are called support vectors. The width of the margin (called ρ) is given by the distance between two support vectors of different labels x_p and x_n as follows:

$$\rho = (x_p - x_n) \cdot \frac{w}{\|w\|} \tag{5.26}$$

Since $x_p \cdot w = 1 - b$ and $x_n \cdot w = 1 + b$ equation 5.26 can be simplified to:

$$\rho = \frac{2}{\|w\|} \tag{5.27}$$

According to 5.27 maximizing the margin of the hyperplane ρ is equivalent to minimizing $\|w\|$ or

$\frac{1}{2}\|w\|^2$. The solution to the maximization problem can then be expressed as the following convex optimization:

$$\min_{w,b} \frac{1}{2}\|w\|^2$$
$$\text{subject to: } y_k(w \cdot x_k + b) \geq 1, \forall i \in [1, m] \tag{5.28}$$

This is a constrained optimization problem that has a unique solution, a very important property for a machine learning algorithm. Lagrange Multipliers can be used to find the solution of this optimization problem. The Lagrangian can then be defined as follows:

$$\mathcal{L} = \frac{1}{2}\|w\|^2 - \sum_{k=1}^{m} \alpha_k [y_k(w \cdot x_k + b) - 1] \tag{5.29}$$

where α_k represents the Lagrange multipliers.

The Karush-Kuhn-Tucker conditions are obtained by taking the partials derivatives of the Lagrangian with respect to the primal variables w and b and setting them to zero as follows:

$$\nabla_w \mathcal{L} = w - \sum_{k=i}^{m} \alpha_k y_k x_k = 0 \implies w = \sum_{k=i}^{m} \alpha_k y_k x_k \tag{5.30}$$

$$\nabla_b \mathcal{L} = -\sum_{k=i}^{m} \alpha_k y_k = 0 \implies \sum_{k=i}^{m} \alpha_k y_k = 0 \tag{5.31}$$

To derive the dual form of the constrained optimization problem presented in equation 5.28, we plug equations 5.30 and 5.31 into the Lagrangian yielding the following:

$$\mathcal{L} = \sum_{k=1}^{m} \alpha_k - \frac{1}{2} \sum_{k,j=1}^{m} \alpha_k \alpha_j y_k y_j (x_k \cdot x_j) \tag{5.32}$$

Analysing equation 5.32 it can be concluded that the optimization only depends on the dot product of pairs of samples that are selected as support vectors. Equation 5.30 can then be used to determine the

decision boundary d of a test sample x_t returned by the SVM as follows:

$$d(x_t) = \text{sgn}(w \cdot x_t + b) = \text{sgn}(\sum_{k=1}^{m} \alpha_k y_k (x_k \cdot x_t) + b) \tag{5.33}$$

Not all classification problems are linearly separable, but the methods explained above can be extended to deal with non-linear classification problems. These extensions consist of the transformation of the original input data into a higher dimensional space using a non-linear mapping. In this new higher dimension space the SVM can search for a linear separating hyperplane using the described technique. The maximal margin hyperplane found in the higher space corresponds to a non-linear separating hyperplane in the original space (Han et al. (2011)).

When solving the optimization problem previously described the training samples only appear in the form of dot products, to extend the linear problem one performs the following computation $\phi(x_k) \cdot \phi(x_j)$, where $\phi(x)$ is the non-linear mapping function. This dot product is equivalent to the application of a kernel function $K(x_k, x_j)$. According to Han et al. (2011), this process can be represented by the following equation:

$$K(x_k, x_j) = \phi(x_k) \cdot \phi(x_j) \tag{5.34}$$

Applying the kernel function to the optimization formula results in the following:

$$\mathcal{L} = \sum_{k=1}^{m} \alpha_k - \frac{1}{2} \sum_{k,j=1}^{m} \alpha_k \alpha_j y_k y_j K(x_k, x_j) \tag{5.35}$$

In this dissertation the following kernel functions are used:

$$\text{Linear kernel } K(x_k, x_j) = x_k^\top \cdot x_j \tag{5.36}$$

$$\text{Polynomial kernel of degree h: } K(x_k, x_j) = (\gamma\, x_k \cdot x_j + \delta)^h \tag{5.37}$$

$$\text{Gaussian radial basis function kernel: } K(x_k, x_j) = -\gamma \|x_k - x_j\|^2 \tag{5.38}$$

$$\text{Sigmoid kernel: } K(x_k, x_j) = \tanh(\gamma\, x_k \cdot x_j + \delta) \tag{5.39}$$

5.7 Multiple Layer Perceptron

Multiple Layer Perceptron (MLP) is a feedforward neural network model whose goal is to approximate a function f^*. Let's consider the following classification problem $y = f^*(x)$, the goal of a MLP is to define a mapping $y = f(x;\theta)$ and learn the values of the parameters θ that result in the best approximation of $f^*(x)$ (Haykin(2009), Goodfellow et al.(2016)).

These models are described as feedforward because the information flows from the input nodes (also known as neurons) to the output nodes, there are no feedback connections in the sense that the output value of a node is never fed back into itself or into nodes that appear earlier in the model. These models are described as networks in the sense that they are represented by a composition of functions. For example, a MLP might be represented by three functions $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ connected in a chain to form $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$, in this example $f^{(1)}$ is called the first layer of the network, $f^{(2)}$ is the second layer, and so on. The final layer of the neural network is called the output layer (Goodfellow et al.(2016)). These networks are constituted by an input layer, one or more hidden layers and one output layer, all these layers have various nodes that are connected to the nodes of the following layer. To increase the modeling power of these models every node in the network, except the input nodes, requires an activation function. The activation function is used to compute the output value of a node, these functions must be differentiable and non-linear which allows the representation of non-linear behavior between the input data and the output. All these characteristics allow MLP models to be universal approximators, with

only one hidden layer and sufficient number of hidden nodes an MLP is able to represent any function (Haykin(2009)). The output of a neuron y_j is then based on the following:

$$z_j = \sum_{i=0}^{N} w_{ji}^{(l)} y_i^{l-1} \tag{5.40}$$

$$g(z_j) = \frac{1}{1 + \exp^{-z_j}} \tag{5.41}$$

$$y_j = g(z_j) \tag{5.42}$$

where N represents the number of nodes in the previous layer $l-1$, $w_{ji}^{(l)}$ is the synaptic weight from node i in layer $l-1$ to node j in layer l, the synaptic weight $w_{j0}^{(l)}$ represents a bias b_j, y_i^{l-1} is the output of neuron i in the previous layer $l-1$, z_j represents the inner output value of the node j and $g(z_j)$ represents the output of the sigmoid activation function. There are various activation functions that can be used, some examples are sigmoid, hyperbolic tangent, radial basis function, rectifier linear unit (ReLU), among others.

A recent study concluded that ReLU achieved the best performance when compared with the other activation functions (Glorot et al.(2011)). Based on this study the used activation function in this dissertation is the ReLU activation function formulated as follows:

$$g(z_j) = max(0, z_j) \tag{5.43}$$

Multiple Layer Perceptron is a supervised learning algorithm, given an input vector x the MLP model should produce a value similar to the desired output y. To learn the mapping from the input data to the desired output the model must decide how to tune the parameters θ (synaptic weights) to achieve the desired output. The goal is to minimize a cost function that describes the performance of the network, if all training examples are correctly classified the cost function is close to zero, otherwise, if the network is classifying the training examples poorly the cost function will have a large value. Various cost functions are used depending on the problem at hand, examples include cross entropy, mean squared error, among others. This work uses cross entropy as the loss function, cross entropy has shown superior results

for classification purposes compared to other loss functions (Golik et al.(2013)). Cross entropy can be mathematically described as follows:

$$J(\theta) = -\sum_{i=0}^{M} p(i) \log(q(i)) \tag{5.44}$$

where M is the number of classes, $p(i)$ is the correct probability value for class i and $q(i)$ is the predicted probability value for class i.

The MLP is trained to obtain the minimum cost function over the training examples. The procedure is as follows: (1) the MLP is designed where the number of hidden layers, hidden nodes, and output nodes is chosen and the synaptic weights are randomly initialized to values close to zero; (2) the training samples are propagated through the network and the cost function $J(\theta)$ is calculated; (3) Finally, the weights must be tuned to reduce the cost function. To tune the weights and reduce the cost function the network performs backpropagation alongside a learning algorithm such as stochastic gradient descent. Backpropagation is used to compute the gradient of the cost function with respect to the learning parameters $\nabla_\theta J(\theta)$. Stochastic gradient descent is then used to tune the weights using the calculated gradients. According to Haykin(2009), the learning process is then composed of two steps, first backpropagation calculates the gradient of the cost function with respect to a synaptic weight $w_j i$ using the chain rule of calculus as follows:

$$\frac{\partial J(w_{ji})}{\partial w_{ji}} = \frac{\partial J(w_{ji})}{\partial y_j} \frac{\partial y_j}{\partial z_j} \frac{\partial z_j}{\partial w_{ji}} \tag{5.45}$$

the calculated gradient is then used by stochastic gradient descent (or other learning algorithm) to tune the synaptic weight w_{ji} as follows:

$$w_{ji} = w_{ji} - \alpha \frac{\partial J(w_{ji})}{\partial w_{ji}} \tag{5.46}$$

where α represents the learning rate (LR) parameter. The learning rate determines how big of an update is performed to the parameters, this value is found empirically.

Finding a good learning rate can be an extremely difficult problem, a small learning rate might be too slow to converge while a large learning rate might overshoot and fluctuate around the minimum. To

address this problem variants of gradient descent were developed, these include Momentum (Qian (1999)), Adagrad (Duchi et al. (2011)), Adadelta (Zeiler (2012)), Adaptive Moment Estimation (Adam) (Kingma and Ba (2014)), among others. In this dissertation the used optimization algorithm was Adam, this optimizer seems to outperform the others in various tasks (Kingma and Ba (2014)). According to Kingma and Ba, this learning algorithm works by computing adaptive learning rates. The optimizer stores an averaged of past gradients and squared gradients, computes a bias-corrected estimate and uses these averages to update the parameters, this algorithm can be described by the following:

$$g_t = \nabla_\theta J(\theta_{t-1}) \tag{5.47}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{5.48}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{5.49}$$

$$\hat{m}_t = \frac{m_t}{(1 - \beta_1^t)} \hat{v}_t = \frac{v_t}{(1 - \beta_2^t)} \tag{5.50}$$

$$\theta_t = \theta_{t-1} - \frac{\alpha\, \hat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)} \tag{5.51}$$

where t represents the time-step, g_t represents the gradients at time-step t, m_t represents the first moment estimate, v_t represents the second moment estimate, $\hat{m}_t$ and $\hat{v}_t$ represent the bias-corrected estimates, α represents the learning rate, $\beta_1, \beta_2 \in [0, 1]$ represent the exponential decay rates, β_1^t and β_2^t denote β_1 and β_2 to the power t and finally, ϵ prevents a zero division. The advised default values for these parameters are $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$.

5.8 Deep Belief Network

A Deep Belief Network (DBN) is a generative model that is constituted by many layers of latent variables. There are no intralayer connections and every node in each layer is connected to every node in

the neighboring layer (Goodfellow et al. (2016)). The introduction of these neural networks in 2006 by Hinton et al. led to the current deep learning renaissance. Deep Belief Networks are built in two separate stages. In the first stage, the DBN is constructed by stacking a number of Restricted Boltzmann Machines (RBMs) that are trained in a greedy layer-wise fashion, this stage is denoted as unsupervised learning. In the second stage, the weights obtained in the first stage are further fine-tuned taking into account label information, this stage is denoted as supervised fine-tuning (Hinton et al. (2006)).

An RBM is an undirected probabilistic graphical model that contains a layer of visible nodes and a single layer of hidden nodes. The connections between the visible nodes and the hidden nodes are undirected and there are no intralayer connections. Restricted Boltzmann machines are common building blocks of deep models and are used to build DBN models (Goodfellow et al. (2016), Fischer and Igel (2012)). According to Fischer and Igel (2012), defining the visible nodes of an RBM as $v = (v_1, v_2, ..., v_m)$ where m is the number of visible nodes and the hidden nodes as $h = (h_1, h_2, ..., h_n)$ where n is the number of hidden nodes, the joint probability distribution of the RBM can be described by the following:

$$p(v, h) = \frac{1}{Z} e^{-E(v,h)} \tag{5.52}$$

where Z is a normalizing constant of the form

$$Z = \sum_{v} \sum_{h} e^{-E(v,h)} \tag{5.53}$$

and $E(v, h)$ is the energy functions of the form

$$E(v, h) = -\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i \tag{5.54}$$

where w_{ij} is the weight value between the visible node v_j and hidden node h_i. The bias terms b_j and c_i are associated with the visible node v_j and hidden node h_i, respectively.

The standard way of estimating the weight parameters is by using maximum-likelihood estimation to maximize $p\left(v \mid \theta\right)$ where $\theta = (b, c, w)$. This training process tries to find the parameters that maximize the likelihood given the training samples. However, in practice, this process is not efficient, a different training process called contrastive divergence learning is then the standard training process for RBMs.

Contrastive divergence (CD or CD-k where k indicates the number of Gibbs steps) approximates the log-likelihood gradient. According to Hinton (2012) given a training sample v CD starts by computing the probability of turning ON a hidden unit i as follows:

$$p(h_i = 1|v) = \sigma(c_i + \sum_j v_j w_{ij}) \tag{5.55}$$

where σ is the logistic function $\sigma(x) = \frac{1}{(1+e^{-x})}$. This process is repeated until all hidden units are sampled. The next step is to sample new visible nodes by computing the probability of turning ON a visible unit j given the hidden vector h as follows:

$$p(v_j = 1|h) = \sigma(b_j + \sum_i h_i w_{ij}) \tag{5.56}$$

The process of computing these two probabilities is called a Gibbs step. Performing only 1 Gibbs step works well in practice and is denoted by CD-1. Let the original vectors be denoted as $v^{(0)}$ and $h^{(0)}$ and the reconstructed vectors be denoted as $v^{(1)}$ and $h^{(1)}$. The weights are then updated using the following equation:

$$\Delta w = \alpha \left(\langle v^{(0)} h^{(0)} \rangle_{data} - \langle v^{(1)} h^{(1)} \rangle_{recon} \right) \tag{5.57}$$

where α denotes the learning rate.

Deep belief networks can now be better understood after this explanation about RBMs and their training process. A DBN is developed by stacking RBMs, the training of these RBMs is done in a layer-wise greedy fashion where each RBM is trained at a time. After training the first RBM its hidden layer is used as the visible layer of the second RBM, this process is repeated for subsequent RBMs.

After this first training stage the DBN can be further developed for classification by adding an output layer and fine-tuning its weights using the same methods that are applied to train a MLP. The DBN can then be trained using back propagation and gradient descent. This is know as the second development stage of a DBN and it is called supervised fine-tuning (Hinton et al. (2006)).

5.9 Convolutional Neural Network

Convolutional Neural Networks (CNNs) was developed in 1989 by LeCun et al. to perform automatic recognition of handwritten zip code digits provided by the U.S. Postal Service (LeCun et al. (1989)). These networks are used to process data of a grid-like structure nature, these models are able to process time-series which can be thought of as a 1D grid of values taken at different time-steps and imaging data which can be thought of as a 2D grid of pixel values or a 3D grid if the color channel is included (Goodfellow et al. (2016)).

As the name indicates the key idea behind CNNs is the convolution operations that are performed to the input data. A CNN is composed of three processing layers, these are the following: convolutional layers, pooling layers, and fully connected layers (Goodfellow et al. (2016)).

5.9.1 Convolutional layer

A convolutional layer is the first layer of a CNN model, these layers are based on the discrete convolution operation. The operation performed by a convolutional layer is the following, an input D represented by a 2D array of size $n_1 \times n_2$ is convolved with a filter (also called kernel) of size $a \times a$. The filter slides across the input D according to the stride parameter. The output of a convolutional layer is a collection of feature maps of size $q_1 \times q_2$, the number of generated feature maps is given by the number of filters i. According to Stutz ((2014)), the i^{th} feature map of the lth convolutional layer denoted $C_i^{(l)}$ is computed as follows:

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{ij}^{(l-1)} * C_j^{(l-1)} \tag{5.58}$$

where $B_i^{(l)}$ is the bias, $K_{ij}^{(l-1)}$ is the filter of size $a \times a$ that connects the jth feature map in layer $(l-1)$ with the ith feature map in layer l, and $C_j^{(l-1)}$ represents the j^{th} feature map of layer $l-1$. The input of the first convolutional layer ($l = 1$) is the input data D, that is $C_1^{(0)} = D$.

After the convolution layer, an activation function is applied to the features maps, the output of the ith feature map is given as follows:

$$Y_i^{(l)} = g(C_i^{(l)}) \tag{5.59}$$

where $g(x)$ represents an activation function. The activation function used in this work was ReLU.

The main benefit of convolutional layers is the ability to learn different patterns from the data by tuning the values of the filters, for example when performing classification of different digits the CNN can tune the filters to identify vertical lines, horizontal lines, circles among other features (Goodfellow et al. (2016)).

5.9.2 Pooling layer

In a CNN model, the pooling layer always follows the convolutional layer, the objective of the pooling layer is to replace features maps' output values at a certain location with a statistical summary of the nearby values reducing the dimensionality of the feature maps. This objective is achieved using a mask of size $b \times b$ to perform a pooling operation on each of the feature maps. The most common pooling operations are the maximum pooling and averaging pooling. In this work the maximum pooling operation is used, this operation outputs the maximum value within a neighborhood (Goodfellow et al. (2016), Stutz (2014)). The maximum pooling mask size used in this work was 2×2.

The pooling operations are able to increase the model's robustness to noise and distortions. After the pooling operations, the representations of the feature maps are able to become approximately invariant to small translations of the input. The pooling layer is then essential to improve the performance of the network on subsequent unseen data. Note that no learning takes place at the pooling layers, the only objective is to summarize the output responses over a neighborhood (Goodfellow et al. (2016)).

5.9.3 Fully connected layers

The final blocks of a CNN are the fully connected layers, after all the convolutional and pooling stages the output of the last pooling layer is flattened and given as input to fully connected layers, in order words the output of the last pooling layer is fed into an MLP. The MLP is the final block of a CNN, the outputs of the pooling layers are flattened and propagated according to the MLP equations described previously. A cost function (e.g. cross-entropy) is used to measure the discrepancy between the output of the network and the ground-truth (class labels) and the weights of the CNN are updated using backpropagation and Adam optimization method previously described (Goodfellow et al. (2016)).

Convolutional Neural Networks have been extremely successful when working with grid-structure data but these models are mostly indicated to process imaging data. The most successful architectures for processing time-series are Recurrent Neural Networks (Goodfellow et al. (2016)). These are described in the following Section 5.10.

5.10 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a class of neural networks that specialize in processing sequential data. These neural networks have a chain-like structure and can be used to process time series and predict their outcome (Goodfellow et al. (2016)). The key idea behind RNNs consists of sharing parameters across different parts of a model, basically, an RNN can be thought of as an MLP that allows recurrent connections between hidden nodes. A traditional fully connected feed-forward neural network (e.g. MLP) analyses input data separately resulting in separate parameters for each input. Contrarily, a RNN shares the same parameters across the different time steps (Goodfellow et al.(2016), Graves(2012)). According to Goodfellow et al. (2016), considering a sequence that contains vectors $x^{(t)}$ with the time index t ranging from 1 to τ. The value of a hidden node can be defined as follows:

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \theta) \tag{5.60}$$

where $h^{(t)}$ represents the state of a hidden node and θ represents the parameters of the network. The hidden state is then able to save a "memory" of previous inputs and therefore influence the output of the network.

According to Graves(2012), the forward pass of an RNN is similar to that of an MLP, the major difference is that the hidden nodes receive two input signals. Given an RNN with I input nodes, H hidden nodes and K output nodes, the forward pass can be computed as follows:

$$u_h^{(t)} = \sum_{i=1}^{I} w_{ih} x_i^{(t)} + \sum_{h=1}^{H} w_{hh} z_h^{(t-1)} \tag{5.61}$$

where w_{ih} represents the weights between the input vector x and the hidden node h, w_{hh} represents the weights between hidden node h at time step $t-1$ and t and u_t represents the internal value of h at time step t. The output value $z_h^{(t)}$ of the hidden node h is computed by applying an activation function $g(x)$ as follows:

$$z_h^{(t)} = g(u_h^{(t)}) \tag{5.62}$$

The output value $a_k^{(t)}$ at time step t is then computed as:

$$a_k^{(t)} = \sum_{h=1}^{H} w_{hk} z_h^{(t)} \tag{5.63}$$

$$z_k^{(t)} = g(a_k^{(t)}) \tag{5.64}$$

The loss of an RNN is usually the sum of the time step losses:

$$J(\theta) = \sum_{t=1}^{\tau} J(\theta) \tag{5.65}$$

where the $J(\theta)$ represents the cost function, similarly to the MLP case the chosen cost function is usually the cross entropy for classification problems.

The training process of an RNN is relatively analogous to the MLP training, backpropagation can be used to calculate the gradients of the cost function with respect to the parameters. In the RNN case, this algorithm is known as backpropagation through time (BPTT), this algorithm operates on the premise that

the temporal operation of an RNN can be unfolded into an MLP, this condition allows the application of standard back-propagation (Haykin(2009)). The gradients of an RNN are then easy to compute, but these neural networks are very difficult to train in practice, especially on long term dependency problems, due to the vanishing or exploding gradient problem where the influence of the early time-steps either decays or blows up exponentially as the sequence is processed by the network (Pascanu et al.(2013), Bengio et al.(1994), Sutskever(2014), Hochreiter(1991)).

The problem is that when applying backpropagation from $z_h^{(t)}$ to $z_h^{(t-1)}$ a multiplication by $W_{hh}^\top$ is performed, this means that many factors of $W_{hh}^\top$ are involved in the gradient computation of $z_h^{(0)}$. The value of the gradients will either vanish or explode depending on the magnitude of $W_{hh}^\top$. According to Sutskever(2014), this can be described as follows:

$$\frac{\partial J}{\partial W_{hh}} = \sum_{t=1}^{\tau} dz_k^{(t)} z_h^{(t-1)\top} \tag{5.66}$$

where

$$dz_k^{(t)} = (\prod_{t=1}^{\tau} W_{hh}^\top g'(z_k^{(t)})) \tag{5.67}$$

If the values of W_{hh} are small the contributions of the past inputs of the time-series will rapidly diminish towards zero as t increases. Many attempts were made to solve the vanishing gradient problem of RNN, and the most favored to tackle the long term dependencies is the Long Short-Term Memory (LSTM) architecture.

5.10.1 Long Short-Term Memory

An LSTM is based on the idea of creating paths through time that have derivatives that neither vanish or explode. This architecture was developed in 1997 by Hochreiter and Schmidhuber, since then it has been further developed and applied to various fields with a high degree of success. The key idea behind LSTMs is the cell state which has a linear self-loop controlled by different gates, more specifically the forget gate and the input gate. The output of these models is computed by the output gate.

According to Goodfellow et al.(2016) the operations of an LSTM can be described by the following equations:

$$\textbf{Forget Gate: } f_i^{(t)} = b_i^f + \sum_{j=1}^{I} U_{i,j}^f \, x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^f \, h_j^{(t-1)} \tag{5.68}$$

$$f_i^{(t)} = \sigma(f_i^{(t)}) \tag{5.69}$$

$$\textbf{Input Gate: } s_i^{(t)} = \sigma(b_i^s + \sum_{j=1}^{I} U_{i,j}^s \, x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^s \, h_j^{(t-1)}) \odot \widetilde{C}_i^{(t)} \tag{5.70}$$

$$\widetilde{C}_i^{(t)} = \tanh(b_i^c + \sum_{j=1}^{I} U_{i,j}^c \, x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^c \, h_j^{(t-1)}) \tag{5.71}$$

$$\textbf{Cell State: } C_i^{(t)} = (f_i^{(t)} \odot C_i^{(t-1)}) + s_i^{(t)} \tag{5.72}$$

$$\textbf{Output Gate: } o_i^{(t)} = \sigma(b_i^o + \sum_{j=1}^{I} U_{i,j}^o \, x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^o \, h_j^{(t-1)}) \tag{5.73}$$

$$h_i^{(t)} = o_i^{(t)} \odot tanh(C^{(t)}) \tag{5.74}$$

where b, U, W are the bias, input weights and recurrent weights of the LSTM gate respectively, σ represents the sigmoid functions, $tanh$ represents the hyperbolic tangent functions, $f_i^{(t)}$ represents the forget gate for time step t and cell i, $s_i^{(t)}$ represents the input gate for time step t and cell i, $\widetilde{C}_i^{(t)}$ represents the new cell state candidate values for time step t and cell i, $C_i^{(t)}$ represents the cell state for time step t and cell i, $o_i^{(t)}$ represents the output gate for time step t and cell i and $h_i^{(t)}$ represents the hidden value of hidden node i at time step t.

Like the classical RNNs, an LSTM can be trained using backpropagation and gradient descent. The vanishing gradient problem is tackled because the LSTM cell state provides a way for the gradient to flow uninterruptedly which allows the network to learn long term dependencies (Goodfellow et al.(2016)).

5.11 Clustering

Clustering analysis refers to a vast set of methods related to segmenting or partitioning a set of observations into subsets or clusters such that the observations within each cluster are similar to one another and dissimilar to observations in the other clusters. Clustering is a form of unsupervised classification where each cluster aims to capture distinct properties of the input observations dividing the data into different subsets. This type of learning is sometimes referred to as data-driven because the characteristics of the subsets arise from the data (Han et al. (2011), Everitt et al. (2011)).

Clustering analysis can be extremely useful when trying to detect previously unknown groups within the data and has been widely applied to many applications ranging from business intelligence to biology (Han et al. (2011)).

Different techniques are available to perform clustering analysis, the two clustering approaches most used in practice are K-means clustering and hierarchical clustering. These methods can be used as a standalone tool or it may serve as a preprocessing step for other algorithms such as classification methods. (Hastie et al. (2009), James et al. (2013)).

5.11.1 K-Means Clustering

K-means clustering is one of the popular approaches for segmenting a data set D of n observations into K distinct non-overlapping clusters, where $K \leqslant n$. In other words, this method divides the data into K clusters such that each cluster contains at least one observation. According to James et al. (2013), considering that $C_1, ..., C_k$ denotes the different clusters the K-means method satisfies the following

properties:

1. $C_1 \cup C_2 \cup ... \cup C_k = 1, ..., n$. This property means that each observation belongs to one of the k clusters.

2. $C_k \cap C_{k'} = \emptyset$ for all $k \neq k'$. This property that the clusters do not overlap, an observation can only belong to one of the K clusters.

An objective function is used to assess the quality of the created clusters, the objective is that the observations within a cluster are similar to one another and dissimilar to observations in the other clusters, in other words, the objective is high intracluster similarity and low intercluster similarity. The quality of a cluster can be assessed by the within-cluster variation (W_{cv}), for a cluster C_k the $W_{cv}(C_k)$ is a measure of the amount by which the observations within the k cluster differ from each other. According to James et al. (2013), the objective of K-means is to minimize the W_{cv}, this can be expressed as follows:

$$min\left(\sum_{k=1}^{K} W_{cv}(C_k)\right) \tag{5.75}$$

Solving 5.75 results in partitioning the observations into K clusters such that the within-cluster variation summed over all clusters is as low as possible. The within-cluster variation of a cluster C_k can be defined using the squared Euclidean distance as follows:

$$W_{cv}(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \tag{5.76}$$

where $|C_k|$ represents the number of observation in cluster k and p represents the number of features. The K-Means optimization can then be defined by combining 5.75 and 5.76 as follows:

$$min\left(\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2\right) \tag{5.77}$$

The process of finding the global minimum of the optimization problem is extremely difficult, there are almost K^n ways to subdivide the n observations into K clusters. Unless the number of K or n is small the number of iterations that must the conducted to find the global optimum is huge, so in practice an algorithm that finds a local optimum by defining centroids (a centroid is the center of a cluster) is used

to tackle the optimization problem with excellent results. According to Han et al. (2011) the algorithm is defined as follows:

Algorithm Input:

- K: the number of clusters;
- D: a dataset of n observations

Algorithm:

1. Arbitrarily choose K observations from D as the initial cluster centroids.
2. Repeat the following steps until convergence:
   (a) Re(assign) each observation to the cluster to which the observation is most similar based on the Euclidean distance to the centroid.
   (b) Update the cluster centroid, that is for each cluster compute the mean value of the observations and set it as the new centroid.

The K-Means algorithm is not guaranteed to converge to the global optimum it instead often terminates at a local optimum and the result depends on the initial cluster assignment. It is then important to run the algorithm multiples times using different initial configurations, then all solutions are analysed and the best solution, that is the one that results in the smallest within-cluster variation, is selected.

As described, the number of clusters is a parameter that must be given explicitly. The optimal number of clusters is often an ambiguous number and there are many possible ways to estimate it. The most common method to determine the number of clusters is the elbow method. This method is based on the observation that increasing the number of clusters diminishes the within-cluster variation. As the number of clusters increases the within-cluster variations decreases because more clusters allow the segmentation of the data into more cohesive subsets, however, the decrease of within-cluster variation may start to stall after a certain number of clusters. Consequently the elbow method plots the measures of the within-cluster variance with respect to the number of clusters and finds the point (number of clusters) in the curve where the decrease of the within-cluster variance starts to stall, this point is then chosen as the optimal number of clusters (Han et al. (2011), Hastie et al. (2009)).

5.11.2 Hierarchical Clustering

Hierarchical clustering is an alternative approach to K-means clustering that does not require the definition of the number of clusters in advance. The data observations are grouped into a hierarchy that results in a tree-based representation called a dendrogram. This clustering method works by creating a hierarchy, at each level of the hierarchy the created clusters are based on clusters created at the level below. At the lowest level, all observations represent a cluster and at the highest level, there is only one cluster containing all the observations (James et al. (2013)).

There are two approaches to hierarchical clustering, these are agglomerative (bottom-up) and divisive (top-down). The agglomerative approach starts at the lowest level and at each level a pair of clusters are merged into a single cluster, the pair of chosen clusters consists of the two most similar subsets. On the contrary, divisive methods start at the highest level and at each level clusters are split into two new clusters. The split is chosen to produce two clusters with the largest intercluster dissimilarity (Hastie et al. (2009), Han et al. (2011)).

The levels of the hierarchy represent a particular grouping of the observations into clusters, the hierarchy represents an ordered sequence of the clustering. The user then decides which level represents the best representation of the data.

The interpretation of the dendrogram is essential for choosing the optimal number of clusters, the dendrogram starts at the leaves and combines clusters up to the trunk. The vertical axis represents the similarity of the clusters, at the lowest level each observation is a cluster, as we move across the vertical axis branches are fused with leaves or other branches, the lower in the vertical axis a fusion occurs the more similar the groups of observations are to each other. Fusion at the end of the vertical axis can represent the clustering of dissimilar observations. In summary, observations that fuse at the bottom of the vertical axis are similar to each other, otherwise, observations that fuse at the top of the vertical axis are quite different from each other. Conclusions can then be drawn based on the point on the vertical axis where branches are fused.

To identify the optimal number of clusters based on the dendrogram a horizontal cut across the vertical axis is made, the groups of observations beneath the cut represent the clusters. The height of the cut

controls the number of clusters, in practice, the cut location is context-dependent, visual interpretation and previous data knowledge are used as assistance tools to decide the cut location (James et al. (2013)).

To create the dendrogram an agglomerative method or a divisive method can be used. To implement these methods a measure of the distance between two clusters needs to be defined. According to Han et al. (2011), the four most widely used measures to assess the distance between clusters are as follows:

$$\textbf{Minimum distance}: \ \text{dist}_{\text{min}}(C_i, C_j) = \min_{p \in C_i, p' \in C_j} \{|p - p'|\} \tag{5.78}$$

$$\textbf{Maximum distance}: \ \text{dist}_{\text{max}}(C_i, C_j) = \max_{p \in C_i, p' \in C_j} \{|p - p'|\} \tag{5.79}$$

$$\textbf{Mean distance}: \ \text{dist}_{\text{mean}}(C_i, C_j) = |m_i - m_j| \tag{5.80}$$

$$\textbf{Average distance}: \ \text{dist}_{\text{max}}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i, p' \in C_j} |p - p'| \tag{5.81}$$

where p and p' represents two objects or points, m_i represents the mean for cluster C_i and n_i represents the number of observations in C_i.

In practice the most used measures are the mean distance and average distance, these are more robust to outliers and noisy data resulting in more balanced dendrograms. The chosen hierarchical clustering approach used in this work is agglomerative clustering, according to James et al. (2013) and Han et al. (2011) this approach can be defined by the following algorithm:

Algorithm:

1. All n observations are treated as its own cluster, the algorithm starts with k clusters, where $k = n$.
2. **Repeat the following steps until** $k = 1$:
   (a) Compute all pairwise intercluster dissimilarities among the k clusters and identify the cluster pair that is most similar. Merge the cluster pair. The dissimilarity between the pair indicates the point in the vertical axis of the dendrogram at which the fusion is placed.
   (b) Update the number of clusters, $k = k - 1$

Part III :

Results, Conclusion and future work

Chapter 6

Data preparation and evaluation metrics

For the analysis of the controls vs. Parkinsonism, IPD vs. VaP and FD without CNS lesions and FD with CNS lesions differentiation tasks four gait datasets were used, these include Raw All Strides dataset, Normalized All Strides dataset, Raw Mean Strides dataset and Normalized Mean Strides dataset. The Mean Strides datasets are derived from the All Strides datasets, these are composed of the arithmetic mean and the coefficients of variation values of the walking events of each subject.

Two cardiac datasets were used for the clustering analysis of FD patients, these were extracted from the Holter and Echocardiogram exams.

For the analysis of controls vs. FD patients, two gait datasets were used, these include Raw Mean Strides and Normalized Mean Strides, these datasets were composed of the arithmetic mean values of the walking events of each subject.

In the following Sections, the data splitting and preparation steps, as well as the quantitative metrics used to evaluate the conducted analysis, are described.

6.1 Data Splitting

Using the same data to train and evaluate a classifier can lead to misleading overoptimistic results due to over-fitting the classifier parameters to the data. The classical approach to tackle this problem is to

split the dataset into a training set, a validation set, and a test set. The training set is used to develop/train the model, the validation set is used to fine-tune the classifier hyperparameters and the test set is used to measure the performance of the classifier in an unbiased manner. The test set is usually locked away until all fine-tuning is complete, then it is used to evaluate the final classifier, this ensures that the evaluation measurements of the test set are unbiased, meaningful and trustworthy (Han et al. (2011), Kuhn and Johnson (2013)).

A better approach to the classical data splinting of training and fine-tuning is to make use of k-Fold cross-validation. In k-Fold cross-validation the data is split into k subsets or folds $D_1, D_2, ..., D_k$ of roughly the same size. Training and validation are performed k times, at iteration i, subset D_i is reserved as the validation set and the remaining subsets are used to train the classifier. At the end of the k iteration, the performance metrics achieved at each iteration are averaged and used to evaluate the training and validation performance of the classifier. The recommended k folds is 10 which leads to good bias and variance properties (Han et al. (2011), Kuhn and Johnson (2013)).

The data splitting strategy used in this work was to split the data into a final test set (test set 25%) and a training (training set 75%). The training set is then used to perform feature selection and develop the machine learning methods. To perform the fine-tuning of hyperparameters K-Fold stratified cross-validation was used on the training set, the number of folds used in this work was 10. In stratified cross-validation the folds are made by preserving the percentage of samples for each class, this means that the number of samples of each class are equally divided across all folds. Stratified cross-validation has shown to be a superior scheme compared to the regular cross-validation approach (Kohavi et al. (1995)). The classifiers that achieved the best validation performance were evaluated using the test set that was locked away from the start, as stated, this was an effort to improve the quality of final results and conclusions.

6.2 Quantitative measures for performance evaluation

To evaluate the developed classifiers a set of performance measures are used, these include accuracy, sensitivity, and specificity. Let's define some terminology to better understand these metrics, suppose that

a classifier is used on a test set of labeled tuples, P is defined as the number of positive tuples and N is the number of negative tuples. For each tuple, the classifier class label prediction is compared to the true class label. A true positive (TP) refers to the positive tuples that were correctly labeled by the classifier, a true negative (TN) refers to the negative tuples that were correctly labeled by the classifier, a false positive (FP) refers to the negative tuples that were incorrectly labeled as positive and false negatives (FN) refers to the positive tuples that were incorrectly labelled as negative tuples. These terms are often summarized in a confusion matrix (Han et al. (2011)).

A confusion matrix (CM) is a tool that allows fast interpretation of a classifier's performance, a CM for a two-class classification problem is defined in Figure 1.

		Predicted Class		
		1	**0**	
Actual Class	**1**	**TP**	**FN**	**P**
	0	**FP**	**TN**	**N**
	Total	**P'**	**N'**	**P+N**

Figure 1: Confusion matrix for a binary classification problem shown with totals for positive and negative tuples.

Accuracy (i.e., the proportion of the total number of cases that were correctly classified), Specificity (i.e., the proportion of actual negative cases which were correctly classified), and Sensitivity (i.e., the proportion of actual positive cases which were correctly classified) can be determined as follows:

$$\text{Accuracy} = \frac{TP + TN}{P + N} \tag{6.1}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{6.2}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6.3}$$

6.3 Multiple Linear Regression gait normalization

Differences in the biometric properties including age, height, weight, gender, speed and stride length of controls vs. IPD patients, controls vs. VaP patients and controls vs. FD patients were accessed using the Mann-Whitney U-test. A p-value less than 0.05 was considered statistically significant. The results are described in the following Table 7.

Table 7: Mann-Whitney U-test p-values comparing controls and group of patients (IPD, VaP and FD) in terms of physical properties, speed and stride length.

Biometric Property	**Controls vs. IPD**	**Controls vs. VaP**	**Controls vs. FD**
Age (years)	$p = 0.301$	$\mathbf{p = 0.014}$	$\mathbf{p < 0.001}$
Weight (kg)	$p = 0.247$	$p = 0.198$	$p = 0.144$
Height (m)	$p = 0.065$	$p = 0.302$	$p = 0.153$
Gender	$\mathbf{p = 0.006}$	$p = 0.078$	$p = 0.345$
Speed	$\mathbf{p < 0.001}$	$\mathbf{p < 0.001}$	$p = 0.412$
Stride Length	$\mathbf{p < 0.001}$	$\mathbf{p < 0.001}$	$p = 0.157$

Variance inflation factors for the independent gait variables were calculated to determine the severity of multicollinearity among the physical properties, speed and stride length. The following Table 8 describes the VIFs for the independent variables of the control group.

Table 8: Variance inflation factor for Age, Weight, Height, Speed, Gender and Stride Length

Multicollinearity	**Age**	**Weight**	**Height**	**Speed**	**Gender**	**Stride Length**
VIF	1.40	2.31	3.29	3.33	1.28	3.58

Combinations between physical characteristics, speed and stride length were considered as an independent variable in the development of the MR models since their VIFs were less than 5 (Belsley (1991)). The stride length was only considered as independent variable for the development of the foot clearance MR models.

For each gait variable, the best MR model was selected based on the AIC and adjusted R^2 values. The following Table 9 describes the best MR models for each gait variable, the estimated coefficients of each independent variable and the corresponding AIC and adjusted R^2 values.

Table 9: Resulting multiple linear regression models for the gait variables. The adjusted R^2 and Akaike information criterion (AIC) are shown. The independent variables are age (A), height (H), speed (S), sex (G), weight (W) and stride length (SL).

Gait variable	**Multiple Linear Regression Model**							**AIC**	**Ajusted R^2**
Spatial-Temporal Variables									
Cycle Duration	$= 1.18$		$+0.190 \cdot H$	$-0.378 \cdot S$	$-0.010 \cdot G$			-2076.96	0.576
Cadence	$= 109.77$	$+0.088 \cdot A$	$-21.56 \cdot H$	$+43.454 \cdot S$				4380.92	0.611
Stance	$= 57.76$	$+0.105 \cdot A$		$-5.80 \cdot S$	$+3.23 \cdot G$			2841.5	0.471
Swing	$= 42.24$	$-0.105 \cdot A$		$+5.80 \cdot S$	$-3.23 \cdot G$			2841.5	0.471
Loading	$= 13.51$	$-0.054 \cdot A$		$3.47 \cdot S$	$-2.41 \cdot G$			2994.1	0.262
Foot Flat	$= 84.64$	$-0.078 \cdot A$		$-23.92 \cdot S$		$+0.128 \cdot W$		3431.89	0.674
Pushing	$= 8.16$	$+0.073 \cdot A$	$-0.133 \cdot H$	$20.49 \cdot S$				3601.62	0.564
Double Support	$= 22.09$	$+0.170 \cdot A$		$-14.06 \cdot S$	$+5.95 \cdot G$			3496.54	0.588
Stride Length	$= 0.147$	$-0.0014 \cdot A$	$+0.323 \cdot H$	$0.547 \cdot S$				-1906.59	0.716
Speed	$= 0.0004$		$+1.045 \cdot H$		$+0.356 \cdot G$	$-0.007 \cdot W$		-573.83	0.082
Peak Swing	$= 261.34$	$-2.29 \cdot A$		$+247.262 \cdot S$	$-27.21 \cdot G$			6214.1	0.799
Foot Clearance Variables									
Strike Angle	$= 24.59$				$+9.44 \cdot G$	$-0.1463 \cdot W$	$-4.10 \cdot SL$	3525.84	0.592
Lift-Off Angle	$= -63.418$	$+0.460 \cdot A$			$-26.78 \cdot G$		$+4.05 \cdot SL$	4046.55	0.60
MaxHC	$= 0.419$	$-0.0035 \cdot A$			$-0.758 \cdot G$		$+0.102 \cdot SL$	-3163.48	0.712
MaxTC1	$= 0.247$	$-0.002 \cdot A$			$-0.037 \cdot G$	$+0.0004 \cdot W$		-3716.12	0.568
MinTC	$= 0.065$	$-0.0002 \cdot A$		$-0.0112 \cdot S$		$+0.000007 \cdot W$		-4833.45	0.100
MaxTC2	$= -0.012$	$-0.0004 \cdot A$			$-0.0295 \cdot G$		$+0.156 \cdot SL$	-4041.64	0.683

MaxHC: Maximum Heel Clearance; MaxTC1: Maximum Toe Clearance 1; MinTC: Minimum Toe Clearance; MaxTC2: Maximum Toe Clearance 2.

From the analysis of Table 9 it can be concluded that the subjects' self-selected walking speed was very significant in the prediction of all spatial-temporal gait variables, this goes in line with the hypothesis that normal ranges for gait variables should be defined with reference to the speed of walking (Kirtley et al. (1985)). Walking speed has also been correlated with most spatial-temporal gait variables previously (Bejek et al. (2006); Mikos et al. (2018); Wahid et al. (2016, 2015)). Another study also concluded that differences in gait parameters between healthy subjects and osteoarthritis patients decrease when walking speed is accounted for in the gait analysis (Zeni Jr and Higginson (2009)). The results also show that stride length is extremely important in the prediction of foot clearance gait variables, the best MR models in terms of adjusted R^2 for the foot clearance variables made use of stride length as an independent variable, in fact, the regression models for strike angle, lift-off angle, maximum toe clearance 1 and maximum toe clearance 2 explained 59.2% to 71.2% in the observed variance using stride length as an independent variable, this goes in line with previous studies that hypothesized that stride length impacts foot clearance measures (Alcock et al. (2018), Ferreira et al. (2019)). A subjects' age remained significant for almost all gait variables, the only exceptions are cycle duration, speed and strike angle. Weight and height seem to be the independent variables that had the least predictive power. These independent variables were

significant in the two worst MR models, weight was significant in the prediction of minimum toe clearance and height was significant in the prediction of speed, these two MR models perform very poorly compared to the others (adjusted $R^2 \leq 0.100$). Finally, with the exception of speed (adjusted R^2 = 0.082), minimum toe clearance (adjusted R^2 = 0.100) and loading (R^2 = 0.262), all other models demonstrated a good ability to predict the gait variables based on the adjusted R^2 (adjusted $R^2 >$ 0.471).

The following Table 10 shows the results of the Spearman's rank order correlations coefficients test before (raw) and after MR normalization for the control subjects.

Table 10: Spearman's correlation coefficients for the controls gait data before (raw) and after MR normalization.

Correlations	**Age**		**Weight**		**Height**		**Speed**		**Stride Length**		**Gender**	
	Raw	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**
Spatial-Temporal Variables												
Cycle Duration	0.03	-0.12	0.07	0.06	0.04	-0.03	-0.57	0.18	-0.09	0.47	-0.19	0.03
Cadence	-0.01	0.13	-0.1	0.03	-0.05	0.08	0.54	-0.05	0.05	-0.46	0.15	0.02
Stance	-0.1	0.04	0.08	0.17	-0.09	0.06	-0.06	0.03	-0.13	-0.0	0.59	-0.06
Swing	0.1	-0.04	-0.08	-0.17	0.09	-0.06	0.06	-0.03	0.13	0.0	-0.59	0.05
Loading	-0.03	-0.02	0.22	0.23	0.26	0.21	-0.0	-0.09	0.25	0.2	-0.51	0.03
Foot Flat	-0.09	0.1	0.13	-0.02	-0.11	0.1	-0.69	0.01	-0.53	-0.1	-0.02	0.08
Pushing	0.03	-0.16	-0.09	-0.05	-0.1	-0.24	0.64	0.25	0.45	0.15	0.24	0.26
Double Support	-0.14	-0.12	0.06	0.07	-0.14	-0.03	-0.18	-0.16	-0.21	-0.3	0.58	-0.05
Stride Length	0.09	-0.12	0.23	-0.01	0.3	-0.06	0.79	-0.05	–	–	0.09	0.01
Speed	-0.02	-0.12	0.21	0.08	0.27	-0.13	–	–	0.79	0.49	0.14	0.17
Peak Swing	-0.3	-0.06	-0.02	-0.12	-0.18	-0.36	0.65	0.03	0.57	0.3	-0.16	-0.05
Foot Clearance Variables												
Strike Angle	0.13	-0.07	-0.11	-0.06	0.14	-0.04	0.01	-0.24	0.5	0.16	-0.41	-0.01
Lift-Off Angle	0.27	0.08	0.06	-0.25	0.09	-0.11	-0.45	0.31	-0.61	0.21	0.13	-0.05
Maximum Heel Clearance	-0.09	-0.12	0.06	0.29	0.2	0.15	-0.02	0.17	0.02	0.19	-0.7	0.04
Maximum Toe Clearance 1	-0.28	0.03	-0.03	0.31	-0.07	0.08	-0.15	-0.08	-0.05	-0.08	-0.55	0.08
Minimum Toe Clearance	-0.33	-0.12	0.1	0.1	-0.31	-0.12	-0.16	-0.04	0.09	0.17	0.09	0.08
Maximum Toe Clearance 2	0.15	-0.08	0.08	0.0	0.28	0.03	0.13	-0.26	0.58	0.19	-0.55	-0.04

From the analysis of Table 10 it can be concluded that before normalization weak to strong correlations ($0.01 < |\rho| < 0.79$) were observed between physical properties, speed, stride length and gait variables,

after normalization these correlations were weak to moderate ($0.0 < |\rho| < 0.49$). The MR normalization reduced almost all correlations to weak values ($|\rho| \leq 0.25$), some weak to moderate correlations were still observed ($0.25 < |\rho| < 0.49$) mostly between spatial-temporal variables and stride length. The results show that stride length was strongly correlated with strike angle ($\rho = 0.5$), lift-off angle ($\rho = -0.61$) and maximum toe clearance 2 ($\rho = 0.58$). After MR normalization using stride length as an independent variable (see Table 9) the correlations between stride length, strike angle, lift-off angle and maximum toe clearance 2 were all weakened ($|\rho| \leq 0.21$). The MR normalization was able to significantly reduce the correlations between the control subjects' physical properties, speed, stride length and gait variables.

The results of the Spearman's rank-order correlations coefficients test before (raw) and after MR normalization for the Parkinsonism patients are shown in Table 11.

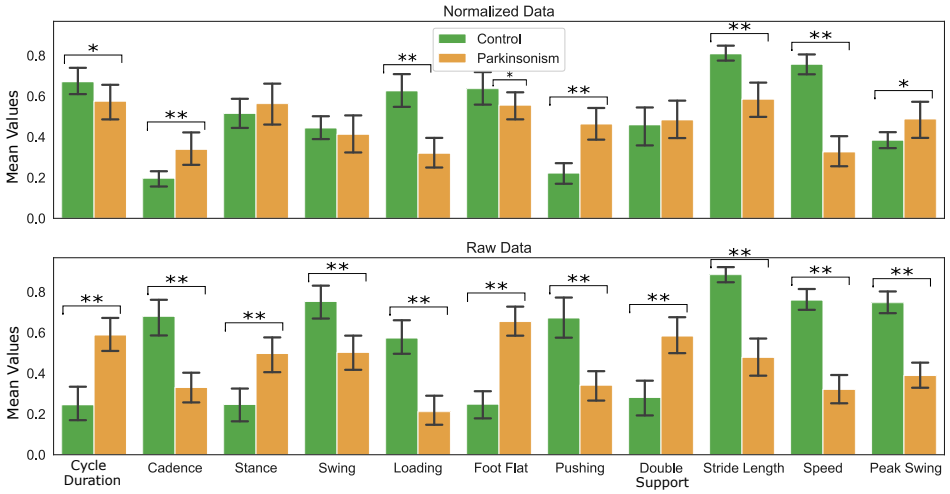
Table 11: Spearman's correlation coefficients for the Parkinsonism patients' gait data before (raw) and after MR normalization.

Correlations	**Age**		**Weight**		**Height**		**Speed**		**Stride Length**		**Gender**	
	Raw	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**
Spatial-Temporal Variables												
Cycle Duration	-0.11	-0.43	0.13	0.19	0.18	0.13	-0.18	0.5	0.06	0.72	-0.08	-0.06
Cadence	0.11	0.4	-0.13	-0.1	-0.17	-0.04	0.18	-0.68	-0.06	-0.86	0.1	0.01
Stance	0.34	0.05	0.17	0.42	0.28	0.59	-0.48	0.05	-0.48	0.06	-0.09	-0.64
Swing	-0.34	-0.06	-0.17	-0.4	-0.28	-0.6	0.48	-0.02	0.48	-0.04	0.09	0.64
Loading	-0.16	-0.01	-0.01	-0.14	-0.17	-0.51	0.6	0.27	0.48	0.15	-0.12	0.41
Foot Flat	0.37	-0.05	0.01	-0.19	-0.01	-0.16	-0.85	0.07	-0.73	0.18	0.14	0.17
Pushing	-0.34	0.21	-0.04	0.16	0.03	0.28	0.78	-0.53	0.66	-0.55	-0.12	-0.18
Double Support	0.44	0.13	0.13	0.38	0.2	0.61	-0.62	0.1	-0.66	0.07	-0.05	-0.68
Stride Length	-0.4	-0.42	0.19	0.16	0.23	0.18	0.94	0.86	–	–	-0.15	-0.03
Speed	-0.42	-0.42	0.15	0.25	0.16	0.07	–	–	0.84	0.82	-0.13	-0.11
Peak Swing	-0.51	0.2	0.14	-0.22	-0.09	-0.59	0.88	-0.39	0.78	-0.47	0.01	0.68
Foot Clearance Variables												
Strike Angle	-0.4	-0.2	0.18	0.31	0.03	-0.24	0.81	0.11	0.84	0.11	-0.15	0.17
Lift-Off Angle	0.47	0.43	-0.2	-0.26	-0.12	-0.43	-0.9	-0.66	-0.88	-0.75	0.07	0.49
Maximum Heel Clearance	-0.5	0.04	0.19	-0.23	0.25	-0.4	0.78	-0.06	0.85	-0.04	-0.22	0.84
Maximum Toe Clearance 1	-0.38	0.14	-0.06	-0.34	-0.03	-0.54	0.14	-0.2	0.2	-0.2	-0.05	0.68
Minimum Toe Clearance	-0.04	-0.03	-0.0	-0.06	-0.21	-0.26	0.4	0.49	0.41	0.49	0.13	0.09
Maximum Toe Clearance 2	-0.41	0.15	0.22	-0.07	0.09	-0.49	0.76	-0.53	0.8	-0.54	-0.19	0.4

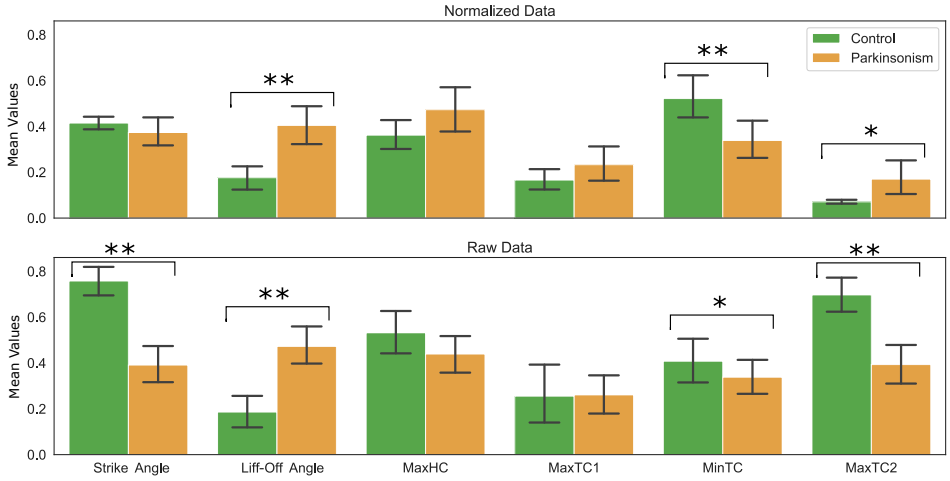
From the analysis of Table 11 it can be concluded that before gait data normalization weak to very strong correlations were observed between physical characteristics, speed, stride length, and gait features ($0.01 < |\rho| < 0.94$), while after MR normalization these correlations were weak to strong ($0.01 < |\rho| < 0.86$). Before normalization speed was moderate to strongly correlated with most spatial-temporal

and foot clearance variables ($|\rho| \geq 0.40$), the exceptions were maximum toe clearance 1 ($\rho = 0.14$), cadence ($\rho = 0.18$) and cycle duration ($\rho = -0.18$). Multiple regression normalization was able to decrease all of these correlations, however, some correlations only decreased slightly, for example, the correlation between speed and lift-off angle decreased from $\rho = -0.90$ to $\rho = -0.66$. Stride length before normalization was also strongly correlated with strike angle ($\rho = 0.84$), lift-off angle ($\rho = -0.88$), maximum heel clearance ($\rho = 0.85$) and maximum toe clearance 2 ($\rho = 0.80$), after normalization these correlations were all weakened slightly ($|\rho| \leq 0.75$). The results show that the MR normalization was able to reduce the overall correlations between Parkinsonism patients' physical properties, speed, stride length, and gait variables, however, strong correlations were also present after normalization. For some gait variables, the MR models were not able to yield good de-correlation results, this may be due to the fact that other factors besides speed, stride length, and physical properties have been shown to impact gait patterns, an example is cognition (Amboni et al. (2013), Hollman et al. (2007)). Studies have shown that Parkinsonism patients display cognitive impairments (Lehosit and Leslie J. (2015)), these impairments might cause correlations that are not captured by the derived MR models. This limitation of the MR models may be the reason why the normalization of parkinsonian gait achieved poorer results compared to the normalization of control subjects' gait. Still, the overall results show a decrease of correlation, however, it is important to note that future works should investigate if results are further improved when the MR models are derived based on more factors (such as cognition) and based on a larger control sample.

The following Figure 2 shows a comparison between the mean value of raw and MR normalized gait features of Parkinsonism patients and controls.



(a) Controls vs. Parkinsonism spatial-temporal gait variables.



(b) Controls vs. Parkinsonism foot clearance gait variables.

Figure 2: Comparison between the mean value of gait features in Parkinsonism patients and controls. Data are shown for the MR normalized gait and the raw gait data. Significant differences in gait features between FD patients and controls are indicated with one asterisk (*p $<$:05) and two asterisk (**p $<$:01). Whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot.

From the analysis of Figure 2 it can be concluded that there were only two raw gait variables that did not present significant differences when comparing controls with Parkinsonism patients, these were

maximum heel clearance (mean difference: 0.10, 95% CI: [0.07, 0.12]) and maximum toe clearance 1 (mean difference: 0.006, 95% CI: [0.04, 0.05]). After gait normalization using MR approach three spatial-temporal variables were no longer significant, these were stance (mean difference: 0.05, 95% CI: [0.02, 0.07]), swing (mean difference: 0.03, 95% CI: [0.002, 0.06]) and double support (mean difference: 0.02, 95% CI: [0.01, 0.03]). Three foot clearance measures were also non significant after normalization, these were maximum toe clearance 1 (mean difference: 0.06, 95% CI: [0.04, 0.10]), maximum heel clearance (mean difference: 0.11, 95% CI: [0.08, 0.14]) and strike angle (mean difference: 0.04, 95% CI: [0.01, 0.08]). The results indicate that MR normalization reduced the number of gait variables that presented significant differences when comparing controls and Parkinsonism patients. Contrary to previous studies were MR normalization was able to uncover significant differences in gait variables when comparing Parkinsonism patients to controls (Wahid et al. (2016, 2015)). The MR normalization in this work seems to diminish parkinsonian gait patterns, for example, parkinsonian gait is characterized by low heel clearance and low maximum toe clearance 2 since the patients display dragging of the feet while walking, however, the results show that after normalization the control subjects displayed lower maximum toe clearance 2 and lower maximum heel clearance than Parkinsonism patients. These results further corroborate the hypothesis laid out in the analysis of Table 11 that the MR models should be improved by being developed using cognition as an independent variable and that the sample size should be larger, these factors might lead to better normalization results for parkinsonian gait.

The results of the Spearman's rank-order correlations coefficients test before (raw) and after MR normalization for the FD patients are shown in Table 12.

From the analysis of Table 12 it can be concluded that before gait data normalization weak to very strong correlations were observed between physical characteristics, speed, stride length, and gait features ($0.0 < |\rho| < 0.94$), while after MR normalization these correlation were weak to strong ($0.01 < |\rho| < 0.79$). The foot clearance variables showed moderate to very strong correlations with stride length ($0.44 < |\rho| < 0.94$), after normalization these correlations were weak to moderate ($0.17 < |\rho| < 0.56$). All spatial-temporal and foot clearance variables showed weak to strong correlations with age ($0.21 < |\rho| < 0.70$) and speed ($0.12 < |\rho| < 0.87$), after normalization these correlations were reduced for age ($0.01 < |\rho| < 0.65$) and speed ($0.01 < |\rho| < 0.51$). The correlations for weight, height and gender before normalization were $0.1 < |\rho| < 0.54$, $0.1 < |\rho| < 0.74$ and $0.04 < |\rho| < 62$, receptively.

Table 12: Spearman's correlation coefficients for the FD patients' gait data before (raw) and after MR normalization.

Correlations	Age		Weight		Height		Speed		Stride Length		Gender	
	Raw	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**	**Raw**	**Norm**
Spatial-Temporal Variables												
Cycle Duration	0.22	-0.39	0.05	0.24	0.1	0.5	-0.66	0.18	-0.09	0.53	-0.28	-0.47
Cadence	-0.21	0.18	-0.04	-0.19	-0.1	-0.39	0.66	0.03	0.1	-0.42	0.27	0.43
Stance	0.62	-0.5	-0.1	0.44	-0.22	0.79	-0.83	0.14	-0.45	0.62	0.04	-0.79
Swing	-0.62	0.45	0.1	-0.46	0.22	-0.79	0.83	-0.1	0.45	-0.6	-0.04	0.78
Loading	-0.54	-0.16	0.29	0.11	0.69	0.32	0.62	0.38	0.89	0.64	-0.64	-0.26
Foot Flat	0.56	0.34	-0.06	-0.23	-0.31	-0.2	-0.76	-0.2	-0.59	-0.32	0.15	0.12
Pushing	-0.32	-0.01	-0.04	0.18	-0.13	-0.23	0.52	-0.28	0.11	-0.22	0.24	0.23
Double Support	0.52	-0.65	0.01	0.45	-0.19	0.87	-0.75	0.39	-0.5	0.73	0.05	-0.57
Stride Length	-0.7	-0.08	0.14	0.2	0.6	0.35	0.87	-0.02	—	0.38	—	-0.42
Speed	-0.68	-0.59	0.11	0.35	0.41	0.26	—	—	0.63	0.56	-0.19	-0.09
Peak Swing	-0.6	0.58	0.37	-0.0	0.2	-0.68	0.79	-0.48	0.51	-0.49	-0.16	0.5
Foot Clearance Variables												
Strike Angle	-0.7	-0.08	0.14	0.2	0.6	0.35	0.87	-0.02	0.76	0.38	-0.47	-0.42
Lift-Off Angle	0.63	0.61	-0.35	-0.21	-0.38	-0.64	-0.58	-0.51	-0.52	-0.54	0.37	0.43
Maximum Heel Clearance	-0.72	0.59	0.29	-0.33	0.61	-0.75	0.62	-0.45	0.78	-0.53	-0.62	0.47
Maximum Toe Clearance 1	-0.31	0.66	-0.07	-0.55	0.26	-0.7	0.52	-0.33	0.46	-0.56	-0.38	0.34
Minimum Toe Clearance	0.57	0.65	-0.54	-0.57	-0.29	-0.34	-0.12	-0.15	-0.44	-0.46	0.37	0.38
Maximum Toe Clearance 2	-0.61	0.15	0.48	0.23	0.8	-0.09	0.49	-0.34	0.94	0.17	-0.7	-0.01

After MR normalization all these correlations slightly increased to $0.0 < |\rho| < 0.67$, $0.2 < |\rho| < 0.77$ and $0.04 < |\rho| < 0.79$, respectively. The MR normalization was able to reduce the overall correlations, however, after MR normalization some correlations slightly increased to strong values. These results might be explained by the small sample size of the control group and by lack of diversity in terms of the physical properties of this group (the control group is composed of elderly individuals), the FD patients are much younger than the control subjects. Since the control group is composed of older individuals the developed MR models might be optimized to predict elderly gait patterns. Furthermore, studies have shown that FD patients might suffer from cognition deficits (Bolsover et al. (2014), Sigmundsdottir et al. (2014)) which, as explained before, might also be a key factor for the development of the MR models. Future MR models should be based on a larger and more diverse (in terms of physical properties) sample size, the cognition factor should also be investigated in future works.

Chapter 7

Results Parkinsonism

This chapter details the feature selection and performance results of the various classifiers developed to tackle the control subjects vs. Parkinsonism (IPD + VaP) patients and IPD vs. VaP patients classification tasks. When performing classification with the SVM, MLP or DBN classifiers based on All Strides datasets the individual test performance is computed based on the following: a subject is classified as belonging to a group if more than 50% of his strides are classified as being of that group.

The following Sections only detail the best performance results, all results can be found in the Appendix.

7.1 Feature selection

In this Section, the feature selection results for the controls vs. Parkinsonism and IPD vs. VaP classification tasks based on the Raw and Normalized All Strides and Raw and Normalized Mean Strides datasets are described. Feature scaling was performed before feature selection ensuring that the range of features does not impact the results.

The approach was the following, three feature selection methods including lasso, gini impurity and backward step-wise were used to select subsets of the original features. The feature selection methods were initially fed all features, at each iteration the least meaningful feature was removed, resulting in a

subset of the original features, this procedure was repeated until there is only one feature left. To evaluate the performance of each subset of features four SVM classifiers were used. At each iteration the SVMs were fed with the selected subset of features, the performance of these classifiers was evaluated using 10-fold stratified cross-validation. In the end, the subset of features that achieved the best classification performance was selected.

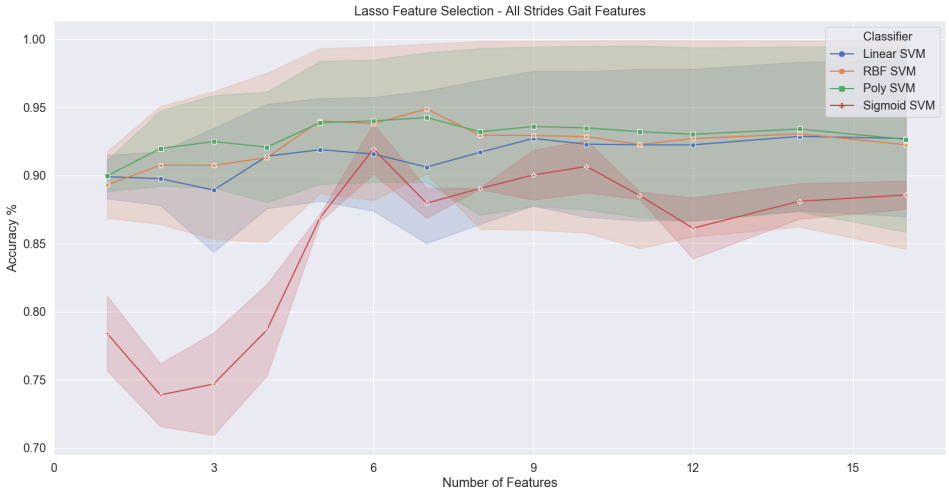
The only hyperparameter difference between the four SVM classifiers was the choice of the kernel function, these included linear, radial basis function, polynomial (third-degree) and sigmoid. The hyperparameters of the SVM classifiers were a c value of 1, a γ value of 0.1 and a δ value of 0.

The feature selection results for each gait dataset are described next. In the following plots, each point represents the mean of the final training and validation accuracies. The width represents the 95% confidence interval (CI). The objective is to choose the subset of features with the highest average accuracy and the lowest CI (width), that is the classifier is neither over-fitting or under-fitting to the training data.

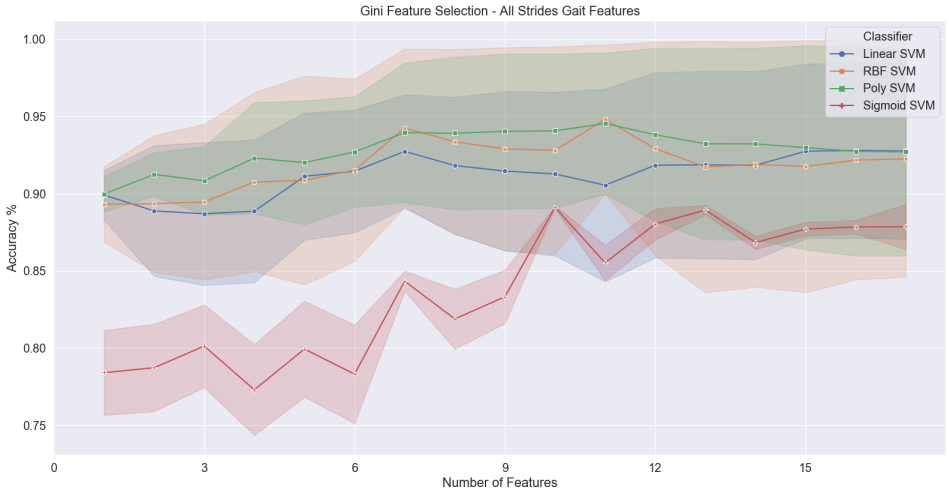
7.1.1 Feature selection - Dataset Raw All Strides

Control subjects vs. Parkinsonism patients

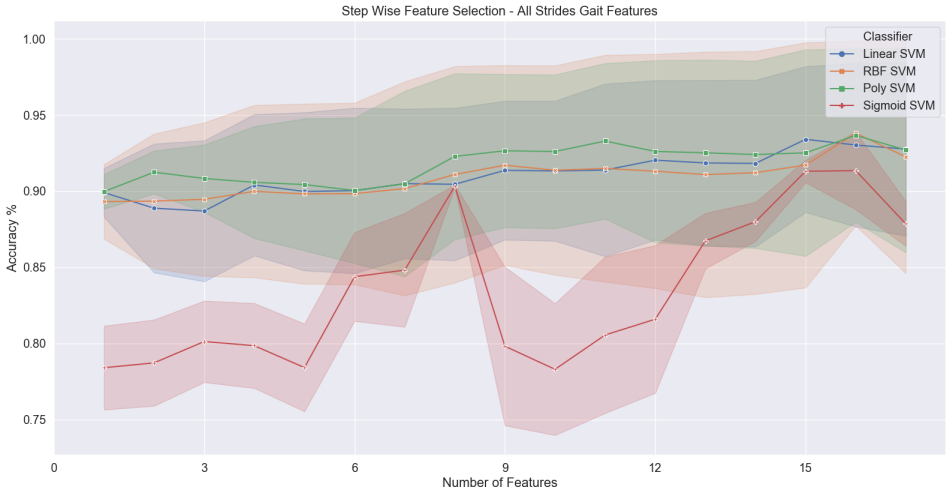
The following Figure 3 shows the feature selection results for the task of distinguishing between control subjects and Parkinsonism patients based on the Raw All Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 3: Raw All Strides dataset feature selection with lasso, gini and backward step-wise for controls vs. Parkinsonism differentiation task.

The following Table 13 summarizes the best results achieved by the feature selection methods for the controls vs. Parkinsonism differentiation task.

Table 13: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the controls vs. Parkinsonism differentiation task based on Raw All Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Sigmoid	90.116%	93.784%	91.95%	**6**
Gini	Linear	96.422%	89.058%	92.74%	**7**
Step-Wise	Sigmoid	90.244%	90.41%	90.33%	**8**

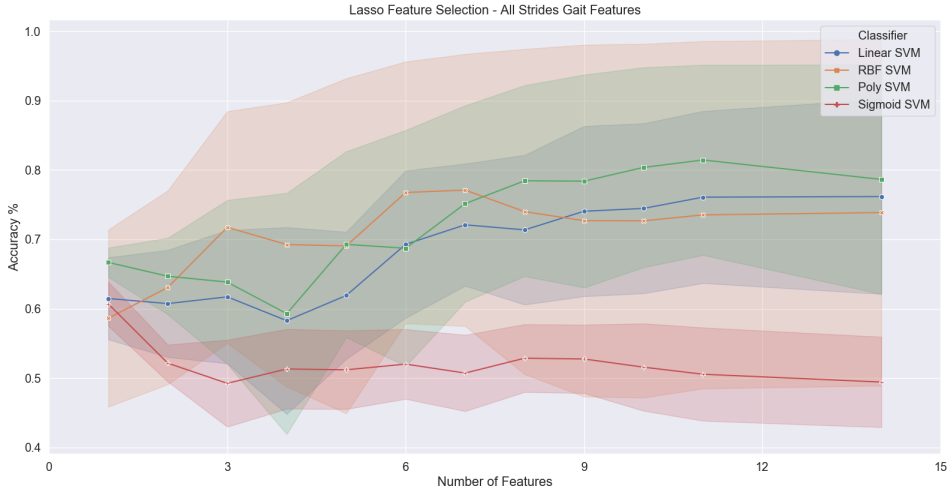
The 6 features selected by the lasso method were **cadence**, **loading**, **speed**, **peak swing**, **strike angle** and **maximum heel clearance**.

The 7 features selected by the gini method were **speed**, **peak swing**, **stride length**, **cadence**, **cycle duration**, **strike angle** and **maximum heel clearance**.

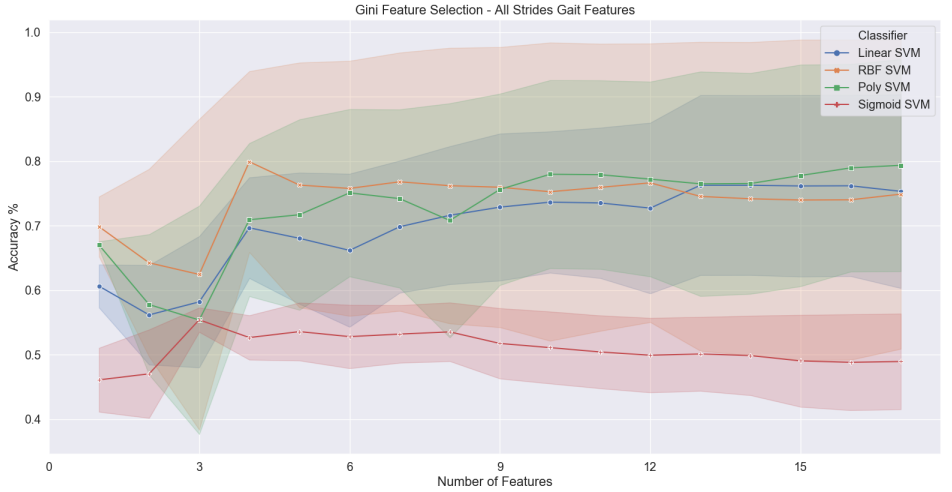
The 8 features selected by the backward step-wise method were **speed**, **foot flat**, **peak swing**, **stride length**, **lift-off angle**, **cycle duration**, **strike angle** and **loading**.

IPD patients vs. VaP patients

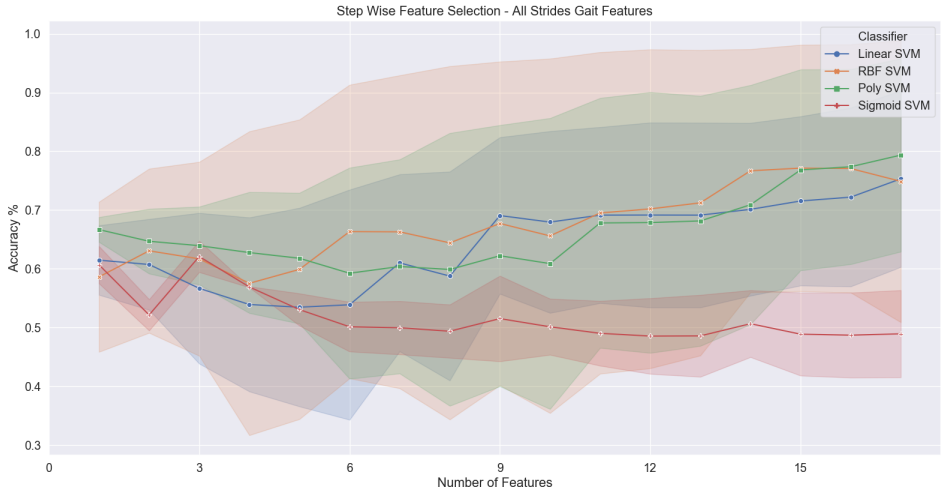
The following Figure 4 shows the feature selection results for the task of distinguishing between IPD and VaP patients based on Raw All Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 4: Raw All Strides dataset feature selection with lasso, gini and backward step-wise for IPD vs. VaP differentiation task.

The following Table 14 summarizes the best results achieved by the feature selection methods for the IPD vs. VaP differentiation task.

Table 14: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the IPD vs. VaP differentiation task based on Raw All Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Linear	80.904%	63.286%	72.10%	**7**
Gini	Linear	77.471%	61.851%	69.99%	**4**
Step-Wise	Sigmoid	64.754%	59.428%	62.10%	**3**

The 7 features selected by the lasso method were **cadence**, **loading**, **peak swing**, **lift-off angle**, **maximum heel clearance**, **minimum toe clearance** and **maximum toe clearance 2**.

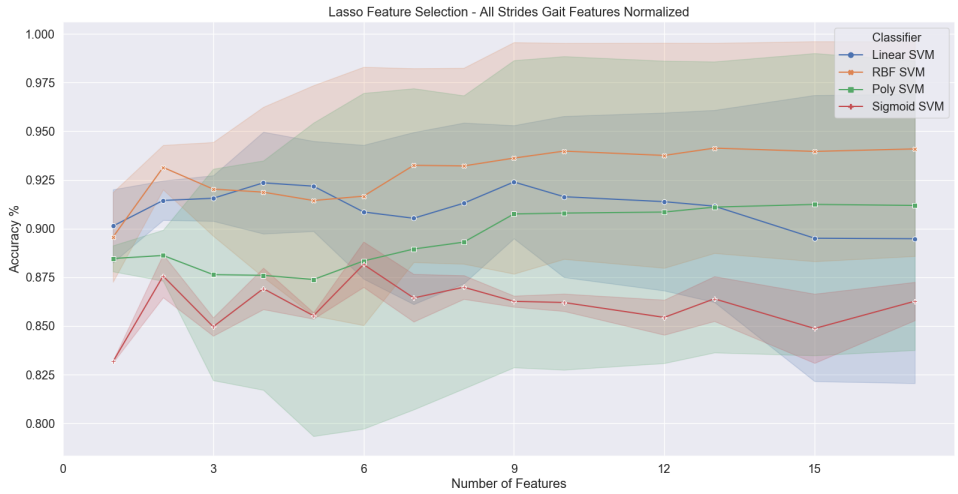
The 4 features selected by the gini method were **maximum toe clearance 2**, **peak swing**, **maximum toe clearance 1** and **loading**.

The 3 features selected by the backward step-wise method were **lift-off angle**, **peak swing** and **stride length**.

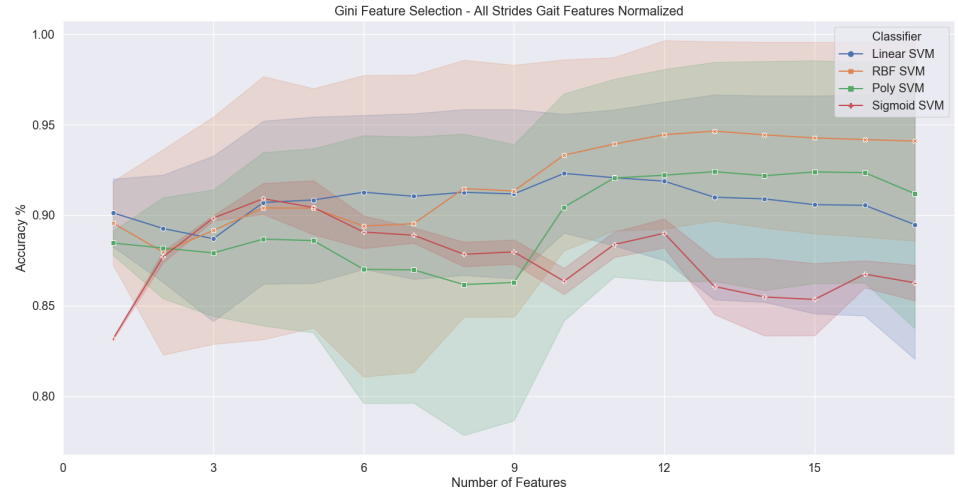
7.1.2 Feature selection - Dataset Normalized All Strides

Control subjects vs. Parkinsonism patients

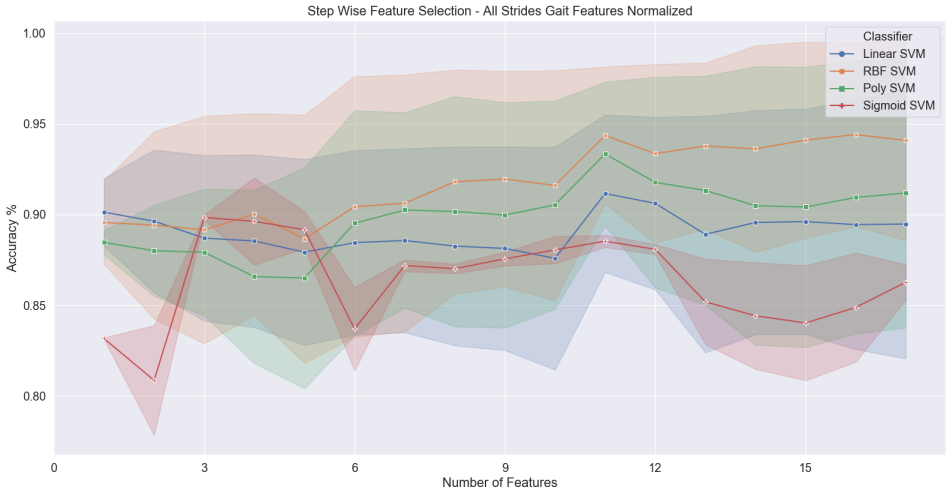
The following Figure 5 shows the feature selection results for the task of distinguishing between control subjects and Parkinsonism patients based on the Normalized All Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 5: Normalized All Strides dataset feature selection with lasso, gini and backward step-wise for controls vs. Parkinsonism differentiation task.

The following Table 15 summarizes the best results achieved by the feature selection methods for the controls vs. Parkinsonism differentiation task.

Table 15: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the controls vs. Parkinsonism differentiation task based on Normalized All Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	RBF	94.287%	92.001%	93.145%	**2**
Gini	Sigmoid	90.051%	91.772%	90.91%	**4**
Step-Wise	Sigmoid	89.695%	89.987%	89.95%	**3**

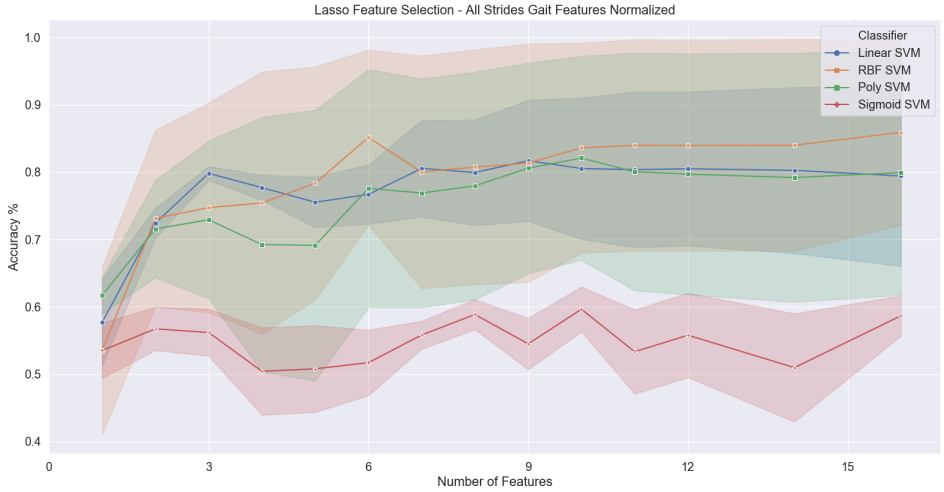
The 2 features selected by the lasso method were **speed** and **maximum heel clearance**.

The 4 features selected by the gini method were **speed**, **stride length**, **strike angle** and **maximum heel clearance**.

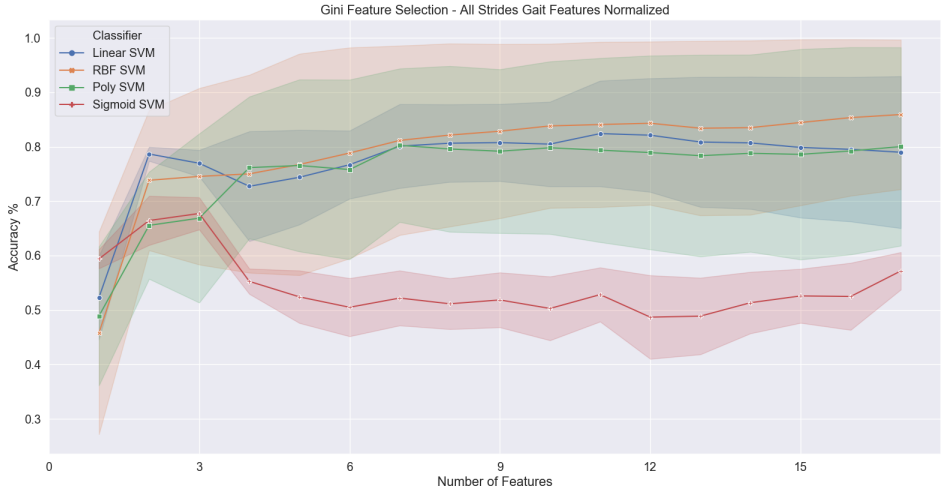
The 3 features selected by the backward step-wise method were **speed**, **stride length** and **strike angle**.

IPD patients vs. VaP patients

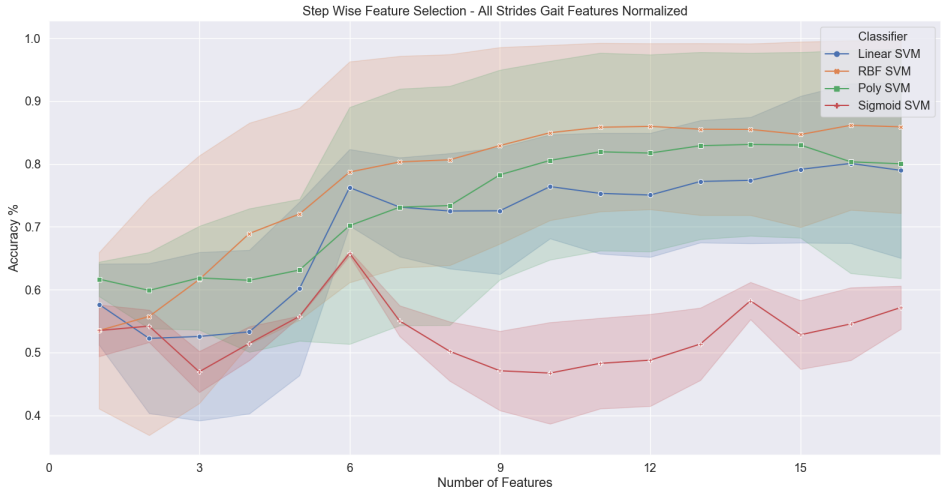
The following Figure 6 shows the feature selection results for the task of distinguishing between IPD and VaP patients based on the Normalized All Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 6: Normalized All Strides dataset feature selection with lasso, gini and backward step-wise for IPD vs. VaP differentiation task.

The following Table 16 summarizes the best results achieved by the feature selection methods for the IPD vs. VaP differentiation task.

Table 16: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the IPD vs. VaP differentiation task based on Normalized All Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Linear	80.864%	78.777%	79.82%	**3**
Gini	Linear	79.964%	77.373%	78.67%	**2**
Step-Wise	Linear	82.343%	70.157%	76.25%	**6**

The 3 features selected by the lasso method were **speed**, **strike angle** and **maximum toe clearance 1**.

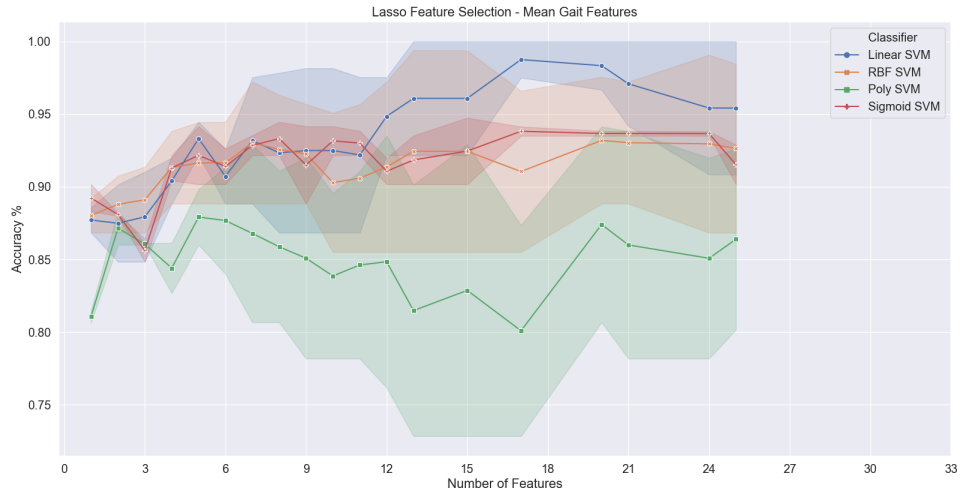
The 2 features selected by the gini method were **maximum toe clearance 1** and **strike angle**.

The 6 features selected by the backward step-wise method were **speed**, **stride length**, **cycle duration**, **strike angle**, **cadence** and **maximum toe clearance 1**.

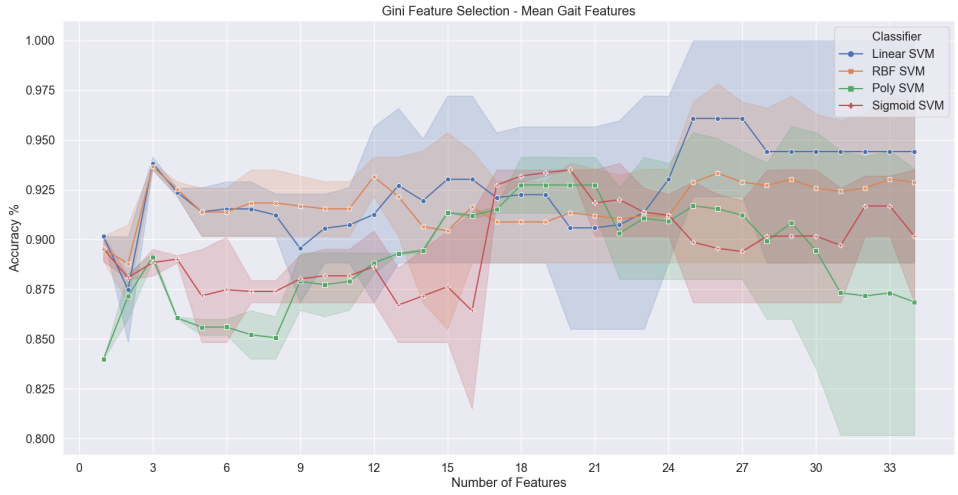
7.1.3 Feature selection - Dataset Raw Mean Strides

Control subjects vs. Parkinsonism patients

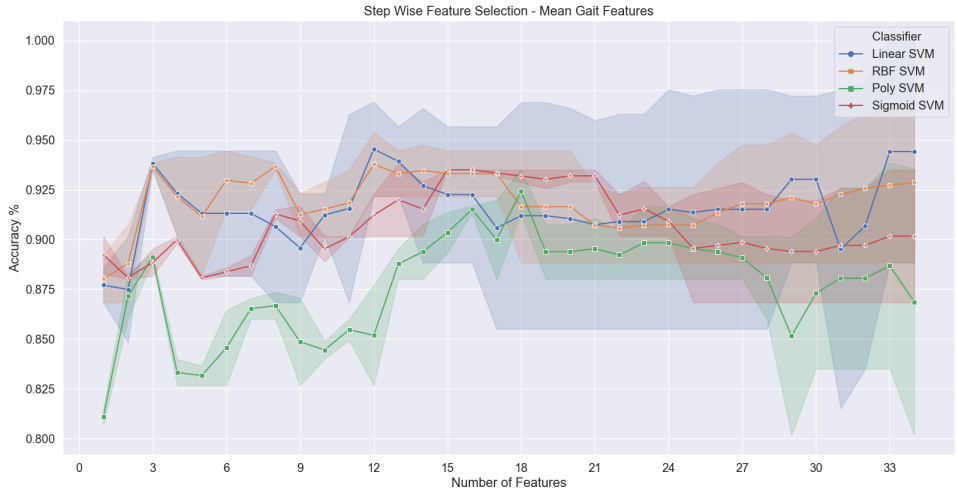
The following Figure 7 shows the feature selection results for the task of distinguishing between control subjects and Parkinsonism patients based on the Raw Mean Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 7: Raw Mean Strides dataset feature selection with lasso, gini and backward step-wise for controls vs. Parkinsonism differentiation task.

The following Table 17 summarizes the best results achieved by the feature selection methods for the controls vs. Parkinsonism differentiation task. The lasso subset of features that achieved the best performance seems to be 17, however this is a high number of features given the sample size, a previous study has shown that the optimal feature size when features present correlation approaches $\sqrt{N}$ (N denotes the sample size) as the correlation increases (Hua et al. (2004)). Given this, the number of selected features was by the lasso method was chosen to be 5.

Table 17: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the controls vs. Parkinsonism differentiation task based on Raw Mean Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Linear	94.458%	92.167%	93.32%	**5**
Gini	Linear	94.146%	93.5%	93.83%	**3**
Step-Wise	Linear	94.146%	93.5%	93.83%	**3**

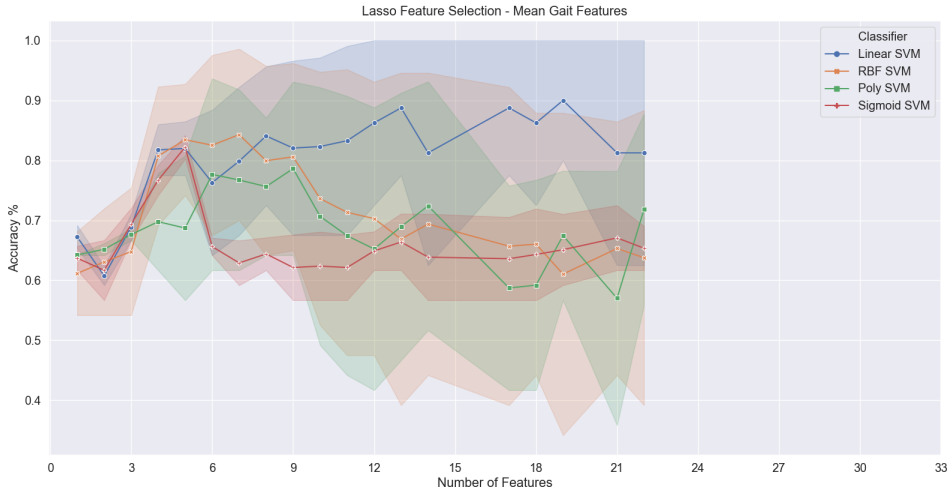
The 5 features selected by the lasso method were **cadence**, **speed**, **peak swing**, **CV of pushing** and **CV of maximum toe clearance 2**.

The 3 features selected by the gini method were **peak swing**, **speed** and **CV of strike angle**.

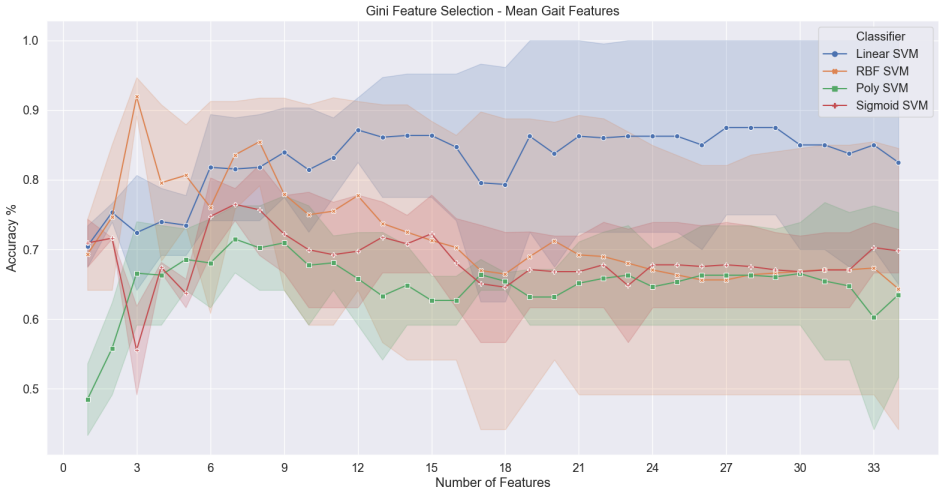
The 3 features selected by the backward step-wise method were **peak swing**, **speed** and **CV of strike angle**.

IPD patients vs. VaP patients

The following Figure 8 shows the feature selection results for the task of distinguishing between IPD and VaP patients based on the Raw Mean Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.

(c) Backward step-wise feature selection.

Figure 8: Raw Mean Strides dataset feature selection with lasso, gini and backward step-wise for IPD vs. VaP differentiation task.

The following Table 18 summarizes the best results achieved by the feature selection methods for the IPD vs. VaP differentiation task.

Table 18: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the IPD vs. VaP differentiation task based on the Raw Mean Strides dataset.

	Kernel	Training Accuracy	Validation Accuracy	Mean Accuracy	Number Features
Lasso	Sigmoid	80.25%	84.17%	82.21%	**5**
Gini	RBF	94.688%	89.167%	91.93%	**3**
Step-Wise	Sigmoid	65.323%	74.167%	69.75%	**13**

The 5 features selected by the lasso method were **lift-off angle**, **maximum toe clearance 2**, **CV of loading**, **CV of speed** and **CV of minimum toe clearance**.

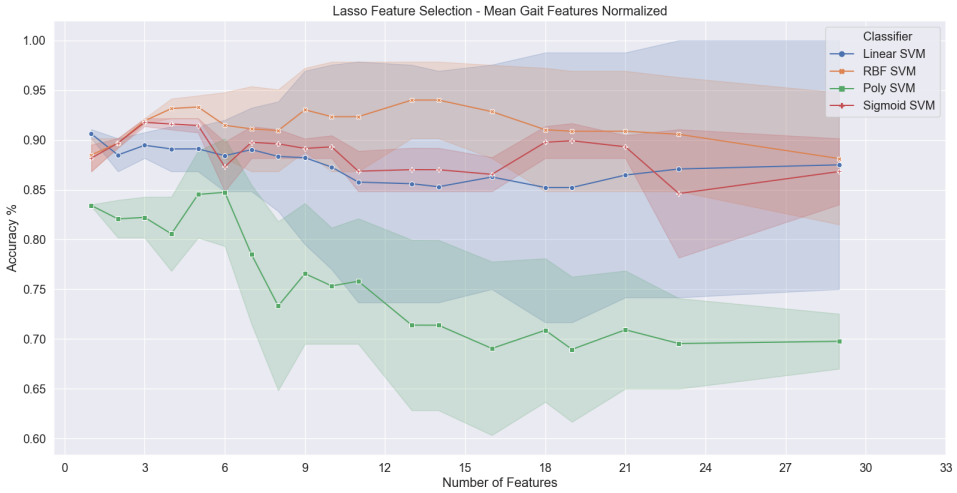
The 3 features selected by the gini method were **CV of minimum toe clearance**, **maximum heel clearance** and **maximum toe clearance 2**.

The 13 features selected by the backward step-wise method were **lift-off angle**, **maximum heel clearance**, **stride length**, **speed**, **CV of speed**, **maximum toe clearance 2**, **peak swing**, **CV of minimum toe clearance**, **strike angle**, **CV of maximum heel clearance**, **CV of stride length**, **CV of lift-off angle** and **pushing**.

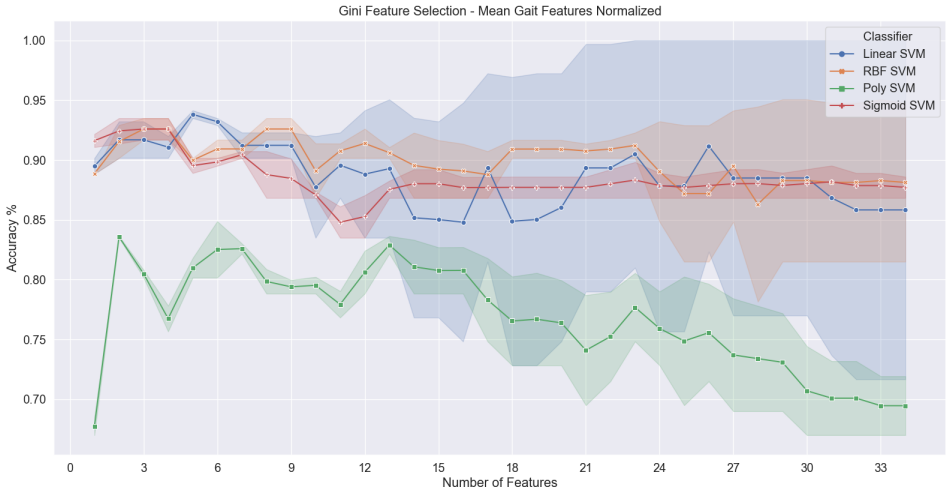
7.1.4 Feature selection - Dataset Normalized Mean Strides

Control subjects vs. Parkinsonism patients

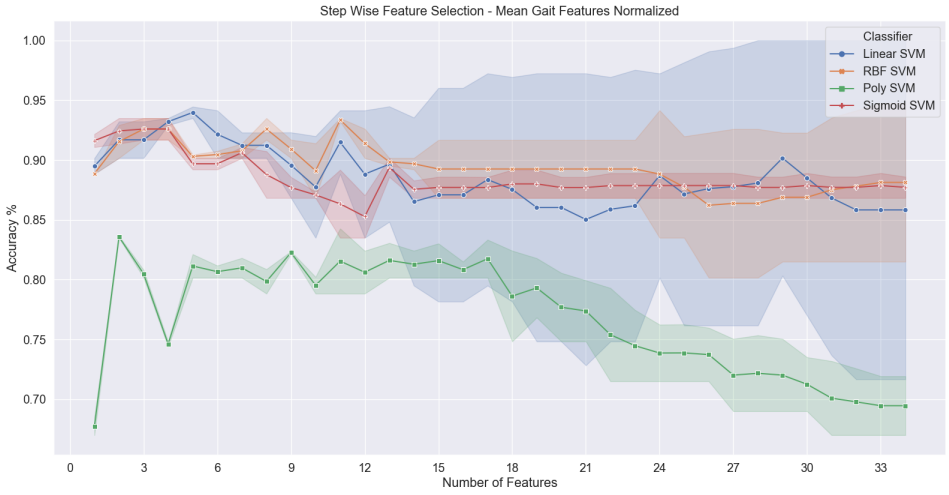
The following Figure 9 shows the feature selection results for the task of distinguishing between control subjects and Parkinsonism patients based on the Normalized Mean Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 9: Normalized Mean Strides dataset feature selection with lasso, gini and backward step-wise for controls vs. Parkinsonism differentiation task.

The following Table 19 summarizes the best results achieved by the feature selection methods for the controls vs. Parkinsonism differentiation task.

Table 19: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the controls vs. Parkinsonism differentiation task based on the Normalized Mean Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	RBF	94.458%	92.167%	93.32%	**5**
Gini	Linear	94.146%	93.5%	93.83%	**5**
Step-Wise	Linear	94.458%	93.5%	93.98%	**5**

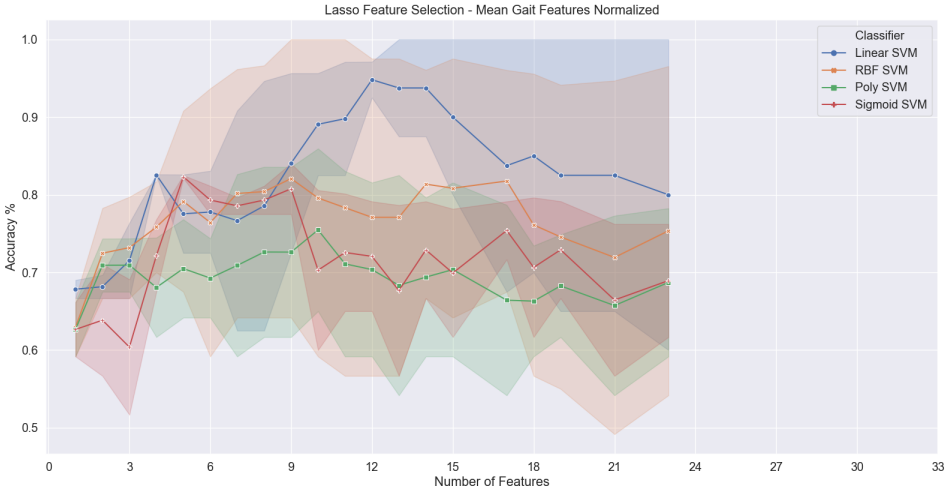
The 5 features selected by the lasso method were **loading**, **speed**, **lift-off angle**, **maximum heel clearance** and **CV of stance**.

The 5 features selected by the gini method were **CV of strike angle**, **speed**, **CV of maximum toe clearance 2**, **CV of lift-off angle** and **CV of swing**.

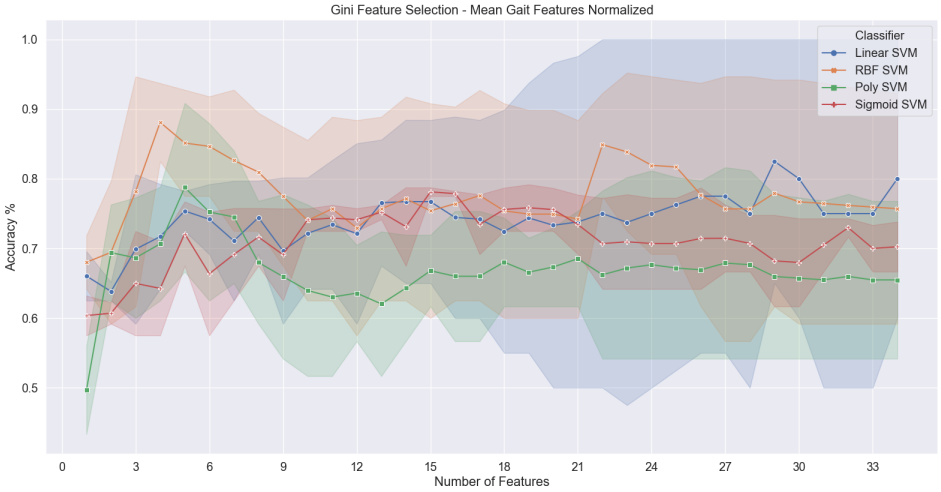
The 5 features selected by the backward step-wise method were **CV of strike angle**, **speed**, **CV of maximum toe clearance 2**, **CV of stride length** and **CV of swing**.

IPD patients vs. VaP patients

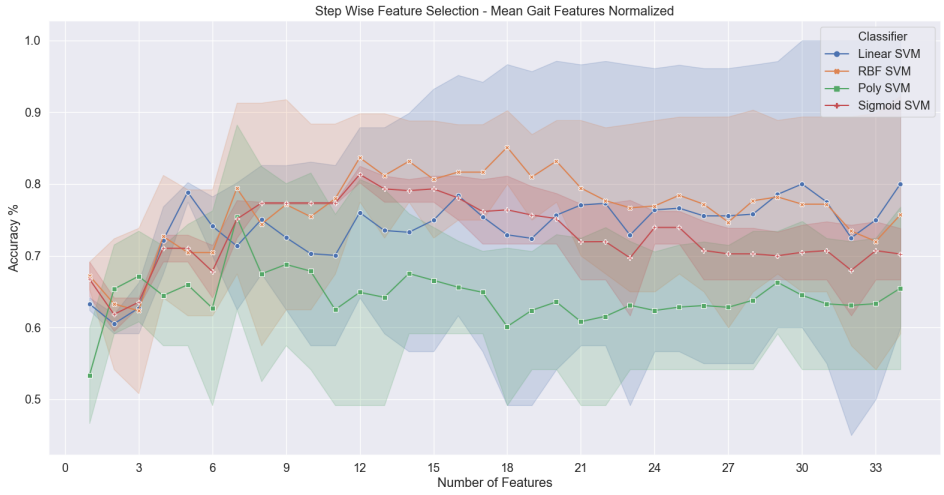
The following Figure 10 shows the feature selection results for the task of distinguishing between IPD and VaP patients based on the Normalized Mean Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 10: Normalized Mean Strides dataset feature selection with lasso, gini and backward step-wise for IPD vs. VaP differentiation task.

The following Table 20 summarizes the best results achieved by the feature selection methods for the IPD vs. VaP differentiation task. The results of the lasso feature selection method with 5 and 4 features were very similar. The final choice of 5 features is based on the fact that 5 features might allow better generalization for future predictions which yields improvements in terms of future performance.

Table 20: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the IPD vs. VaP differentiation task based on the Normalized Mean Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Sigmoid	82.135%	82.5%	82.32%	**5**
Gini	RBF	93.712%	82.5%	88.11%	**4**
Step-Wise	Linear	80.231%	77.5%	78.87%	**5**

The 5 features selected by the lasso method were **strike angle**, **maximum toe clearance 1**, **CV of loading**, **CV of speed** and **CV of minimum toe clearance**.

The 4 features selected by the gini method were **CV of minimum toe clearance**, **maximum toe clearance 1**, **cycle duration** and **peak swing**.

The 5 features selected by the backward step-wise method were **CV of speed**, **cycle duration**, **peak swing**, **maximum toe clearance 1** and **stride length**.

7.2 Performance evaluation for the SVM classifiers

Different hyperparameters were exhaustively combined which resulted in the development of 1769 SVM classifiers for each classification task and feature subset. This work comprises four different datasets, each dataset is used to perform two classification tasks and each classification task is evaluated using the three best feature subsets previously selected. In total, the number of developed SVM classifiers is 42456.

The search of hyperparameters included 29 different c values, 12 different γ values, 4 different kernel functions, and 3 different polynomial degrees. The δ value was set to 0. All combinations between these hyperparameters were evaluated.

From all the developed SVM classifiers the one with the best validation performance on each classification task and feature subset was selected and used to make predictions on the test set. Finally, for each classification task the best overall SVM classifier, based on the validation and test performances, was selected. The results are presented in the following sub-sections.

7.2.1 Performance with Raw All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 21: Performance and hyperparameters of the best SVM classifier developed with the 6 Raw All Strides features selected by the lasso method for the controls vs. Parkinsonism differentiation task.

Lasso Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.8%	83.53%	95.73%
Validation Set	96.23%	95.68%	96.95%
Test Set	97.96%	92.65%	99.75%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	1.1	0.1	––

IPD patients vs. VaP patients

Table 22: Performance and hyperparameters of the best SVM classifier developed with the 4 Raw All Strides features selected by the gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.98%	95.07%	94.88%
Validation Set	68.86%	65.07%	56.39%
Test Set	78.44%	92.37%	65.34%
Test Set Individual Performance	83.33%	100%	75.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	4.5	0.00002	––

7.2.2 Performance with Normalized All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 23: Performance and hyperparameters of the best SVM classifier developed with the 4 Normalized All Strides features selected by the gini method for the controls vs. Parkinsonism differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	90.99%	79.56%	94.65%
Validation Set	94.65%	91.42%	98.17%
Test Set	97.41%	90.65%	99.75%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	0.9	0.2	––

IPD patients vs. VaP patients

Table 24: Performance and hyperparameters of the best SVM classifier developed with the 3 Normalized All Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	80.00%	78.49%	83.31%
Validation Set	80.16%	84.42%	66.65%
Test Set	95.89%	93.85%	100%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	30	0.02	––

7.2.3 Performance with Raw Mean Strides Dataset

Control subjects vs. Parkinsonism patients

Table 25: Performance and hyperparameters of the best SVM classifier developed with the 3 Raw Mean Strides features selected by the backward step-wise method for the controls vs. Parkinsonism differentiation task.

Step-Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	93.54%	83.91%	100%
Validation Set	95.5%	91.67%	100%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.8	0.04	––

IPD patients vs. VaP patients

Table 26: Performance and hyperparameters of the best SVM classifier developed with the 13 Raw Mean Strides features selected by the backward step-wise method for the IPD vs. VaP differentiation task.

Step-Wise Method - 13 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	67.2%	72.55%	64.64%
Validation Set	74.17%	80.0%	75.0%
Test Set	83.33%	100%	75.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	3.0	0.002	––

7.2.4 Performance with Normalized Mean Strides Dataset

Control subjects vs. Parkinsonism patients

Table 27: Performance and hyperparameters of the best SVM classifier developed with the 5 Normalized Mean Strides features selected by the gini method for the controls vs. Parkinsonism differentiation task.

Gini Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.29%	82.6%	99.09%
Validation Set	93.5%	90.0%	100%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.9	0.03	––

IPD patients vs. VaP patients

Table 28: Performance and hyperparameters of the best SVM classifier developed with the 5 Normalized Mean Strides features selected by the backward step-wise method for the IPD vs. VaP differentiation task.

Step-Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	71.51%	72.14%	71.36%
Validation Set	81.67%	85.0%	85.0%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	3.0	0.2	––

7.3 Performance evaluation for the MLP classifiers

Different hyperparameters were combined which resulted in the development of 1500 MLP classifiers for each classification task and feature subset. This work comprises four different datasets, each dataset

is used to perform two classification tasks and each classification task is evaluated using the three best feature subsets previously selected. In total, the number of developed MLP classifiers is 36000.

The search of hyperparameters included 13 different values for the first hidden layer of neurons, 15 different values for the second hidden layer of neurons, 4 different values for the learning rate, 4 different values for the dropout rate (DR), 3 different values for the batch size, and 9 different values for the epochs. From all the possible 84240 combinations between these hyperparameters, 1500 were randomly chosen to develop the MLP classifiers.

From all the developed MLP classifiers the one with the best validation performance on each classification task and feature subset was selected and used to make predictions on the test set. Finally, for each classification task the best overall MLP classifier, based on the validation and test performances, was selected. The results are presented in the following sub-sections.

7.3.1 Performance with Raw All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 29: Performance and hyperparameters of the best MLP classifier developed with the 7 Raw All Strides features selected by the gini method for the controls vs. Parkinsonism differentiation task.

Gini Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	99.27%	97.98%	99.68%
Validation Set	91.48%	89.33%	94.67%
Test Set	99.44%	98.44%	99.76%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.1	24/16	32	26

IPD patients vs. VaP patients

Table 30: Performance and hyperparameters of the best MLP classifier developed with the 4 Raw All Strides features selected by the gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	94.86%	95.07%	94.54%
Validation Set	71.52%	64.14%	67.17%
Test Set	79.88%	88.76%	69.09%
Test Set Individual Performance	83.33%	100%	75.0%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.002	0.1	20/14	32	36

7.3.2 Performance with Normalized All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 31: Performance and hyperparameters of the best MLP classifier developed with the 2 Normalized All Strides features selected by the lasso method for the controls vs. Parkinsonism differentiation task.

Lasso Method - 2 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	94.54%	81.41%	99.87%
Validation Set	90.15%	83.29%	96.77%
Test Set	99.63%	100.0%	99.52%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.002	0.2	26/18	32	36

IPD patients vs. VaP patients

Table 32: Performance and hyperparameters of the best MLP classifier developed with the 3 Normalized All Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	82.53%	82.26%	83.20%
Validation Set	65.77%	70.81%	69.46%
Test Set	96.71%	99.66%	92.35%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.001	0.3	12/16	64	36

7.3.3 Performance with Raw Mean Strides Dataset

Control subjects vs. Parkinsonism patients

Table 33: Performance and hyperparameters of the best MLP classifier developed with the 3 Raw Mean Strides features selected by the gini/step-wise method for the controls vs. Parkinsonism differentiation task.

Gini/Step-Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	94.16%	85.22%	100%
Validation Set	93.5%	88.33%	100%
Test Set	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.001	0.2	20/16	64	26

IPD patients vs. VaP patients

Table 34: Performance and hyperparameters of the best MLP classifier developed with the 13 Raw Mean Strides features selected by the backward step-wise method for the IPD vs. VaP differentiation task.

Step-Wise Method - 13 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	84.61%	80.95%	89.14%
Validation Set	84.17%	81.67%	81.67%
Test Set	83.33%	75.0%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.002	0.2	18/26	4	26

7.3.4 Performance with Normalized Mean Strides Dataset

Control subjects vs. Parkinsonism patients

Table 35: Performance and hyperparameters of the best MLP classifier developed with the 5 Normalized Mean Strides features selected by the lasso method for the controls vs. Parkinsonism differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	94.14%	85.06%	100%
Validation Set	95.5%	91.67%	100%
Test Set	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.001	0.2	30/26	4	26

IPD patients vs. VaP patients

Table 36: Performance and hyperparameters of the best MLP classifier developed with the 5 Normalized Mean Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.49%	87.64%	86.41%
Validation Set	77.5%	86.67%	80.0%
Test Set	83.33%	100%	75.0%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.001	0.3	22/12	2	26

7.4 Performance evaluation for the DBN classifiers

Different hyperparameters were combined which resulted in the development of 1500 DBN classifiers for each classification task and feature subset. This work comprises four different datasets, each dataset is used to perform two classification tasks and each classification task is evaluated using the three best feature subsets previously selected. In total, the number of developed DBN classifiers is 36000.

The search of hyperparameters included 13 different values for the first hidden layer of neurons, 15 different values for the second hidden layer of neurons, 10 different values for the learning rate of the unsupervised training stage (Stage 1), 3 different values for the learning rate of the fine-tuning stage (Stage 2), 4 different values for the dropout rate, 3 different values for the batch size, 6 different values for the epochs of the Stage 1, and 6 different values for the epochs of the Stage 2. From all the possible 2 525 200 combinations between these hyperparameters, 1500 were randomly chosen to develop the DBN classifiers.

From all the developed DBN classifiers the one with the best validation performance on each classification task and feature subset was selected and used to make predictions on the test set. Finally, for each classification task the best overall DBN classifier, based on the validation and test performances,

was selected. The results are presented in the following sub-sections.

7.4.1 Performance with Raw All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 37: Performance and hyperparameters of the best DBN classifier developed with the 6 Raw All Strides features selected by the lasso method for the controls vs. Parkinsonism differentiation task.

Lasso Method - 6 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	93.37%	80.39%	98.33%
Validation Set	91.70%	84.98%	98.93%
Test Set	99.81%	100%	99.76%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.004	0.001	0.0	18/16	32	10	100

IPD patients vs. VaP patients

Table 38: Performance and hyperparameters of the best DBN classifier developed with the 7 Raw All Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 7 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	64.52%	66.97%	61.89%
Validation Set	66.66%	63.66%	66.0%
Test Set	86.24%	81.99%	100%
Test Set Individual Performance	83.33%	75.0%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0055	0.0015	0.2	24/16	32	10	100

7.4.2 Performance with Normalized All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 39: Performance and hyperparameters of the best DBN classifier developed with the 4 Normalized All Strides features selected by the gini method for the controls vs. Parkinsonism differentiation task.

Gini Method - 4 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	93.61%	89.89%	99.15%
Validation Set	93.03%	88.10%	99.28%
Test Set	98.89%	98.40%	99.04%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0015	0.002	0.2	14/12	64	8	100

IPD patients vs. VaP patients

Table 40: Performance and hyperparameters of the best DBN classifier developed with the 3 Normalized All Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	80.44%	78.70%	84.09%
Validation Set	78.37%	67.21%	89.47%
Test Set	94.66%	92.15%	100%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0015	0.0015	0.1	16/18	32	11	100

7.4.3 Performance with Raw Mean Strides Dataset

Control subjects vs. Parkinsonism patients

Table 41: Performance and hyperparameters of the best DBN classifier developed with the 5 Raw Mean Strides features selected by the lasso method for the controls vs. Parkinsonism differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	91.98%	82.91%	98.14%
Validation Set	96.0%	95.0%	100%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.007	0.0015	0.1	18/16	2	8	100

IPD patients vs. VaP patients

Table 42: Performance and hyperparameters of the best DBN classifier developed with the 5 Raw Mean Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	63.32%	71.89%	65.57%
Validation Set	74.17%	86.67%	76.67%
Test Set	83.33%	100.0%	75.0%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.006	0.0015	0.1	22/16	2	10	110

7.4.4 Performance with Normalized Mean Strides Dataset

Control subjects vs. Parkinsonism patients

Table 43: Performance and hyperparameters of the best DBN classifier developed with the 5 Normalized Mean Strides features selected by the gini method for the controls vs. Parkinsonism differentiation task.

Gini Method - 5 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	90.15%	80.18%	98.75%
Validation Set	95.5%	91.67%	100%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0075	0.002	0.1	18/24	2	8	100

IPD patients vs. VaP patients

Table 44: Performance and hyperparameters of the best DBN classifier developed with the 5 Normalized Mean Strides features selected by the backward step-wise method for the IPD vs. VaP differentiation task.

Step-Wise Method - 5 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	67.18%	67.94%	68.21%
Validation Set	76.67%	90.0%	86.67%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0075	0.002	0.2	24/18	2	10	120

7.5 Performance evaluation for the LSTM classifiers

Different hyperparameters were combined which resulted in the development of 1500 LSTM classifiers for each classification task and feature subset. This work comprises two different datasets, each dataset

is used to perform two classification tasks and each classification task is evaluated using the three best feature subsets previously selected. In total, the number of developed LSTM classifiers is 18000.

The search of hyperparameters included 15 different values for the first LSTM hidden layer of neurons, 17 different values for the second LSTM hidden layer of neurons, 15 different values for the first MLP hidden layer of neurons, 17 different values for the second MLP hidden layer of neurons, 4 different values for the learning rate, 4 different values for the dropout rate, 3 different values for the batch size, and 6 different values for the epochs. From all the possible 18 727 200 combinations between these hyperparameters, 1500 were randomly chosen to develop the LSTM classifiers.

From all the developed LSTM classifiers the one with the best validation performance on each classification task and feature subset was selected and used to make predictions on the test set. Finally, for each classification task the best overall LSTM classifier, based on the validation and test performances, was selected. The results are presented in the following sub-sections.

7.5.1 Performance with Raw All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 45: Performance and hyperparameters of the best LSTM classifier developed with the 8 Raw All Strides features selected by the backward step-wise method for the controls vs. Parkinsonism differentiation task.

Step-Wise Method - 8 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	89.25%	77.34%	99.02%
Validation Set	95.50%	91.67%	100%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.2	22/14	22/16	4	30

IPD patients vs. VaP patients

Table 46: Performance and hyperparameters of the best LSTM classifier developed with the 4 Raw All Strides features selected by the gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	67.75%	73.91%	76.65%
Validation Set	74.17%	76.67%	85.0%
Test Set	83.33%	75.0%	100%

LSTM Classifier Hyperparameters

LR	DR	LSTM Neurons	MLP Neurons	Batch Size	Epochs
0.001	0.2	22/14	22/16	4	30

7.5.2 Performance with Normalized All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 47: Performance and hyperparameters of the best LSTM classifier developed with the 3 Normalized All Strides features selected by the backward step-wise method for the controls vs. Parkinsonism differentiation task.

Step-Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	82.37%	77.87%	88.82%
Validation Set	91.0%	83.33%	95.0%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	DR	LSTM Neurons	MLP Neurons	Batch Size	Epochs
0.002	0.2	22/14	22/16	4	30

IPD patients vs. VaP patients

Table 48: Performance and hyperparameters of the best LSTM classifier developed with the 3 Normalized All Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	77.80%	73.14%	84.18%
Validation Set	82.50%	81.67%	80.0%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	22/16	18/14	2	30

7.6 Performance evaluation for the CNN classifiers

Different hyperparameters were combined which resulted in the development of 1500 CNN classifiers for each classification task and feature subset. This work comprises two different datasets, each dataset is used to perform two classification tasks and each classification task is evaluated using the three best feature subsets previously selected. In total, the number of developed CNN classifiers is 18000.

The search of hyperparameters included 15 different values for the first CNN hidden layer of neurons, 17 different values for the second CNN hidden layer of neurons, 15 different values for the first MLP hidden layer of neurons, 17 different values for the second MLP hidden layer of neurons, 3 different values for the Kernel value (length of the convolution window), 4 different values for the learning rate, 4 different values for the dropout rate, 3 different values for the batch size and 6 different values for the epochs. From all the possible 56 181 600 combinations between these hyperparameters, 1500 were randomly chosen to develop the CNN classifiers.

From all the developed CNN classifiers the one with the best validation performance on each classification task and feature subset was selected and used to make predictions on the test set. Finally, for each classification task the best overall CNN classifier, based on the validation and test performances,

was selected. The results are presented in the following sub-sections.

7.6.1 Performance with Raw All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 49: Performance and hyperparameters of the best CNN classifier developed with the 7 Raw All Strides features selected by the gini method for the controls vs. Parkinsonism differentiation task.

Gini Method - 7 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	95.06%	92.25%	96.80%
Validation Set	94.67%	86.67%	96.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	DR	CNN Neurons	Kernel	MLP Neurons	Batch Size	Epochs
0.001	0.2	22/16	3	18/14	4	30

IPD patients vs. VaP patients

Table 50: Performance and hyperparameters of the best CNN classifier developed with the 4 Raw All Strides features selected by the gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	79.38%	87.36%	77.63%
Validation Set	71.67%	70.0%	76.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	DR	CNN Neurons	Kernel	MLP Neurons	Batch Size	Epochs
0.001	0.4	22/14	3	22/16	2	30

7.6.2 Performance with Normalized All Strides Dataset

Control subjects vs. Parkinsonism patients

Table 51: Performance and hyperparameters of the best CNN classifier developed with the 3 Normalized All Strides features selected by the backward step-wise method for the controls vs. Parkinsonism differentiation task.

Step-Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	90.10%	82.53%	95.01%
Validation Set	90.17%	88.33%	96.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	DR	CNN Neurons	Kernel	MLP Neurons	Batch Size	Epochs
0.001	0.4	22/14	3	22/16	2	30

IPD patients vs. VaP patients

Table 52: Performance and hyperparameters of the best CNN classifier developed with the 3 Normalized All Strides features selected by the lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	79.68%	78.28%	81.18%
Validation Set	82.50%	81.67%	80.0%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	DR	CNN Neurons	Kernel	MLP Neurons	Batch Size	Epochs
0.001	0.2	24/16	5	18/14	2	20

7.7 Performance evaluation for the CNN classifiers based on the On and Off medication gait data

The CNN classifier based on the On and Off medication gait data is constructed based on two CNN models, one processes the On medication gait data and the other processes the Off medication gait data. Both models share the same hyperparameters, the features of the last layer of the each CNN model are added and fed to the feed-forward network (MLP) whose output will be the prediction of each patient gait patterns. The following Figure 11 summarizes the architecture of the developed classifier.

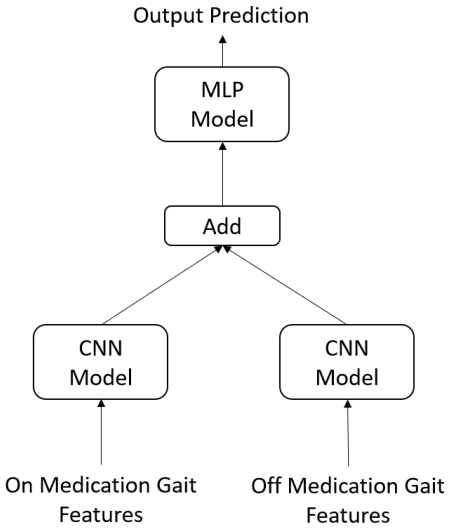


Figure 11: Architecture of the CNN classifier based on the On and Off medication gait data for the classification task of differentiating between IPD and VaP gait patterns.

Different hyperparameters were combined which resulted in the development of 1500 On/Off CNN classifiers for each feature subset of the IPD vs. VaP differentiation task. This work comprises two different datasets and each dataset is evaluated using the three best feature subsets previously selected. In total, the number of developed On/Off CNN classifiers is 9000.

The search of hyperparameters included 15 different values for the first CNN hidden layer of neurons, 17 different values for the second CNN hidden layer of neurons, 15 different values for the first MLP hidden layer of neurons, 17 different values for the second MLP hidden layer of neurons, 3 different values for the

Kernel value (length of the convolution window), 4 different values for the learning rate, 4 different values for the dropout rate, 3 different values for the batch size, and 6 different values for the epochs. From all the possible 56 181 600 combinations between these hyperparameters, 1500 were randomly chosen to develop the On/Off CNN classifiers.

From all the developed On/Off CNN classifiers, the one with the best validation performance on each feature subset was selected and used to make predictions on the test set. Finally, for each dataset the best overall On/Off CNN classifier, based on the validation and test performances, was selected. The results are presented in the following sub-sections.

7.7.1 Performance with Raw All Strides Dataset

IPD patients vs. VaP patients

Table 53: Performance and hyperparameters of the best CNN classifier developed with the 3 Raw All Strides features selected by the backward step-wise method for the IPD vs. VaP differentiation task. The features were collected when the patients were On and Off medication.

Step-Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.85%	92.0%	91.82%
Validation Set	89.17%	96.67%	91.67%
Test Set	100%	100%	100%

On/Off CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.2	32/16	5	32/16	2	30

7.7.2 Performance with Normalized All Strides Dataset

IPD patients vs. VaP patients

Table 54: Performance and hyperparameters of the best CNN classifier developed with the 2 Normalized All Strides features selected by the gini method for the IPD vs. VaP differentiation task. The features were collected when the patients were On and Off medication.

Gini Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.97%	83.49%	90.98%
Validation Set	87.50%	91.67%	95.0%
Test Set	100%	100%	100%

On/Off CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.3	20/16	5	20/16	4	24

7.8 Performance evaluation for the CNN/MLP classifiers based on the On and Off medication gait data and biometric data

The CNN/MLP classifier based on the On and Off medication gait and biometric data (age, weight, height, and gender) of each patient is constructed based on two CNN models, one processes the On medication gait data and the other processes the Off medication gait data, and an MLP model that processes the biometric data. Both CNN models share the same hyperparameters, the features of the last layer of each CNN model are added and the features of the last layer of the MLP model that processes the biometric data are concatenated to the added features of the CNN models. The concatenated features are then fed to the final feed-forward network (MLP) whose output will be the prediction of each patients' gait patterns. The following Figure 12 summarizes the architecture of the developed classifier.

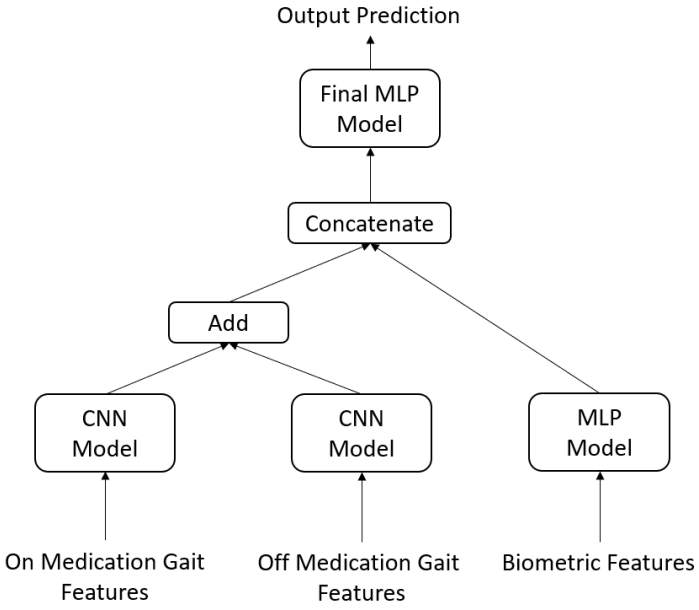


Figure 12: Architecture of the CNN Classifier based on the On and Off medication gait data and the biometric data for the classification task of differentiating between IPD and VaP gait patterns.

Different hyperparameters were combined which resulted in the development of 1500 CNN/MLP classifiers for each feature subset of the IPD vs. VaP differentiation task. This work comprises one dataset (Raw All Strides and biometric data of each patient) that is evaluated using the three best gait feature subsets previously selected. In total, the number of developed CNN/MLP classifiers is 4500.

The search of hyperparameters included 15 different values for the first CNN hidden layer of neurons, 17 different values for the second CNN hidden layer of neurons, 15 different values for the first MLP hidden layer of neurons, 17 different values for the second MLP hidden layer of neurons, 3 different values for the Kernel value (length of the convolution window), 4 different values for the learning rate, 4 different values for the dropout rate, 3 different values for the batch size, and 6 different values for the epochs. The number of neurons of the MLP that processes the biometric data was fixed. From all the possible 56 181 600 combinations between these hyperparameters, 1500 were randomly chosen to develop the CNN/MLP classifiers.

From all the developed CNN/MLP classifiers the one with the best validation performance on each feature subset was selected and used to make predictions on the test set. Finally, the best overall CNN/MLP classifier, based on the validation and test performances, was selected. The results are described next.

7.8.1 Performance with Raw All Strides Dataset

IPD patients vs. VaP patients

Table 55: Performance and hyperparameters of the best CNN/MLP classifier developed with the 7 Raw All Strides features selected by the lasso method for the IPD vs. VaP differentiation task. The classifier was developed based on gait features collected while the patients were On and Off medication and biometric data of each patient.

Lasso Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	87.17%	84.46%	91.67%
Validation Set	92.50%	96.67%	95.0%
Test Set	100%	100%	100%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.001	0.3	20/16	7	16/8	20/16	4	24

7.9 Performance comparison among the different classifiers

In this Section, a comparison between the overall performances of the best classifiers for the differentiation task of controls vs. Parkinsonism and the differentiation task of IPD vs. VaP is summarized in the following tables.

7.9.1 Controls vs. Parkinsonism performance comparison

The following Table 56 summarizes the best results for the Controls vs. Parkinsonism classification task based on the Raw and Normalized All Strides datasets features.

Table 56: Validation and test performance comparison between the best classifiers when trained with All Strides datasets features for the controls vs. Parkinsonism differentiation task.

	Raw All Strides			Normalized All Strides		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM Classifier	96.23%	95.68%	96.95%	**94.65%**	**91.42%**	**98.17%**
Validation/Test	97.96%	92.65%	99.75%	**97.41%**	**90.65%**	**99.75%**
MLP Classifier	91.48%	79.33%	94.67%	90.15%	83.29%	96.77%
Validation/Test	99.44%	98.44%	99.76%	99.63%	100%	99.52%
DBN Classifier	91.70%	84.98%	98.93%	93.03%	88.10%	99.28%
Validation/Test	99.81%	100%	99.76%	98.89%	98.40%	99.04%
LSTM Classifier	**95.50%**	**91.67%**	**100%**	91.0%	83.33%	95.0%
Validation/Test	**100%**	**100%**	**100%**	100%	100%	100%
CNN Classifier	94.67%	88.67%	96.67%	90.17%	88.33%	96.67%
Validation/Test	100%	100%	100%	100%	100%	100%

For each classifier, validation performance is displayed on the first row, and test performance is displayed on the second row.

The following Table 57 summarizes the best results for the Controls vs. Parkinsonism classification task based on the Raw and Normalized Mean Strides datasets features.

Table 57: Validation and test performance comparison between the best classifiers when trained with Mean Strides datasets features for the controls vs. Parkinsonism differentiation task.

	Raw Mean Strides			Normalized Mean Strides		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM Classifier	95.50%	91.67%	100%	93.50%	90.0%	100%
Validation/Test	100%	100%	100%	100%	100%	100%
MLP Classifier	93.50%	88.33%	100%	**95.50%**	**91.67%**	**100%**
Validation/Test	100%	100%	100%	**100%**	**100%**	**100%**
DBN Classifier	**96.0%**	**95.0%**	**100%**	**95.50%**	**91.67%**	**100%**
Validation/Test	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

For each classifier, validation performance is displayed on the first row, and test performance is displayed on the second row.

7.9.2 IPD vs. VaP performance comparison

The following Table 58 summarizes the best results for the IPD vs. VaP classification task based on the Raw and Normalized All Strides datasets.

Table 58: Validation and test performance comparison between the best classifiers when trained with All Strides datasets for the IPD vs. VaP differentiation task.

	Raw All Strides			Normalized All Strides		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM Classifier	68.86% 78.44%	65.07% 92.37%	56.39% 65.34%	80.16% 95.89%	84.42% 93.85%	66.65% 100%
MLP Classifier	71.52% 79.88%	64.14% 88.76%	67.17% 69.09%	65.77% 96.71%	70.81% 99.66%	69.46% 92.35%
DBN Classifier	66.66% 86.24%	63.66% 81.99%	66.00% 100%	78.37% 94.66%	67.21% 92.15%	89.47% 100%
LSTM Classifier	74.17% 83.33%	76.67% 75.0%	85.0% 100%	82.50% 100%	81.67% 100%	80.0% 100%
CNN Classifier	71.67% 100%	70.0% 100%	76.67% 100%	82.50% 100%	81.67% 100%	80.0% 100%
CNN Classifier On/Off gait data	89.17% 100%	96.67% 100%	91.67% 100%	**87.50%** **100%**	**91.67%** **100%**	**95.0%** **100%**
CNN Classifier Bio data + On/Off gait data	**92.50%** **100%**	**96.67%** **100%**	**95.0%** **100%**	– –	– –	– –

For each classifier, validation performance is displayed on the first row, and test performance is displayed on the second row.

The following Table 59 summarizes the best results for the IPD vs. VaP classification task based on the Raw and Normalized Mean Strides datasets.

Table 59: Validation and test performance comparison between the best classifiers when trained with Mean Strides datasets for the IPD vs. VaP differentiation task.

	Raw Mean Strides			Normalized Mean Strides		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM Classifier	74.17%	80.0%	75.0%	**81.67%**	**85.0%**	**85.0%**
Validation/Test	83.33%	100%	75.0%	**100%**	**100%**	**100%**
MLP Classifier	**84.17%**	**81.67%**	**81.67%**	77.50%	86.37%	80.0%
Validation/Test	**83.33%**	**75.0%**	**100%**	83.33%	100%	75.0%
DBN Classifier	74.17%	86.67%	76.67%	76.67%	90.0%	86.67%
Validation/Test	83.33%	100%	75.0%	100%	100%	100%

For each classifier, validation performance is displayed on the first row, and test performance is displayed on the second row.

7.10 Clustering of Parkinsonism patients based on gait data

K-means clustering and hierarchical clustering based on the Mean Strides datasets were performed to identify subgroups of patients with parkinsonism (IPD + VaP) with similar gait characteristics. Two subsets of gait features were used: (1) the subset of gait features that achieved the best classification performance results; (2) toe clearance and speed-related gait features such as cadence, speed and stride length. This second choice of features was based on previous studies showing that Parkinsonism patients display lower speed, shorter stride length and lower cadence (Lord et al. (2013), Winikates and Jankovic (1999), Bäzner et al. (2000)), toe clearance measures were also selected since these have proven to be powerful predictors of gait patterns based on the feature selection and classification results of this work.

The clustering was performed with both raw and MR normalized gait features. The approach was the following: (1) perform feature extraction using PCA with the objective of extracting two PCs, this allows graphical visualization of the results, (2) assess the performance of the two clustering methods (k-means and hierarchical clustering) using the raw and normalized subset of gait features with the number of clusters pre-defined to two, (3) select the best clustering method based on the clustering classification performance, (4) perform clustering analysis with the selected method, the number of optimal clusters is identified, depending on the selected clustering method, using the elbow method or the dendrogram and finally (5) interpret the clustering results.

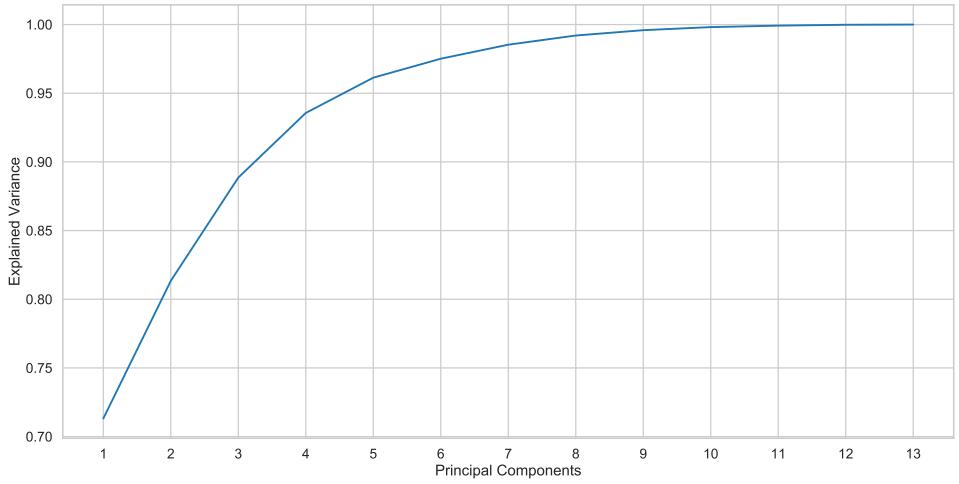
7.10.1 Clustering using the best performance classification features

The best normalized Mean Strides gait features, based on the classification performance of the previously developed classifiers, were **CV of speed**, **ground cycle duration**, **peak swing**, **maximum toe clearance 1** and **stride length**.

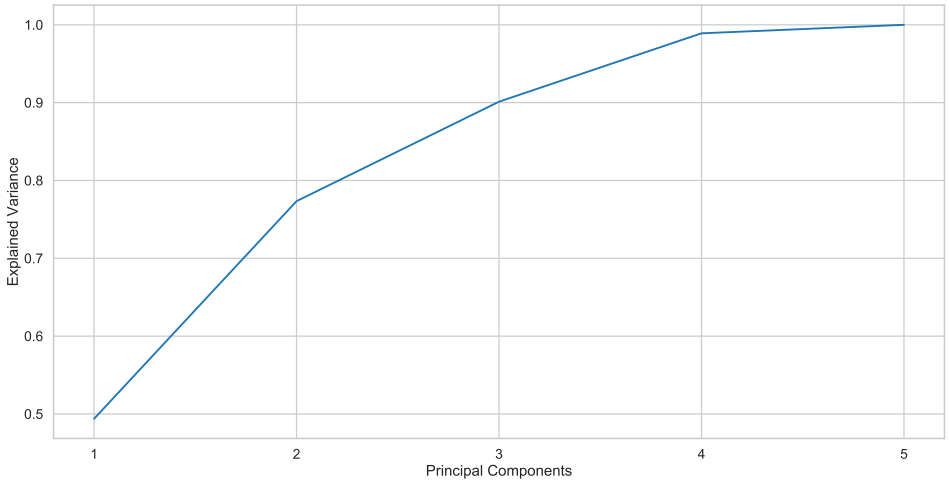
The best raw Mean Strides gait features, based on the classification performance of the previously developed classifiers, were **lift-off angle**, **maximum heal clearance**, **stride length**, **speed**, **CV of speed**, **maximum toe clearance 2**, **peak swing**, **CV of minimum toe clearance**, **strike angle**,

CV of maximum heel clearance, **CV of stride length**, **CV of lift-off angle** and **pushing**.

Principal components analysis was applied to both the raw and normalized gait features, the resulting PCs are summarized in the following Figure 13.



(a) Raw Principal Components



(b) Normalized Principal Components

Figure 13: Principal Components and the respective cumulative sum of the explained variance.

The raw and normalized gait features were compressed to the first two PCs. The first two raw PCs

account for 81.31% of the total variance and the first two normalized PCs account for 77.34% of the total variance. The selected raw and normalized PCs were used to perform the clustering analysis.

To assess the classification performance of the clustering methods the number of clusters was pre-defined to two. The results are summarized in the following Table 60.

Table 60: Clustering performance for the IPD vs. VaP classification task based on the two raw and normalized PCs with the number of clusters pre-defined to two.

	K Means Clustering			Hierarchical Clustering		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Raw PCs	68.97 %	77.78 %	65.00 %	62.07 %	71.43 %	59.09 %
Normalized PCs	65.52 %	75.00 %	61.90 %	62.07 %	80.00 %	58.33 %

From the analysis of Table 60 it can be concluded that the best performance was achieved by the K-means method based on the two PCs obtained from the raw gait features. The K-means clustering was used to perform clustering analysis of patients with parkinsonism based on the two raw PCs. The optimal number of clusters was selected with the assistance of the elbow method, the following Figure 14 shows the elbow method results.

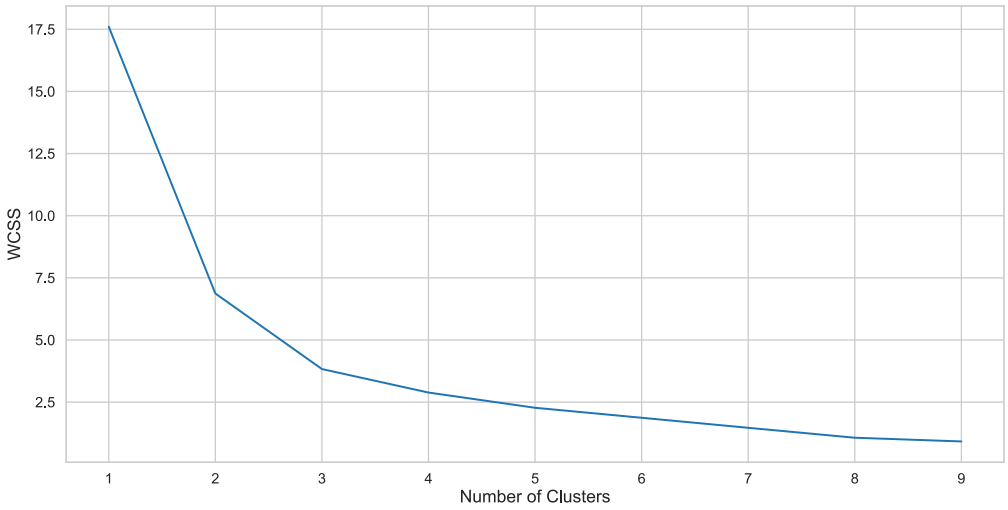


Figure 14: Elbow Method using the two raw PCs.

With the assistance of the elbow method results (see Figure 14), the selected optimal number of clusters was three. The following figure 15 and Table 61 show the results of K-means clustering based on the two raw PCs with the number of clusters set to three.

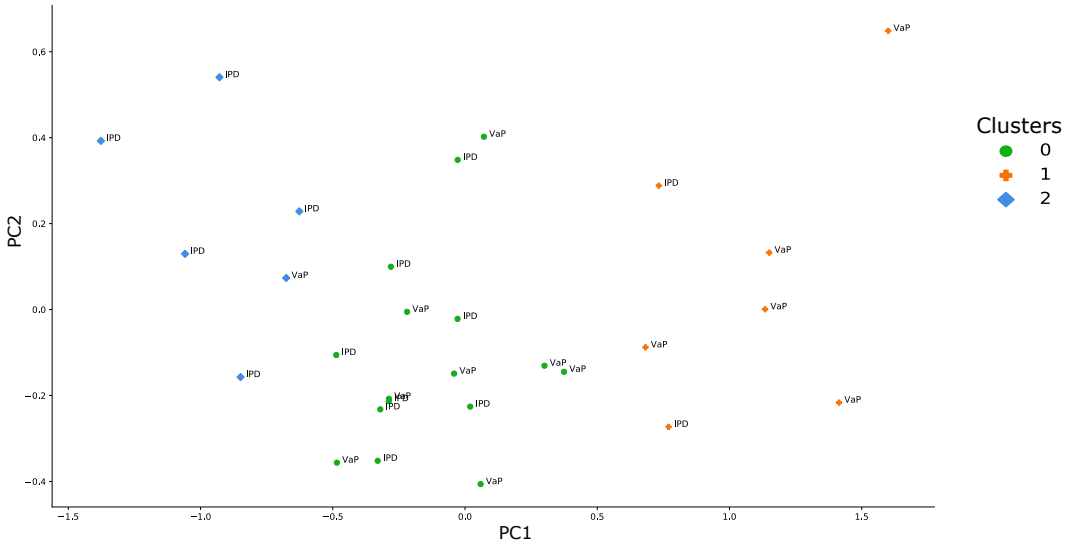


Figure 15: K-means clustering based on the two raw PCs with the number of clusters set to three.

Table 61: Mean values of each gait feature for each cluster found by the K-means clustering analysis based on the two raw PCs.

Feature	**Cluster 1** (8/7)	**Cluster 2** (5/2)	**Cluster 3** (1/5)
Lift-Off Angle	-49.05	-32.09	-59.49
Max Heal Clearance	0.219	0.169	0.280
Stride Length	0.863	0.503	1.21
Speed	0.729	0.448	1.09
CV of Speed	7.65	15.36	5.203
Max Toe Clearance 2	0.089	0.048	0.138
Peak Swing	269.47	191.64	334.66
CV of Min Toe Clearance	10.49	9.30	10.37
Strike Angle	12.84	4.31	22.25

CV of Max Heal Clearance	5.65	15.06	3.83
CV of Stride Length	5.78	15.61	3.87
CV of Lift-Off Angle	-5.87	-12.01	-3.82
CV of Pushing	24.24	18.49	26.94

The following conclusions can be drawn from the analysis of Table 61, the first cluster is composed of 8 VaP patients and 7 IPD patients, this cluster is evenly composed of patients of both disorders and is characterized by mildly impacted gait patterns. The speed is low and stride length is small, the foot clearance measures indicate restricted foot dexterity and flexibility, finally, the CVs indicate that the patients are able to maintain consistent gait patterns throughout the walk.

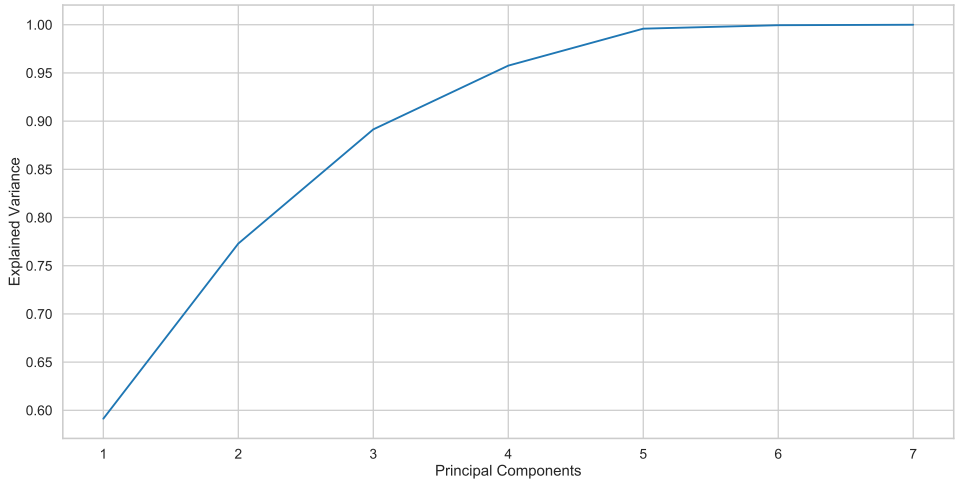
The second cluster is composed of 5 VaP patients and 2 IPD patients. This cluster represents the patients with the highest gait impairment, the speed is extremely low, stride length is extremely small, the foot clearance measures indicate poor foot dexterity, flat-footed gait patterns and dragging of the feet. Additionally, from the analysis of the CVs, it can be concluded that the gait patterns described by this cluster are very inconsistent, the patients do not seem to be able to maintain a constant rhythm throughout the walking course.

Finally, the third cluster is composed of 1 VaP patient and 5 IPD patients. This cluster represents the best gait patterns among all clusters. The speed and stride length are the highest among all clusters, the foot clearance measures indicate good foot dexterity, balance, and flexibility, the patients seem to be able to land on their heel and push with their toes. The gait patterns are also steady and consistent. Overall this cluster represents the less impaired gait patterns and is mainly composed of IPD patients.

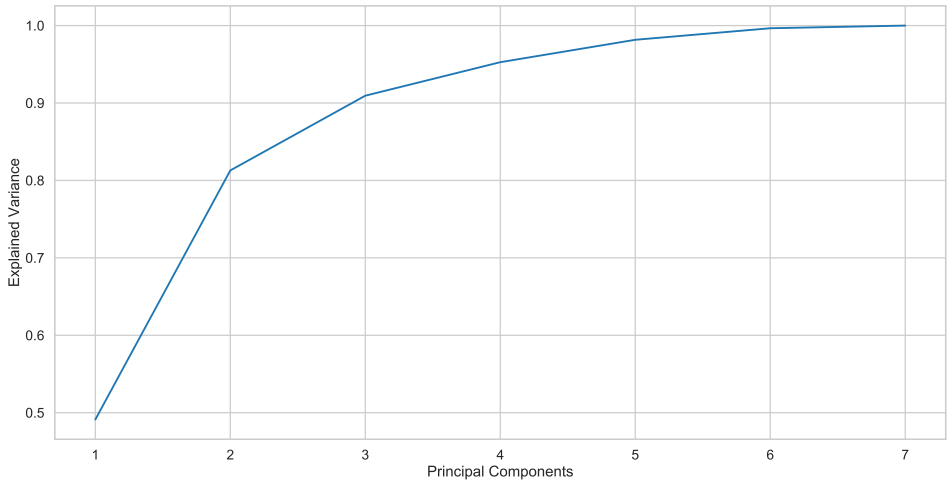
7.10.2 Clustering using toe clearance and speed related gait features

The Mean Strides gait features used in this analysis were **cadence**, **speed**, **stride length**, **maximum toe clearance 1**, **maximum heel clearance**, **maximum toe clearance 2** and **minimum toe clearance**. The analysis was performed using the raw and normalized values of this subset of features.

Principal components analysis was applied to both the raw and normalized gait features, the resulting PCs are summarized in the following Figure 16.



(a) Raw Principal Components



(b) Normalized Principal Components

Figure 16: Principal components and the respective cumulative sum of the explained variance.

The raw and normalized gait features were compressed to the first two PCs. The first two raw PCs account for 77.29% of the total variance and the first two normalized PCs account for 81.28% of the total variance. The selected raw and normalized PCs were used to perform the clustering analysis.

To assess the classification performance of the clustering methods the number of clusters was pre-defined to two. The results are summarized in the following Table 62.

Table 62: Clustering performance for the IPD vs. VaP classification task based on the two raw and normalized PCs with the number of clusters pre-defined to two.

	K Means Clustering			Hierarchical Clustering		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Raw PCs	62.07 %	57.89 %	70.00 %	58.62 %	54.17 %	80.00%
Normalized PCs	65.52 %	75.00 %	61.90 %	65.52 %	75.00 %	61.90 %

From the analysis of Table 62 it can be concluded that the performance results of both methods are similar for the normalized PCs, however, with the raw PCs the K-means method achieved a greater performance being the overall superior method. K-means clustering was then used to perform clustering analysis of patients with parkinsonism based on the two normalized PCs. The optimal number of clusters was selected with the assistance of the elbow method, the following Figure 17 shows the elbow method results.

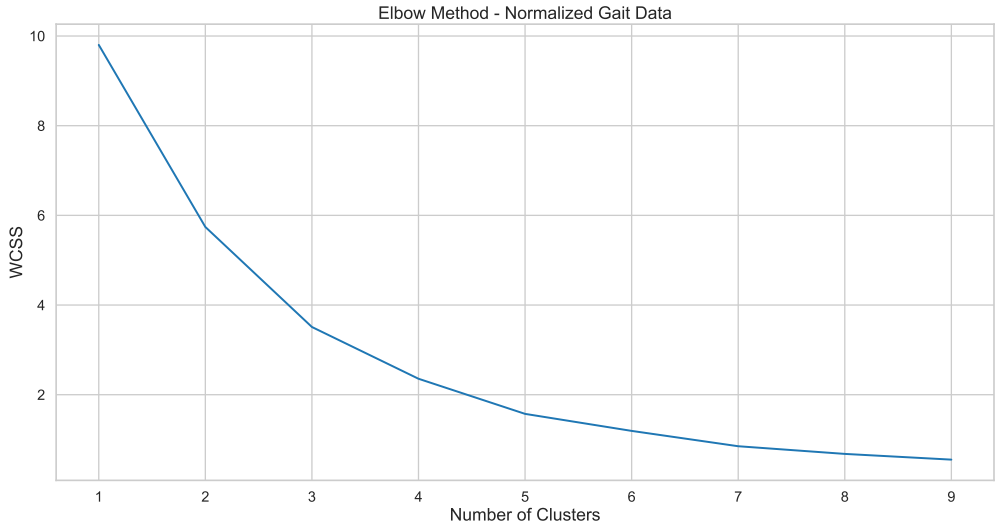


Figure 17: Elbow Method using the two Normalized PCs.

With the assistance of the elbow method results (see Figure 17), the selected optimal number of

clusters was three. The following Figure 18 and Table 63 show the results of K-means clustering based on the two normalized PCs with the number of clusters set to three.

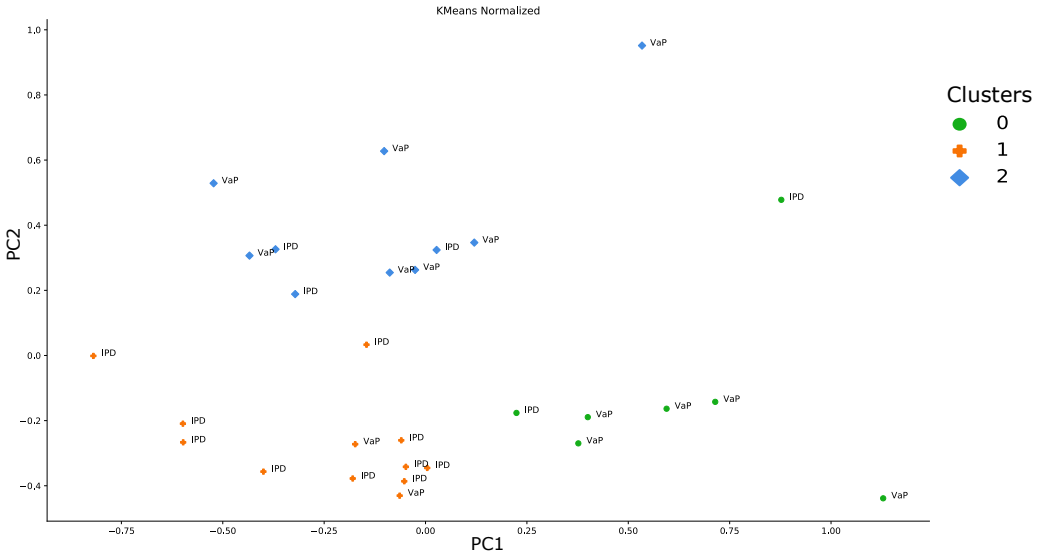


Figure 18: K-Means Clustering using the two normalized PCs and the number of clusters set to three.

Table 63: Mean values of each gait feature for each cluster found with K-means clustering based on the two normalized PCs.

Feature	**Cluster 1** (5/2)	**Cluster 2** (2/10)	**Cluster 3** (7/3)
Cadence	107.61	103.74	101.50
Stride Length	0.529	1.02	0.865
Speed	0.473	0.889	0.737
Max Heel Clearance	0.181	0.235	0.228
Max Toe Clearance 1	0.070	0.069	0.087
Min Toe Clearance	0.030	0.036	0.037
Max Toe Clearance 2	0.047	0.108	0.096

The following conclusions can be drawn from the analysis of Table 63, the first cluster is composed of 5 VaP patients and 2 IPD patients. This cluster is characterized by the lowest speed, smallest stride

length and the highest cadence among all clusters. The foot clearance metrics indicate low foot agility and dexterity causing flat-footed gait patterns. The maximum toe clearance 2 and maximum heel clearance are low which indicates that the patients are not landing with the heel, the patients are instead lading and pushing with a flat foot. Overall this cluster represents the worst gait patterns comparing to the other clusters.

The second cluster is composed of 2 VaP patients and 10 IPD patients. This cluster represents the highest stride length and speed among the clusters. The foot clearance metrics indicate good foot agility and balance. The maximum toe clearance 2 measure indicates that the patients are landing with the heel, the maximum heel clearance measure is also a good indicator that the patients are pushing off the ground with the front of the foot. Overall this cluster represents the best gait patterns among the three clusters and is mainly composed of IPD patients.

The third and final cluster is composed of 7 VaP patients and 3 IPD patients. This cluster represents good foot clearance measures, similar to those of the second cluster. However, the speed, cadence and stride length measures are lower than those of the second cluster. Overall this cluster represents slightly impaired gait patterns, the cluster is composed of patients that display slightly slow gait patterns associated with good balance and foot agility.

7.11 Parkinsonism Results Discussion

7.11.1 General discussion

One of the main goals of this work was to exhaustively investigate the performance of machine learning classifiers based on gait data acquired by wearable sensors to identify differences between two groups of individuals: controls vs. Parkinsonism patients and IPD vs. VaP patients. Previous studies have shown that a gait normalization approach using MR has the potential to minimize the effect of inter-subject physical differences and self-selected speed on gait features which improves the classifiers' ability to differentiate normal gait patterns from parkinsonian gait patterns (Mikos et al. (2018); Wahid et al. (2016, 2015)). This

hypothesis was tested in the present work, MR gait normalization was implemented which generated two normalized gait datasets, resulting in a total of four gait datasets. For the development of the classifiers, the four datasets including the Raw All Strides, Normalized All Strides, Raw Mean Strides, and Normalized Mean Strides were used and their respective performance results were compared.

Before addressing the developed machine learning classifiers it is important to address the dimensionality of the datasets used in this work. The collected datasets present a high number of features, some of these features might be irrelevant to the classification tasks being tackled, irrelevant features should be addressed and removed since these increase the computational cost of the classifiers and might also lead to performance decreases (Alpaydin (2010)). Dimensionality reduction is also important given the sample size of this work, a large number of features given a small sample size might lead to over-fitting which also decreases performance. Over-fitting is a consequence of learning with the data instead of learning the patterns that underlies the data (Mohri et al. (2012)). To improve the performance of the classifiers dimensionality reduction using feature selection methods was implemented. This work made use of three feature selection methods: backward step-wise, lasso and gini. Feature selection has been shown to result in enhanced performance, a reduced hypothesis search space (increased interpretability), and reduced computational cost (Hall and Smith (1998), Alpaydin (2010)).

The implemented feature selection approach consisted of three feature selection methods and four SVM classifiers, the only difference between the SVM classifiers was the kernel function. This approach worked as follows, the feature selection methods reduced the number of features one step at a time, at each iteration the performance of the four SVM classifiers was evaluated based on the selected features using 10-fold stratified cross validation (approach detailed in section 7.1). For each dataset, each feature selection method and each differentiation task the subset of features that achieved the best performance was chosen. The best subsets of features were used to develop the machine learning classifiers.

The proposed feature selection approach achieved great results, the performance of the four SVM classifiers always increased when based on a subset of the original features. The SVM kernel that seems to achieve the best performance was the linear kernel, the feature selection method that finds the best subset of features based on the SVMs performances was the gini (mean accuracy: 87.49 ± 8.12), the lasso method performed slightly worse (mean accuracy: 86.02 ± 7.52) and finally, the backward step wise method achieved the worst performance overall (mean accuracy: 81.88 ± 11.21). These results

further corroborate previous findings where feature selection methods have shown a positive effect on the performance of machine learning classifiers when working with gait variables for the classification of IPD patients (Javed et al. (2018), Rovini et al. (2018)).

For the classification of controls vs. Parkinsonism and IPD vs. VaP, five machine learning classifier architectures were used these include SVMs, MLPs, DBNs, LSTMs, and CNNs. Each classifier was implemented based on the selected subsets of gait features for each classification task using various combinations of hyperparameters. The performance of each classifier was evaluated based on the average of the validation (obtained by 10-fold stratified cross-validation) and test performances.

7.11.2 Controls vs. Parkinsonism

Classification performance of controls vs. Parkinsonism patients was extremely good for all datasets. The best overall performance was achieved by the DBN classifier based on the 5 Raw Mean Strides features selected by the lasso method, the classifier obtained an accuracy of 98.0% $\pm$ 2.0, a sensibility of 97.5% $\pm$ 2.5 and a specificity of 100% $\pm$ 0.0 (results are shown as mean of validation and test performance $\pm$ standard deviation). The 5 Raw Mean Strides features were cadence, speed, peak swing, CV of pushing and CV of maximum toe clearance 2. These results show that slow cadence, low speed, slow peak swing and variations in pushing and maximum toe clearance 2 while walking are powerful predictors of parkinsonian gait patterns, these findings are aligned with previous studies showing that Parkinsonism patients exhibit slowness of gait, shorter steps and shuffling of the feet (Lord et al. (2013), Winikates and Jankovic (1999), Bäzner et al. (2000)).

The classifier that achieved the worst results was the MLP based on 2 Normalized All Strides features selected by the lasso method, the classifier obtained an accuracy of 94.89% $\pm$ 4.74, a sensibility of 91.65% $\pm$ 8.35 and a specificity of 98.15% $\pm$ 1.38, the 2 Normalized All Strides features were speed and maximum heel clearance. These findings indicate that the classification results are extremely consistent, the worst classifier was able to obtain a mean accuracy of 94.89% which is comparable to the best initial results obtained by this work (Fernandes et al. (2018)). Foot clearance metrics, especially maximum heel clearance and maximum toe clearance 2, have proven to be extremely powerful predictors of parkinsonian

gait patterns, these features were constantly among the best subset of features selected by the feature selection methods. It can be concluded that these foot clearance features are extremely important and should be considered in future analysis of parkinsonian gait.

All classifiers achieved a higher specificity than sensitivity when differentiating between controls and Parkinsonism patients. These results indicate that the classifiers do not predict that a Parkinsonism subject displays normal gait patterns. The mispredictions that the classifiers make occur when the classifier predicts that controls display parkinsonian gait patterns, this is a desired characteristic because the cost associated with the misclassification of a Parkinsonism patient is greater compared to the misclassification of a healthy subject. Furthermore, the overall performance of the normalized gait datasets was very similar to the performance obtained using the raw gait datasets.

To conclude, the classification results achieved in this work are comparable to results achieved by previous studies that used ANNs, SVMs, RFs, K-Nearest Neighbours, among other classifiers based on gait data to differentiate between controls and IPD patients. Using ANNs classifiers Manap et al. achieved between 81.25% and 95.63% accuracy (Manap et al. (2011)). Other study conducted by Tahir and Manap achieved between 66.40% and 100% accuracy using SVMs and ANNs classifiers (Tahir and Manap (2012)). More recently Wahid et al. implemented an RF classifier that achieved an accuracy of 92.6% when classifying controls vs. IPD patients based on MR normalized gait data. (Wahid et al. (2015)). All these studies trained the classifiers using various configurations of hyperparameters, the performance metrics where evaluated based on k-fold cross-validation.

7.11.3 VaP vs. IPD

When trying to distinguish between VaP and IPD patients, the performance of the classifiers slightly decreased and a higher spectrum of performances was observed compared to the controls vs. Parkinsonism classification task. The decrease in performance shows the difficulty in the differentiation between these disorders. Nevertheless, good performance results were obtained when making predictions, the best overall classification results were obtained when using All Strides datasets and time-series specific classifiers such as the LSTM and CNN, specifically when taking into account the effect of levodopa medi-

cation. These results indicate that time-series specific classifiers are better able to capture the underlying gait patterns of VaP and IPD patients, this might result from the fact that walking events of IPD and VaP patients present high variability so, time-series classifiers ability to connect walking events seem to be an advantage when detecting the underlying gait patterns displayed by IPD and VaP patients.

The best classifier was the CNN based on 3 Raw All Strides dataset features collected while the patients were on and off medication. The 3 features were lift-off angle, peak swing and stride length, these were selected by the step-wise method. This CNN classifier achieved an accuracy of 94.59% $\pm$ 5.41, a sensibility of 98.34% $\pm$ 1.66 and a specificity of 95.84% $\pm$ 4.16. These classification results were further improved when the biometric data of each patient was used as input alongside the gait features collected while the patients were on and off medication, the CNN classifier based on the biometric data and the same 3 on/off medication Raw All Strides features achieved an accuracy of 96.25% $\pm$ 3.75, a sensibility of 98.33% $\pm$ 1.66 and a specificity of 97.50% $\pm$ 2.50. These results show that the response of gait data to levodopa presents valuable information when classifying VaP vs. IPD patients, these findings are aligned with previous studies showing that VaP patients do not respond well to levodopa medication compared to IPD patients, the walking events of VaP patients with and without medication are similar while the walking events of IPD patients improve while they have the effect of levodopa medication (Miguel-Puga et al. (2017), Gupta and Kuruvilla (2011)). Feeding the physical properties of the patients as input along side the gait features was an effort to make the classifier learn the impact of physical properties on gait patterns. The architecture proposed in this work was successful and achieved the best VaP vs. IPD classification results. These findings suggest that this approach might be useful to simultaneously perform classification and normalization of gait data according to the patients' physical properties, however, further investigation is required.

Still focusing on the VaP vs. IPD classification task, in general, the normalized gait datasets achieved slightly greater results than the raw gait datasets. However, the differences in performance between these gait datasets were not very accentuated. Regarding the SVMs, MLPs and DBNs architectures, the best classifier was the SVM based on 5 Normalized Mean Strides features selected by the step-wise method, this classifier achieved an accuracy of 90.84% $\pm$ 9.17, a sensibility of 92.50% $\pm$ 7.50 and a specificity of 92.50% $\pm$ 7.50, the 5 features were CV of speed, cycle duration, peak swing, maximum toe clearance 1 and stride length. These results provide further contributions to the research started in 2018 which resulted

in the first publication that included the analysis and prediction of a VaP and an IPD group (Fernandes et al. (2018)).

7.11.4 Clustering of Parkinsonism patients

A recent study performed clustering analysis over motor and non-motor features of 951 IPD patients and has concluded the existence of homogeneous subgroups of IPD patients (Mu et al. (2017)). In this work, clustering analysis was applied to explore the existence of subgroups between IPD and VaP patients based on mean strides gait measurements. Two clustering analysis were conducted, the first used the subset of Mean Strides features that achieved the best performance when classifying between IPD and VaP patients, the second clustering analysis used toe clearance and speed-related features (speed, stride length, and cadence). This second choice of features was based on previous studies showing that Parkinsonism patients display lower speed, shorter stride length and lower cadence (Lord et al. (2013), Winikates and Jankovic (1999), Bäzner et al. (2000)), toe clearance measures were also selected since these have proven to be powerful predictors of gait patterns based on the feature selection and classification results of this work.

The first step of the clustering analysis was data dimensionality reduction using PCA, which resulted in the extraction of two raw PCs and two normalized PCs. The classification performance of K-means and hierarchical clustering was conducted with the number of clusters pre-defined to two. In the first clustering analysis, the superior clustering method was the K-means clustering based on the two raw PCs, this method achieved an accuracy of 68.97%. The K-means method based on the two raw PCs was used to implement the clustering analysis. From the analysis of the elbow method, the optimal number of clusters was three. The clustering analysis found the following 3 subgroups: the first cluster was composed of 8 IPD and 7 VaP patients (53.33%/46.67%), this cluster was characterized by slow, rigid but relatively steady (not much variation) gait patterns, the second cluster was composed of 2 IPD and 5 VaP patients (28.57%/71.43%) and was characterized by extremely slow, highly variable and limited range of motion gait patterns, the final group was composed of 5 IPD patients and 1 VaP patient (83.33%/16.67%), this last cluster was characterized by good, consistent and stable gait patterns, this third cluster represented

the best gait patterns among all clusters.

In the second clustering analysis, the superior clustering method was the K-means clustering based on two normalized PCs, the K-means achieved a classification accuracy of 65.52%. From the analysis of the elbow method, the optimal number of clusters was three. The resulting clusters captured similar characteristics to those of the first clustering analysis, the main difference was the number of patients composing each cluster, in this second analysis there was not a cluster with an even number of VaP and IPD patients. The first cluster was composed of 2 IPD and 5 VaP patients (28.57%/71.43%), the second cluster was composed of 2 IPD and 10 VaP patients (16.67%/83.33%) and finally, the third cluster was composed of 7 IPD and 3 VaP patients (70.0%/30.0%). Similar to the first clustering analysis, the third cluster was characterized by the best gait patterns.

These results are very consistent, both clustering analysis based on different gait features achieved very similar results. These findings indicate that subgroups of homogeneous patients between IPD and VaP might exist and that these diseases affect gait patterns in multiples ways. The underlying detected patterns indicated that VaP patients presented greater difficulties and displayed worst gait patterns than the IPD patients, however, there were IPD patients that also presented extremely affected gait patterns since there were no clusters composed of only one disease. These findings are aligned with previous studies stating that VaP patients display greater gait impairments when comparing to IPD patients (Lehosit and Leslie J. (2015), Zijlmans et al. (2004)). In both cluster analysis the number of optimal clusters was three and the characteristics of the clusters indicated that there was a group that displayed extremely affect gait patterns (this group was mainly composed of VaP patients), a group that displayed mildly affected gait patterns (this group was more evenly composed but the majority were VaP patients) and finally, a group that displayed impacted but still reasonably good gait patterns (this group was mainly composed of IPD patients).

Chapter 8

Results Fabry Disease

This chapter details the feature selection results and the performance results of the various classifiers developed to tackle the FD with CNS lesions vs. FD without CNS lesions classification task. When performing classification with the SVM, MLP or DBN classifiers based on All Strides datasets the individual test performance is computed based on the following: a subject is classified as belonging to a group if more than 50% of his strides are classified as being of that group.

The following sections only detail the best performance results, all results can be found in the Appendix.

8.1 Feature selection

In this Section, the feature selection results for the FD without CNS lesions vs. FD with CNS lesions classification task based on the four FD gait datasets composed of 20 FD patients (7 without CNS lesions and 13 CNS with lesions) are described. Feature scaling has been performed before feature selection ensuring that the range of features does not impact the results.

The FD gait feature selection follows the procedure detailed in Section 7.1.

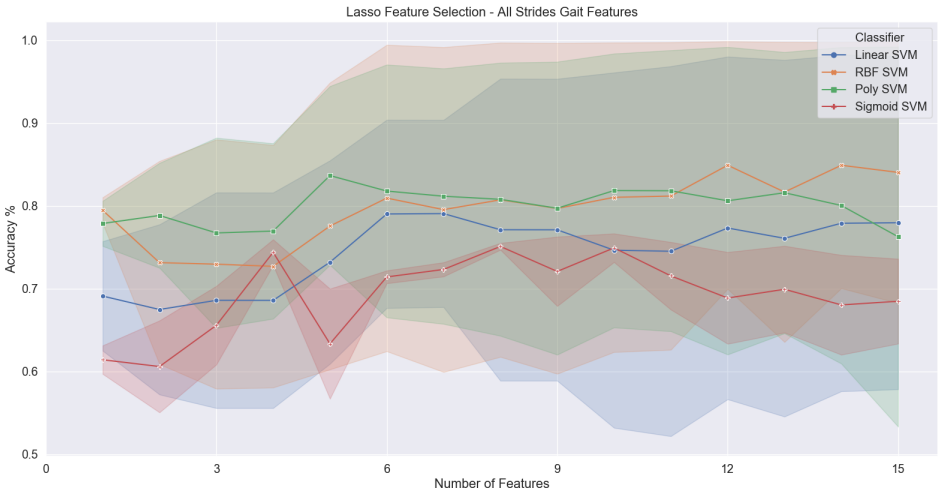
The feature selection results for each gait dataset are described next. In the following plots, each point represents the mean of the final training and validation accuracies. The width represents the 95% CI. The objective is to choose the subset of features with the highest average accuracy and the lowest CI (width),

that is the classifier is neither over-fitting or under-fitting to the training data.

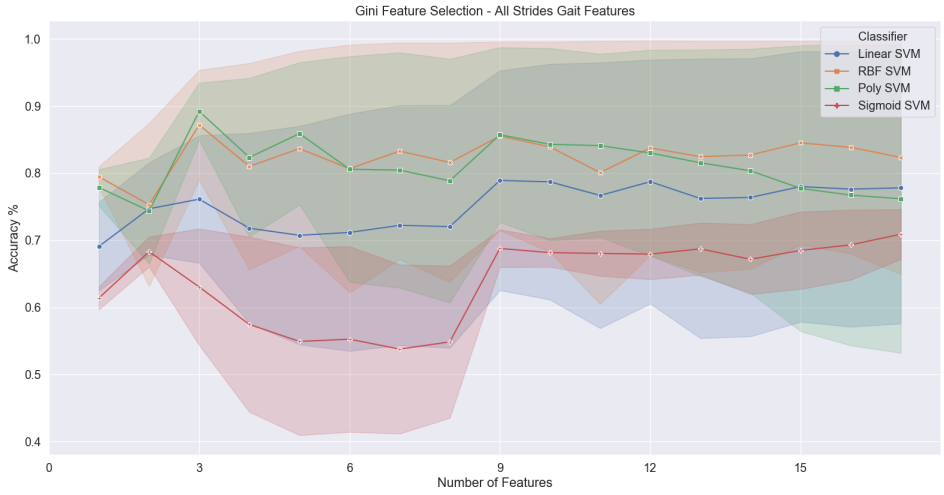
8.1.1 Feature selection - Dataset Raw All Strides

Patients with CNS lesions vs. patients without CNS lesions

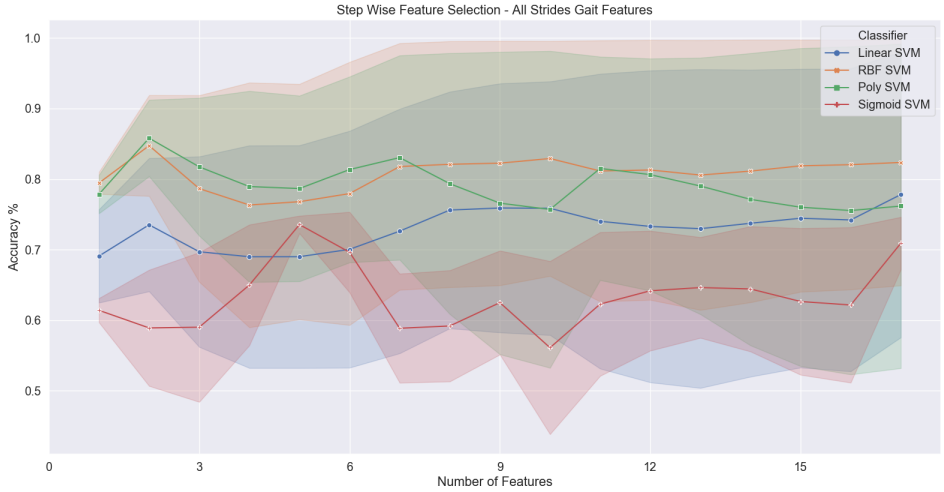
The following Figure 19 shows the feature selection results for the task of distinguishing between FD patients without CNS lesions and FD patients with CNS lesions based on the Raw All Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 19: Raw All Strides dataset feature selection with lasso, gini and backward step-wise for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task.

The following Table 64 summarizes the best results achieved by the feature selection methods for the FD without CNS lesions vs. FD with CNS lesions differentiation task. The selection of 5 features with the lasso method is based on the trade-off between average accuracy and CI, in this case, having the highest average accuracy was considered most important than the smaller CI.

Table 64: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task based on Raw All Strides dataset.

	Kernel	Training Accuracy	Validation Accuracy	Mean Accuracy	Number Features
Lasso	Polynomial	94.46%	72.86%	83.65%	**5**
Gini	Polynomial	93.48%	84.98%	89.23%	**3**
Step-Wise	Polynomial	91.23%	80.39%	85.81%	**2**

The 5 features selected by the lasso method were **stance**, **peak swing**, **strike angle**, **lift-off angle** and **minimum toe clearance**.

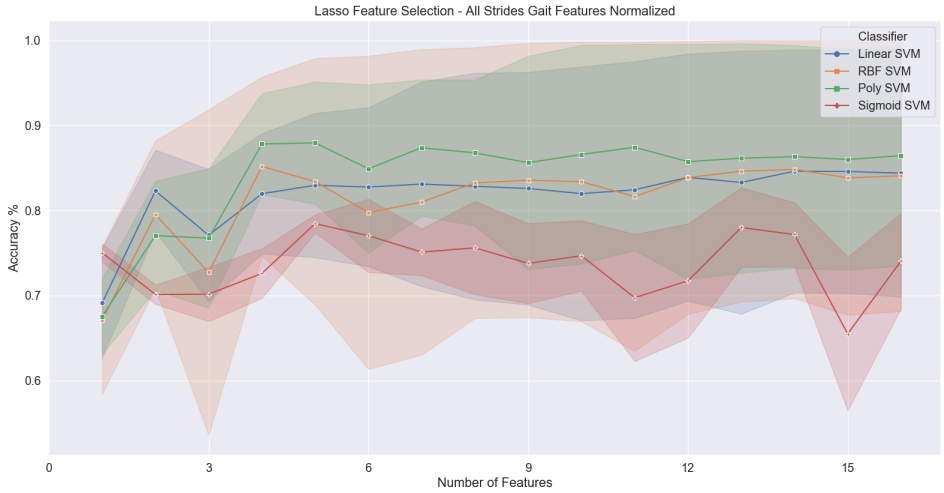
The 3 features selected by the gini method were **lift-off angle**, **stride length** and **maximum toe clearance 2**.

The 2 features selected by the backward step-wise method were **lift-off angle** and **stride length**.

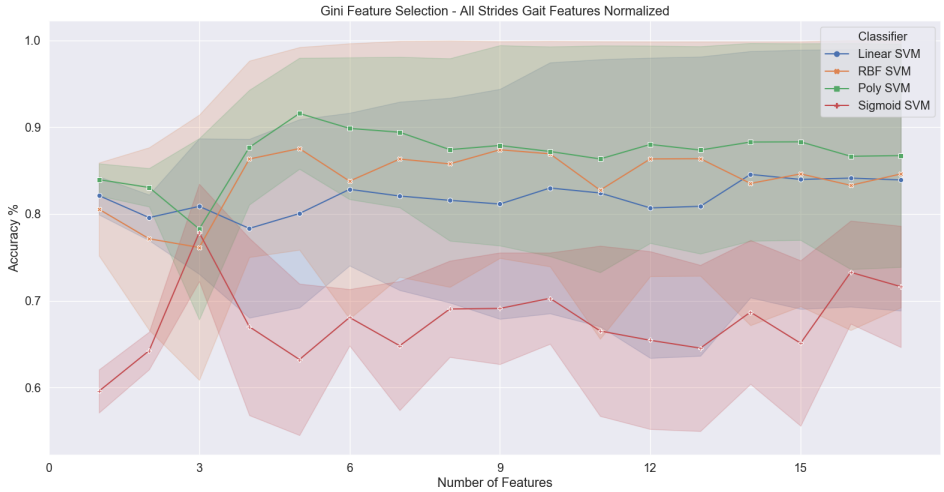
8.1.2 Feature selection - Dataset Normalized All Strides

Patients with CNS lesions vs. patients without CNS lesions

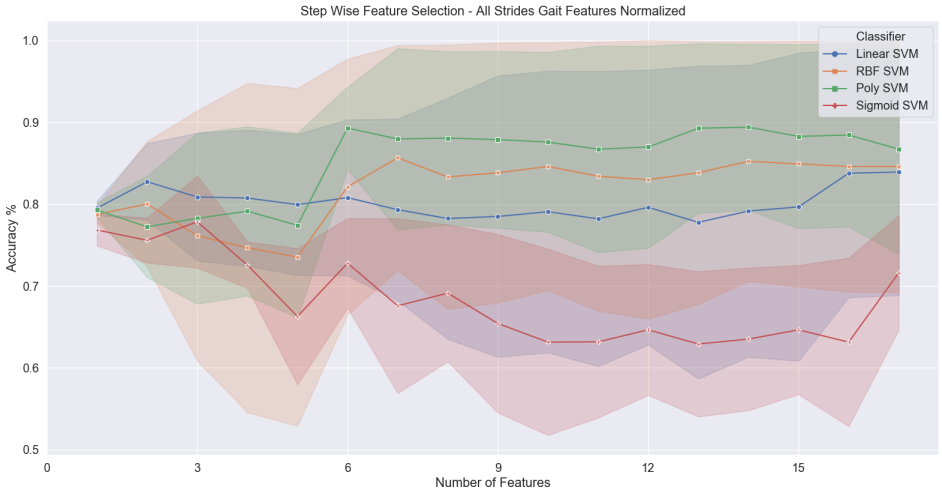
The following Figure 20 shows the feature selection results for the task of distinguishing between FD patients without CNS lesions and FD patients with CNS lesions based on the Normalized All Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 20: Normalized All Strides dataset feature selection with lasso, gini and backward step-wise for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task.

The following Table 65 summarizes the best results achieved by the feature selection methods for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task.

Table 65: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task based on the Normalized All Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Polynomial	93.79%	81.86%	87.82%	**4**
Gini	Polynomial	97.99%	85.18%	91.59%	**5**
Step-Wise	Polynomial	94.33%	84.26%	89.30%	**6**

The 4 features selected by the lasso method were **double support**, **speed**, **strike angle** and **minimum toe clearance**.

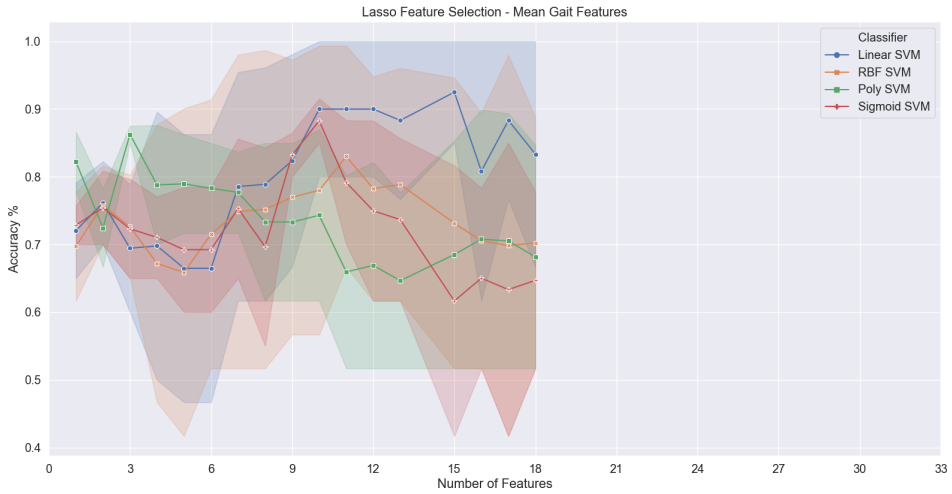
The 5 features selected by the gini method were **maximum heel clearance**, **maximum toe clearance 1**, **speed**, **pushing** and **strike angle**.

The 6 features selected by the backward step-wise method were **maximum toe clearance 1**, **speed**, **maximum heel clearance**, **double support**, **minimum toe clearance** and **strike angle**.

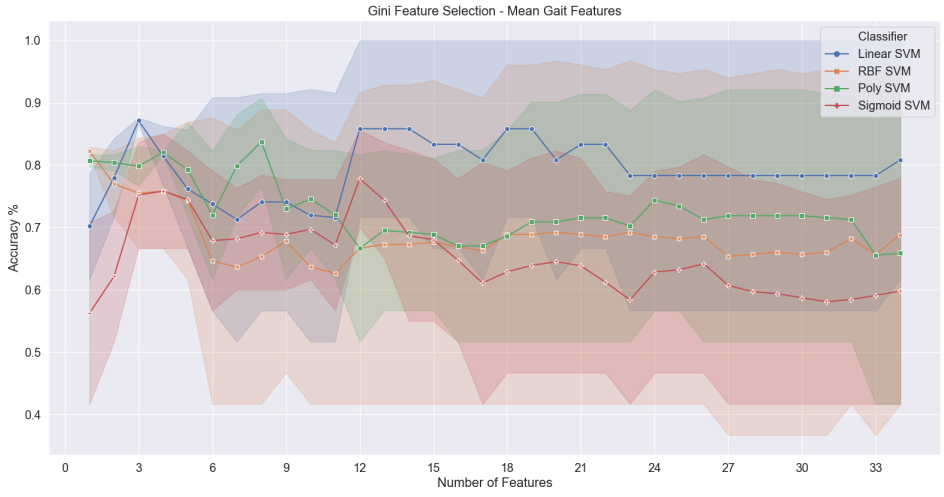
8.1.3 Feature selection - Dataset Raw Mean Strides

Patients with CNS lesions vs. patients without CNS lesions

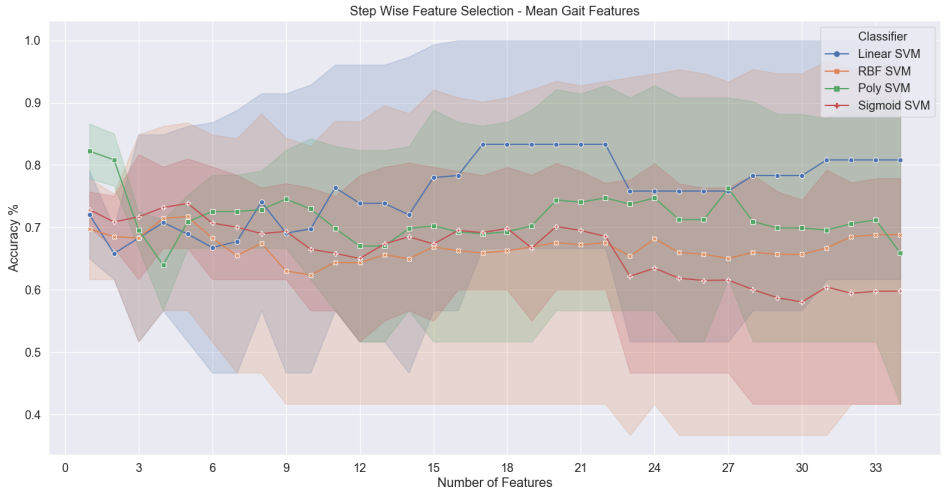
The following Figure 21 shows the feature selection results for the task of distinguishing between FD patients without CNS lesions and FD patients with CNS lesions based on the Raw Mean Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 21: Raw Mean Strides dataset feature selection with lasso, gini and backward step-wise for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task.

The following Table 66 summarizes the best results achieved by the feature selection methods for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task. The results of the backward step-wise feature selection method show that the number of features that achieved the best results is 1, however, since the results for 1 and 2 features are very similar the final choice of feature subset is 2. This choice is based on the fact that when performing fine-tuning the performance of the classifiers will improve the most with 2 features.

Table 66: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for the FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task based on Raw Mean Strides dataset.

	Kernel	**Training Accuracy**	**Validation Accuracy**	**Mean Accuracy**	**Number Features**
Lasso	Polynomial	87.52%	85.0%	86.26%	**3**
Gini	Linear	87.58%	86.67%	87.13%	**3**
Step-Wise	Polynomial	85.04%	76.67%	80.85%	**2**

The 3 features selected by the lasso method were **peak swing**, **lift-off angle** and **CV of maximum toe clearance 2**.

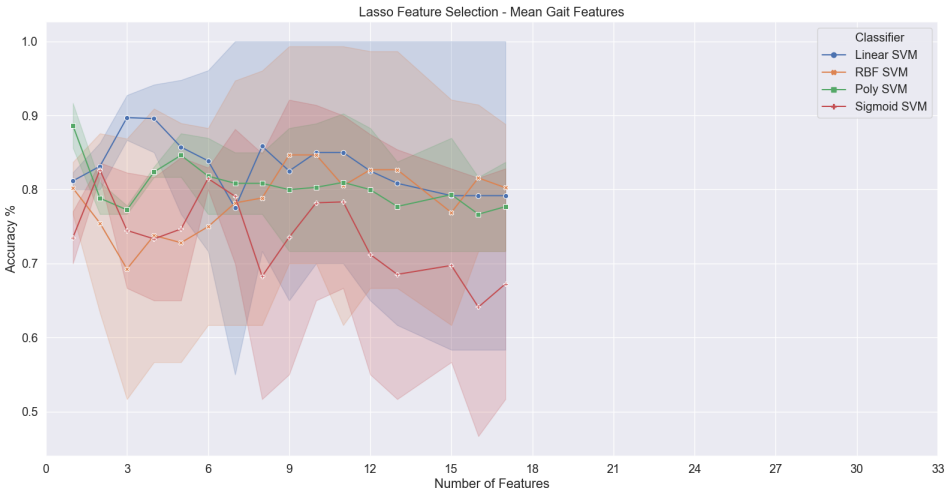
The 3 features selected by the gini method were **stride length**, **CV of strike angle** and **CV of maximum toe clearance 2**.

The 2 features selected by the backward step-wise method were **CV of maximum toe clearance 2**, **CV of strike angle**.

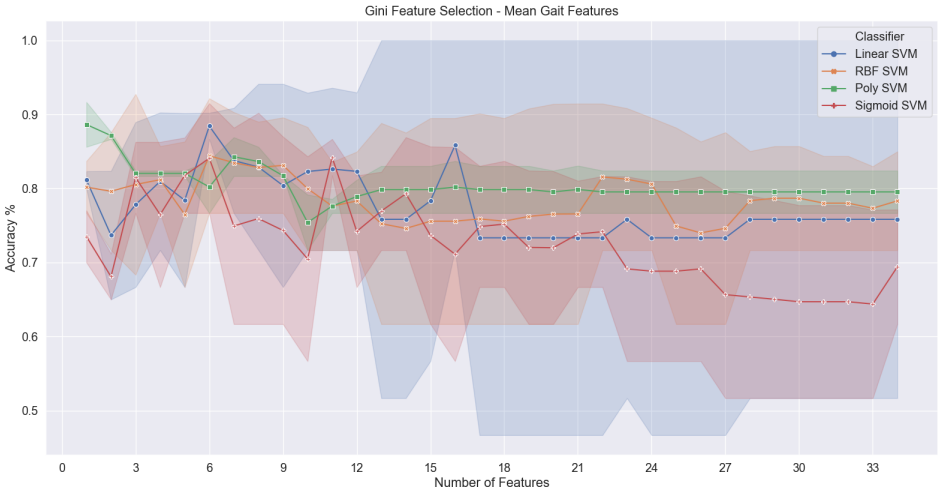
8.1.4 Feature selection - Dataset Normalized Mean Strides

Patients with CNS lesions vs. patients without CNS lesions

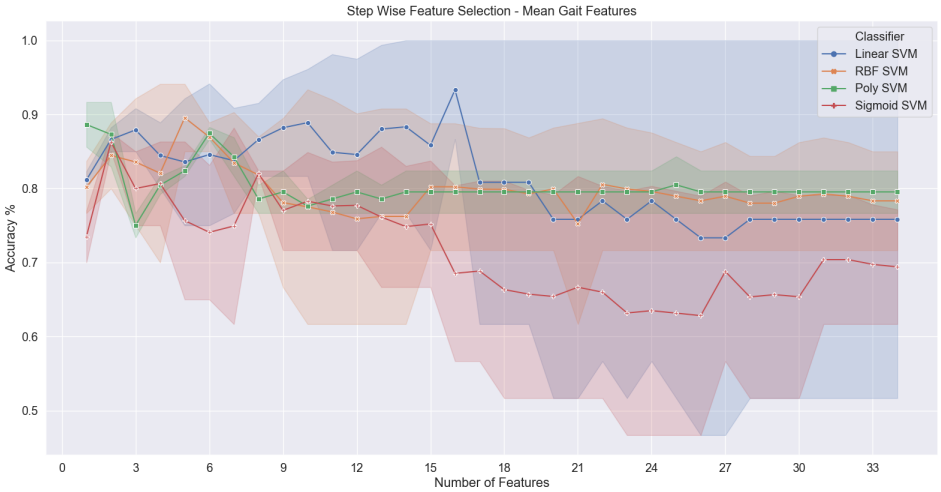
The following Figure 22 shows the feature selection results for the task of distinguishing between FD patients without CNS lesions and FD patients with CNS lesions based on the Normalized Mean Strides dataset.



(a) Lasso feature selection.



(b) Gini feature selection.



(c) Backward step-wise feature selection.

Figure 22: Normalized Mean Strides dataset feature selection with lasso, gini and backward step-wise for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task.

The following Table 67 summarizes the best results achieved by the feature selection methods for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task.

Table 67: Best results achieved by the different feature selection methods based on the performance of the four SVM classifiers for FD patients without CNS lesions vs. FD patients with CNS lesions differentiation task based on the Normalized Mean Strides dataset.

	Kernel	Training Accuracy	Validation Accuracy	Mean Accuracy	Number Features
Lasso	Linear	92.74%	86.66%	89.70%	**3**
Gini	Linear	90.20%	86.67%	88.44%	**6**
Step-Wise	RBF	94.12%	85.0%	89.56%	**5**

The 3 features selected by the lasso method were **minimum toe clearance**, **CV of minimum toe clearance** and **CV of maximum toe clearance 2**.

The 6 features selected by the gini method were **CV of maximum toe clearance 2**, **CV of swing**, **CV of maximum heel clearance**, **CV of maximum toe clearance 1**, **speed** and **CV of strike angle**.

The 5 features selected by the backward step-wise method were **CV of strike angle**, **CV of swing**, **CV of maximum toe clearance 2**, **speed** and **CV of maximum toe clearance 1**.

8.2 Performance evaluation for the SVM classifiers

Different hyperparameters were exhaustively combined which resulted in the development of 1769 SVM classifiers for each classification task and feature subset. This work comprises four different datasets, each dataset is used to perform one classification task using the three best feature subsets previously selected. In total, the number of developed SVM classifiers is 21228.

The search of hyperparameters included 29 different c values, 12 different γ values, 4 different kernel functions, 3 different polynomial degrees, and δ was set to 0. All the combinations between these hyperparameters were developed and evaluated.

From all the developed SVM classifiers the one with the best validation performance on each feature subset was selected and used to make predictions on the test set. Finally, the best overall SVM classifier for each dataset, based on the validation and test performances, was chosen. The results are summarized in the following sub-sections.

8.2.1 Performance with Raw All Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 68: Performance and hyperparameters of the best SVM classifier developed with the 5 Raw All Strides features selected by the lasso method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	83.19%	99.48%	80.37%
Validation Set	79.15%	70.0%	79.01%
Test Set	96.06%	100%	94.79%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	1.5	0.1	4

8.2.2 Performance with Normalized All Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 69: Performance and hyperparameters of the best SVM classifier developed with the 5 Normalized All Strides features selected by the gini method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Gini Method - 5 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	95.99%	99.66%	94.57%
Validation Set	86.27%	89.74%	85.85%
Test Set	83.46%	94.12%	81.82%
Test Set Individual Performance	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	C	Gamma	Degree
Polynomial	1.5	0.2	5

8.2.3 Performance with Raw Mean Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 70: Performance and hyperparameters of the best SVM classifier developed with the 3 Raw Mean Strides features selected by the lasso method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	87.57%	100.0%	83.94%
Validation Set	90.0%	100.0%	90.0%
Test Set	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	C	Gamma	Degree
Polynomial	0.5	0.17	3

8.2.4 Performance with Normalized Mean Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 71: Performance and hyperparameters of the best SVM classifier developed with the 6 Normalized Mean Strides features selected by the gini method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Gini Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.49%	86.71%	95.18%
Validation Set	86.67%	95.0%	91.67%
Test Set	66.67%	50.0%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	10	0.01	––

8.3 Performance evaluation for the DBN classifiers

Different hyperparameters were combined which resulted in the development of 1500 DBN classifiers for each Mean Strides dataset and feature subset. This work comprises two datasets, each dataset is used to perform one classification task based on the three best feature subsets previously selected. In total, the number of developed DBN classifiers is 9000.

The search of hyperparameters included 13 different values for the first hidden layer of neurons, 15 different values for the second hidden layer of neurons, 10 different values for the learning rate of the unsupervised training stage (Stage 1), 3 different values for the learning rate of the fine-tuning stage (Stage 2), 4 different values for the dropout rate, 3 different values for the batch size, 6 different values for the epochs of the first learning stage, and 6 different values for the epochs of the fine-tuning stage. From all the possible 2 525 200 combinations between these hyperparameters, 1500 were randomly chosen to develop the DBN classifiers.

From all the developed DBN classifiers the one with the best validation performance on each feature subset was selected and used to make predictions on the test set. Finally, the best overall DBN classifier for the Raw and Normalized Mean Strides datasets, based on the validation and test performances, was chosen. The results are summarized in the following sub-sections.

8.3.1 Performance with Raw Mean Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 72: Performance and hyperparameters of the best DBN classifier developed with the 3 Raw Mean Strides features selected by the lasso method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	66.76%	100.0%	66.43%
Validation Set	76.67%	100.0%	76.67%
Test Set	66.67%	100%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.0015	0.1	24/16	2	10	120

8.3.2 Performance with Normalized Mean Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 73: Performance and hyperparameters of the best DBN classifier developed with the 3 Normalized Mean Strides features selected by the lasso method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	68.24%	98.33%	68.75%
Validation Set	75.0%	100%	75.0%
Test Set	66.67%	100%	66.67%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.007	0.0015	0.1	18/16	2	8	100

8.4 Performance evaluation for the LSTM classifiers

Different hyperparameters were combined which resulted in the development of 1500 LSTM classifiers for each All Strides dataset and feature subset. This work comprises two different datasets, each dataset is used to perform one classification task based on the three best feature subsets previously selected. In total, the number of developed LSTM classifiers is 9000.

The search of hyperparameters included 15 different values for the first LSTM hidden layer of neurons, 17 different values for the second LSTM hidden layer of neurons, 15 different values for the first MLP hidden layer of neurons, 17 different values for the second MLP hidden layer of neurons, 4 different values for the learning rate, 4 different values for the dropout rate, 3 different values for the batch size, and 6 different values for the epochs. From all the possible 18 727 200 combinations between these hyperparameters, 1500 were randomly chosen to develop the LSTM classifiers.

From all the developed LSTM classifiers the one with the best validation performance on each feature

subset was selected and used to make predictions on the test set. Finally, the best overall LSTM classifier for the Raw and Normalized All Strides datasets, based on the validation and test performances, was chosen. The results are summarized in the following sub-sections.

8.4.1 Performance with Raw All Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 74: Performance of the best LSTM classifier developed with the 2 Raw All Strides features selected by the backward step-wise method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Step-Wise Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	68.60%	100%	67.44%
Validation Set	71.67%	100%	71.67%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	22/14	22/16	4	30

8.4.2 Performance with Normalized All Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 75: Performance of the best LSTM classifier developed with the 6 Normalized All Strides features selected by the backward step-wise method for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Step-Wise Method - 6 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	81.07%	100%	77.50%
Validation Set	86.67%	100%	86.67%
Test Set	66.67%	100%	66.67%

LSTM Classifier Hyperparameters

LR	DR	LSTM Neurons	MLP Neurons	Batch Size	Epochs
0.001	0.2	22/14	22/16	4	30

8.5 Performance evaluation for the CNN/MLP classifiers based on gait and biometric data

The CNN/MLP classifier based on the gait and biometric data of each FD patient is constructed based on a CNN model that processes the gait data and an MLP model that processes the biometric data (age, weight, height, and gender) of each patient. The features of the last layer of the CNN model and the features of the last layer of the MLP model that processes the biometric data are concatenated. The concatenated features are then fed to the final feed-forward network (MLP) whose output will be the gait pattern prediction of each patient. The following Figure 23 summarizes the architecture of the developed classifier.

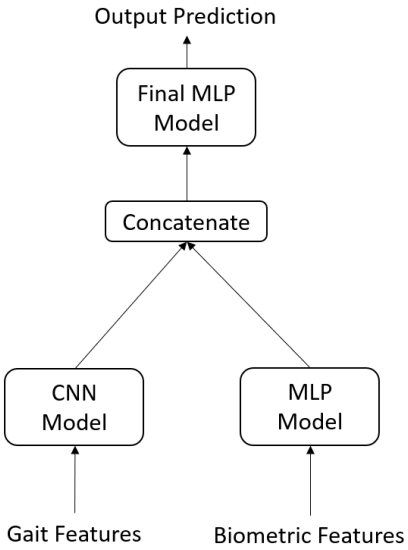


Figure 23: Architecture of the CNN classifier based on gait and biometric data for the classification task of differentiating between FD without CNS lesions and FD with CNS lesions gait patterns.

Different hyperparameters were combined which resulted in the development of 1500 CNN/MLP classifiers for each feature subset of the FD without CNS lesions vs. FD with CNS lesions differentiation task. This work comprises one dataset to perform one classification task based on the three best feature subsets previously selected. In total, the number of developed CNN classifiers is 4500.

The search of parameters included 15 different values for the first CNN hidden layer of neurons, 17 different values for the second CNN hidden layer of neurons, 3 different values for the first hidden layer of the MLP Biometric model, 3 different values for the second hidden layer of the MLP Biometric model, 15 different values for the first MLP hidden layer of neurons, 17 different values for the second MLP hidden layer of neurons, 3 different values for the kernel value (length of the convolution window), 4 different values for the learning rate, 4 different values for the dropout rate, 3 different values for the batch size, and 6 different values for the epochs. From all the possible 505 634 400 combinations between these hyperparameters, 1500 were randomly chosen to develop the CNN classifiers.

From all the developed CNN/MLP classifiers the one with the best validation performance on each feature subset was selected and used to make predictions on the test set. Finally, the best overall CNN/MLP classifier, based on the validation and test performances, was chosen. The results are summarized in the following sub-section.

8.5.1 Performance with Raw All Strides Dataset

Patients with CNS lesions vs. patients without CNS lesions

Table 76: Performance of the best CNN/MLP classifier developed with the 2 Raw All Strides features selected by the backward step-wise method and biometric data of each patient for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

Step-Wise Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	88.15%	87.40%	90.36%
Validation Set	86.67%	100%	86.67%
Test Set	100%	100%	100%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.002	0.3	32/16	5	16/8	32/16	4	24

8.6 Comparison of performance among the different classifiers

In this section a comparison between the overall performance of the best classifiers for the differentiation task of FD patients without CNS lesions vs. FD patients with CNS lesions is displayed in the following tables.

The following Table 77 summarizes the best results for the classification task based on the Raw and Normalized All Strides datasets features.

Table 77: Validation and test performance comparison between the best classifiers when trained with All Strides datasets for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

	Raw All Strides			Normalized All Strides		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM Classifier	79.15% 96.06%	70.0% 100%	79.01% 94.79%	**86.27%** **83.46%**	**89.74%** **94.12%**	**85.85%** **81.82%**
LSTM Classifier	71.67% 100%	100% 100%	71.67% 100%	86.67% 66.67%	100% 100%	86.67% 66.67%
CNN Classifier Biometric + gait data	**86.67%** **100%**	**100%** **100%**	**86.67%** **100%**	— —	— —	— —

For each classifier, validation performance is displayed on the first row, and test performance is displayed on the second row.

The following Table 78 summarizes the best results for the classification task based on the Raw and Normalized Mean Strides datasets features.

Table 78: Validation and test performance comparison between the best classifiers when trained with Mean Strides datasets for the FD without CNS lesions vs. FD with CNS lesions differentiation task.

	Raw Mean Strides			Normalized Mean Strides		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM Classifier	**90.0%** **66.67%**	**100%** **100%**	**90.0%** **66.67%**	**86.67%** **66.67%**	**95.0%** **66.67%**	**91.67%** **100%**
DBN Classifier	76.67% 66.67%	100% 100%	76.67% 66.67%	75.0% 66.67%	100% 100%	75.0% 66.67%

For each classifier, validation performance is displayed on the first row, and test performance is displayed on the second row.

8.7 Clustering of Fabry Disease patients with/without CNS lesions based on cardiac data

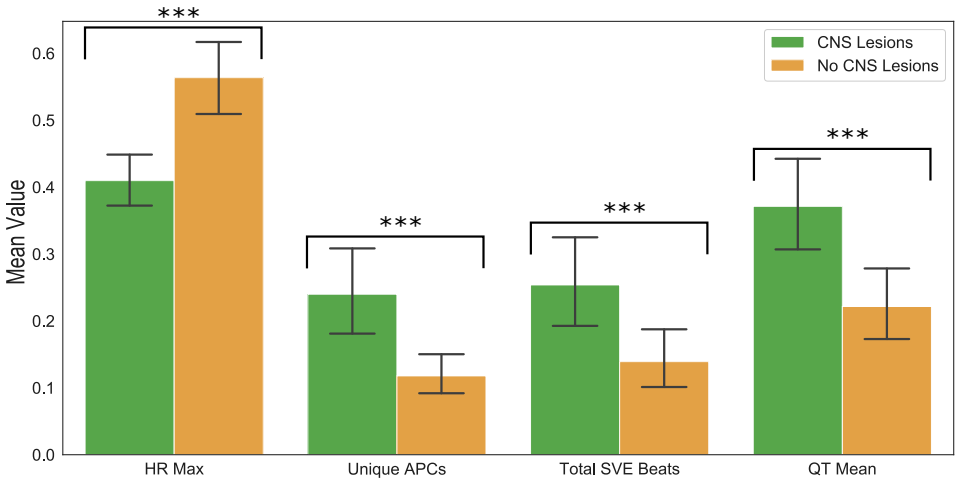
K-means clustering and hierarchical clustering were performed to identify subgroups of FD patients based on cardiac characteristics. The cardiac and neurological characteristics of the found subgroups were analysed. Two cardiac datasets were used for the clustering analysis, these were Holter cardiac dataset and Echocardiogram dataset.

The clustering approach was the following: (1) acquire information about the features composing each of the used datasets, Mann Whitney U-test was used to assess the significance of each feature when comparing between FD without CNS lesions and FD with CNS lesions, features with a significance level superior to 0.05 were discarded, (2) further reduce the data dimensionality using PCA (feature scaling was performed before PCA), the PCs were selected based on a threshold of around 95% for the total explained variance, (3) assess the performance of the two clustering methods (k-means and hierarchical clustering) based on the Holter and Echo PCs with the number of clusters pre-defined to two, (4) select the best clustering method based on the clustering classification performance, (5) perform clustering analysis with the selected method, the number of optimal clusters is identified, depending on the selected clustering method, using the elbow method or the dendrogram, and finally, (6) interpret the clustering results.

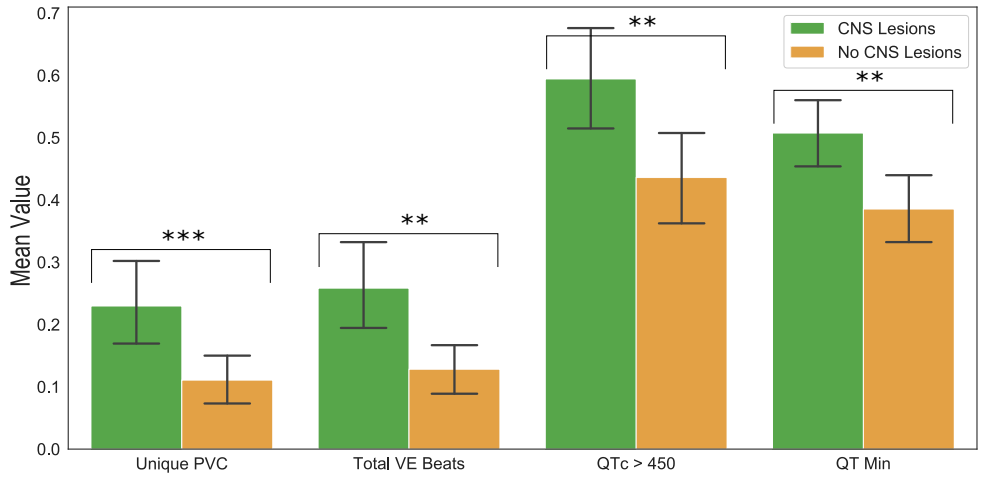
8.7.1 Clustering using Holter cardiac data

The Table containing all the Mann Whitney U-test p-values for the different Holter features when comparing FD patients without CNS lesions and FD patients with CNS lesions can be found in the Appendix.

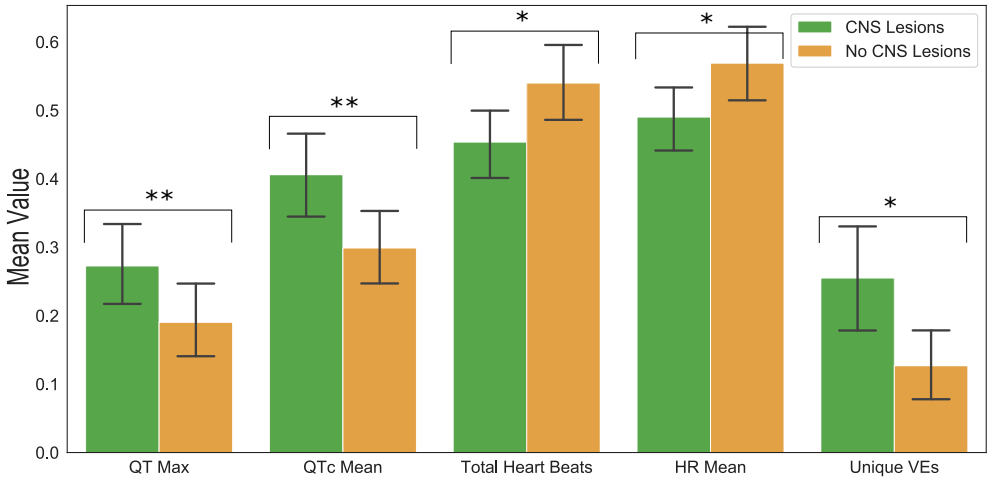
A p-value of 0.05 or less was considered significant which resulted in 13 selected features. The following Figure 24 represents the differences between FD patients without CNS lesions and FD patients with CNS lesions for each of the selected Holter features.



(a) Bar Plot of HR Max, Unique APCs, Total SVE Beats and QT Mean.



(b) Bar Plot of Unique PVC, Total VE Beats, QTc $>$ 450 and QT Min.



(c) Bar Plot of QT Max, QTc Mean, Total Heart Beats, HR Mean and Unique VEs.

Figure 24: Comparison between the mean value of Holter cardiac features of FD patients without CNS lesions and FD patients with CNS lesions. Significant differences in gait features between FD patients and controls are indicated with one asterisk (*$p < 0.05$), two asterisk (**$p < 0.01$) and three asterisk (***$p < 0.001$). Whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot.

A pair plot of the four most significant Holter features when comparing between FD patients without CNS lesions and FD patients with CNS lesions according to the Mann Whitney U-test is represented in Figure 25.

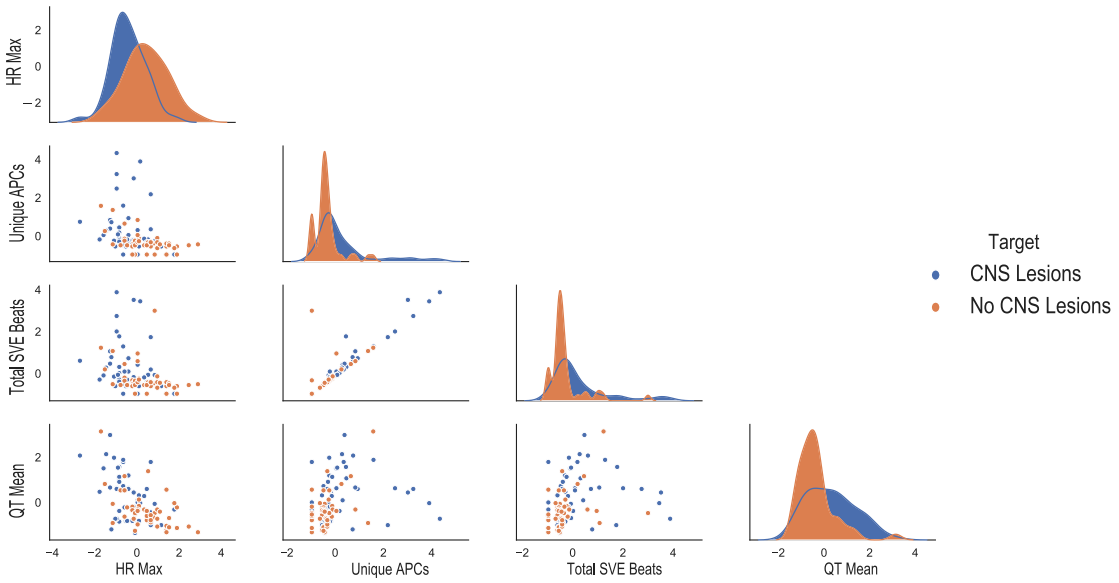


Figure 25: Pair Plot of HR Max, Unique APCs, Total SVE Beats and QT Mean. All data are standardized and dimensionless.

Principal components analysis was applied to the cardiac Holter features, the resulting principal components are summarized in the following Figure 26.

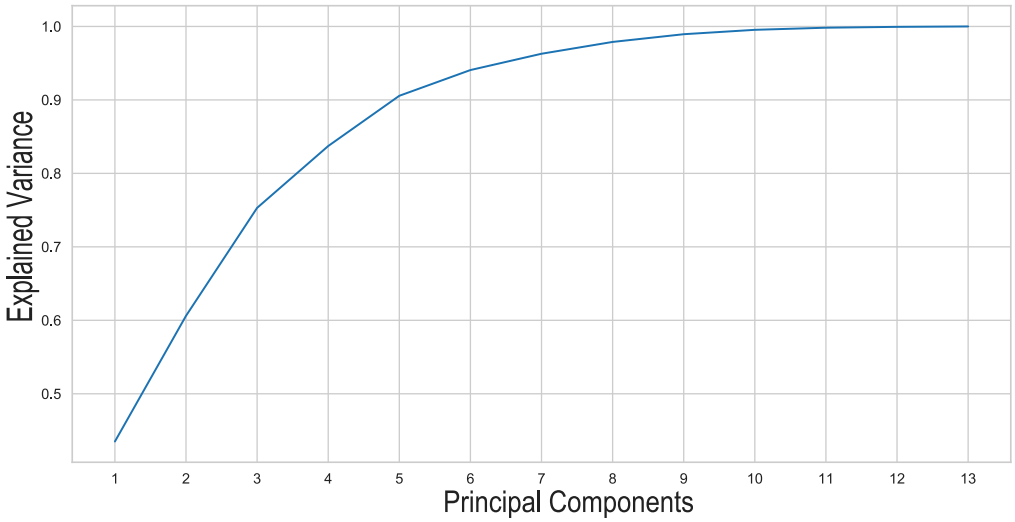


Figure 26: Principal components and the respective cumulative sum of the variance.

The first seven PCs account for 96.28% of the total variance. These were used to perform the clustering analysis.

To choose which method would be used to perform clustering analysis the classification performance of both K-means and hierarchical clustering was evaluated. These methods were implemented with the number of clusters pre-defined to two, the results are summarized in the following Table 79.

Table 79: Clustering performance for the FD without CNS lesions vs. FD with CNS lesions classification task based on the seven Holter PCs with the number of clusters pre-defined to two.

	K-Means Clustering			Hierarchical Clustering		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
7 PCs	65.26 %	59.68 %	75.76%	66.32 %	61.02 %	75.0%

From the analysis of Table 79 it can be concluded that the hierarchical clustering was the superior method. Based on these results hierarchical clustering was used to perform the clustering analysis. The following Figure 27 shows the dendrogram for the hierarchical clustering based on the seven Holter PCs.

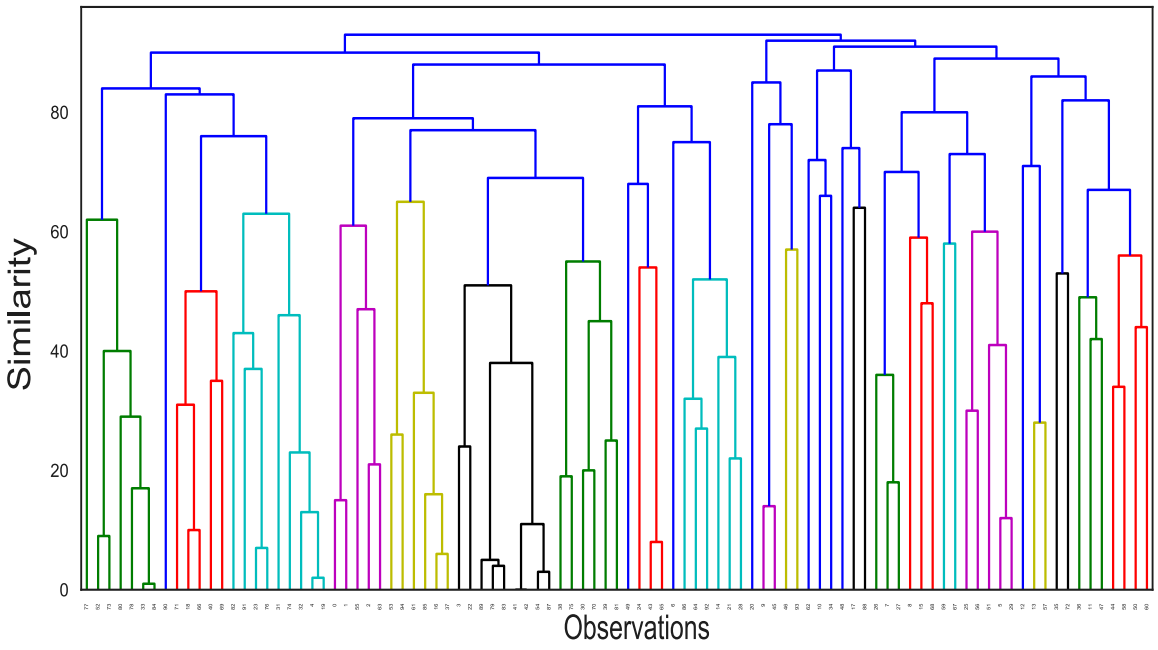


Figure 27: Dendrogram obtained with the seven Holter PCs.

With the assistance of the dendrogram results (see Figure 27), the selected optimal number of clusters was five.

The following Table 80 shows the results of hierarchical clustering based on the seven Holter PCs with the number of clusters set to five.

To better understand the following results, it is important to briefly discuss the Holter features used to conduct the clustering analysis.

The QT interval metrics provide information about the cardiac conduction system, the QT interval is the time from ventricular depolarization to ventricular re-polarization. A long QT interval means that the heart is taking longer than normal to recharge between beats, this is called the long QT syndrome (LQTS) which can be lethal (Johnson and Ackerman (2009)). The QT interval is usually analysed using the QTc Mean metric, this metric should be between 350 and 440 if the QTc Mean is above 500 the risk of sudden cardiac death is high. A QTc Mean lower than 350 is very rare and it is denoted as Short QT syndrome (SQTS), this syndrome increases the risk of atrial and ventricular arrhythmias (Rudic et al. (2014)).

The metrics regarding the ventricular and supraventricular ectopic events should be as low as possible, various studies show that individuals with a high number of ectopic events are at a greater risk of stroke and sudden death (Simpson et al. (2017) Nguyen and Thomas (2010)). Finally, the heart rate metrics also provide information about cardiac condition, it is difficult to define an optimal heart rate range, the normal range is between 60 and 90 or 100, above this range, it might be an indication of different cardiac pathologies like ventricular arrhythmias (Fox et al. (2007)).

Table 80: Mean of the Holter cardiac features for each cluster found with hierarchical clustering method based on the seven Holter PCs.

Feature	**Cluster 1** (8/17)	**Cluster 2** (23/14)	**Cluster 3** (0/5)	**Cluster 4** (1/5)	**Cluster 5** (13/9)
HR Max	119.16	132.08	118.2	102.33	138.0
Unique APCs	52.08	46.16	2345.6	119.33	7.14
Total SVE Beats	97.92	123.27	3078.8	140.17	21.36
QT Mean	420.68	377.14	400.8	467.33	374.64
Unique PVC	160.48	3.54	462.6	194.17	1.5

Total VE Beats	218.16	5.27	521.4	274.33	2.59
QTc $>$ 450	52.88	4.14	49.4	59.83	24.32
QT Min	334.76	291.14	304.0	285.33	286.09
QT Max	497.28	451.14	495.4	674.0	448.64
QTc Mean	452.88	416.57	445.4	478.33	441.23
Total Heart Beats	98339.0	101817.68	102837.0	87672.0	116117.27
HR Mean	71.88	74.81	75.6	63.83	84.73
Unique VEs	7.64	1.1351	24.2	13.33	0.27

The following conclusions can be drawn from the analysis of Table 80, the first cluster is composed of 8 FD patients without CNS lesions and 17 FD patients with CNS lesions. This cluster is characterized by a slightly high number of abnormal cardiac events (ectopic events), a QTc Mean slightly higher than the normal range of 350 to 440, and a high QTc $>$ 450 (QT interval of these patients is usually above 450). The HR Mean is in the normal range, comparing to the other clusters it is at an intermediate level. This cluster seems to represent mainly patients with minor cardiac complications.

The second cluster is composed of 23 FD patients without CNS lesions and 14 FD patients with CNS lesions. This cluster is characterized by a low number of abnormal cardiac events, the second-highest HR Max and very good QT intervals. These characteristics seem to describe FD patients without cardiac complications.

The third cluster is composed of 5 FD patients with CNS lesions. This cluster represents patients with a very high number of abnormal cardiac events, in fact, the highest number of ectopic beats among all clusters. The QTc Mean is slightly over the normal range and the QTc $>$ 450 is also high around 50%. This cluster seems to represent FD patients with cardiac complications that have origin at the atrium and ventricles.

The fourth cluster is composed of 1 FD patient without CNS lesions and 5 FD patients with CNS lesions. This cluster is characterized by a low HR Max (the lowest among all clusters), a QTc Mean above the normal range (the highest QTc Mean among all the clusters) a very high QT Max (also the highest among all the clusters), the highest QTc $>$ 450 among all cluster and the lowest number of Total Heart

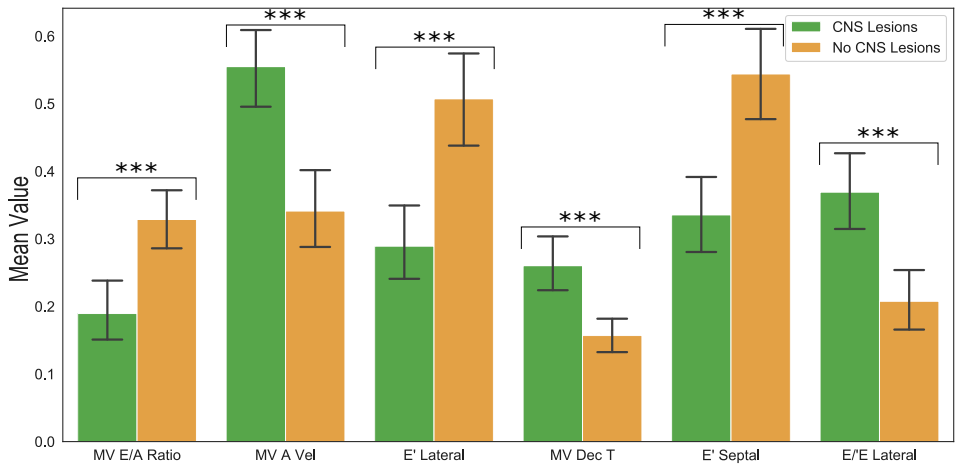
Beats among all the clusters. This cluster seems to represent patients with abnormal QT intervals which indicates serious risk of severe cardiac events.

The fifth cluster is composed of 13 FD patients without CNS lesions and 9 FD patients with CNS lesions. This cluster is characterized by the lowest number of abnormal cardiac events, a QTc Mean very slightly over the normal range and a QTc $>$ 450 slightly high. This cluster seems to represent patients with very minor cardiac complications.

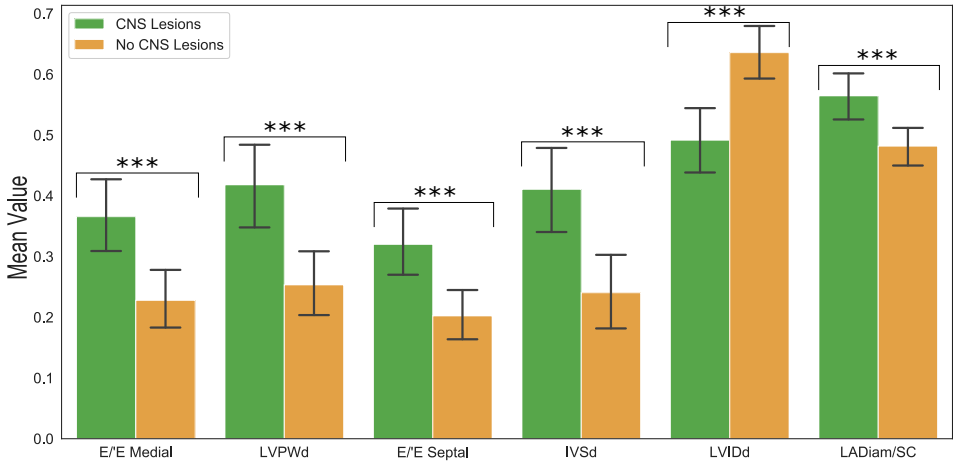
8.7.2 Clustering using Echocardiogram data

The Table containing all the Mann Whitney U-test p-values for the different Echocardiogram features when comparing FD patients without CNS lesions and FD patients with CNS lesions can be found in the Appendix.

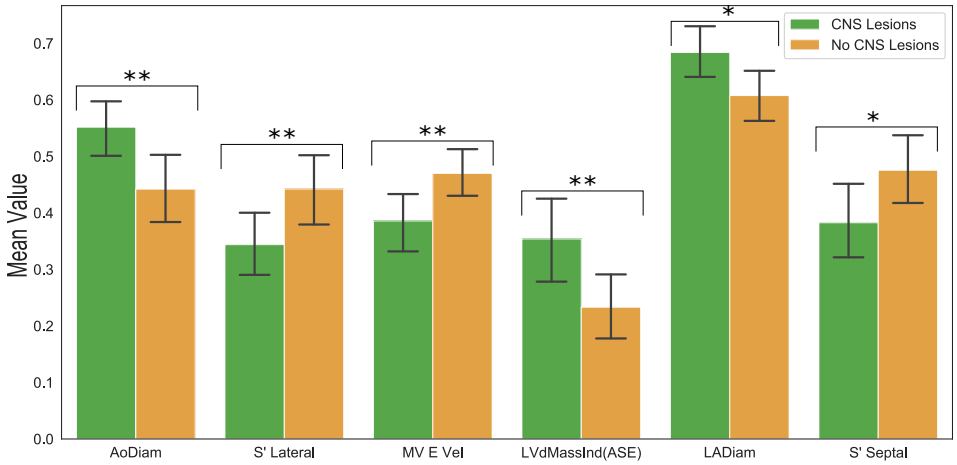
A p-value of 0.05 or less was considered significant which resulted in 21 selected features. The following Figure 28 represents the differences between FD patients without CNS lesions and FD patients with CNS lesions for each of the selected Echocardiogram features.



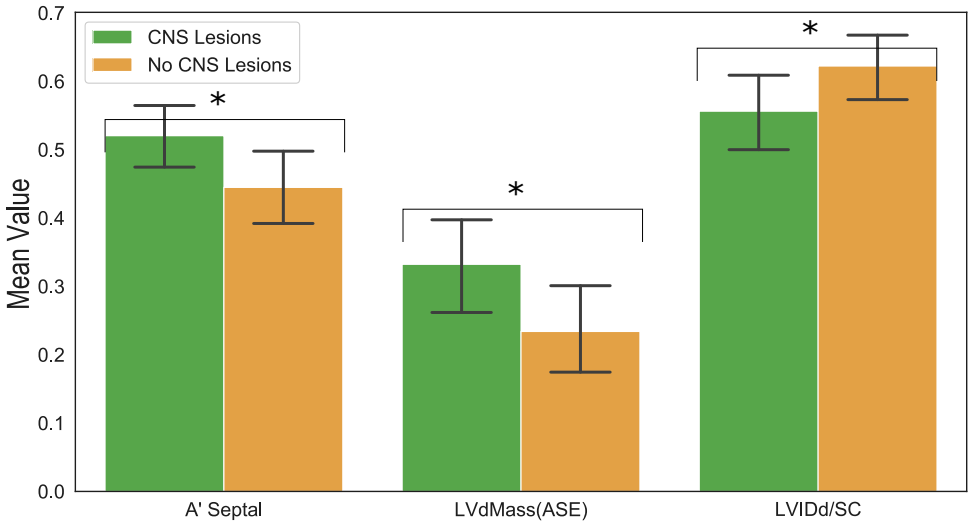
(a) Bar Plot of MV E/A Ratio, MV A Vel, E' Lateral MV Dec T, E' Septal and E/E' Lateral.



(b) Bar Plot of E/E' Medial, LVPWd, E/E' Septal, IVSd, LVIDd and LADiam/SC



(c) Bar Plot of AoDiam, S' Lateral, MV E Vel, LVdMassInd(ASE), LADiam and S' Septal.



(d) Bar Plot of A' Septal, LVdMass(ASE) and LVIDd/SC.

Figure 28: Comparison between the mean value of Echocardiogram features of FD patients without CNS lesions and FD patients with CNS lesions. Significant differences in gait features between FD patients and controls are indicated with one asterisk (*$p < 0.05$), two asterisk (**$p < 0.01$) and three asterisk (***$p < 0.001$). Whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot.

A pair plot of the four most significant Echocardiogram features when comparing between FD patients without CNS lesions and FD patients with CNS lesions according to the Mann Whitney U-test is represented in Figure 29.

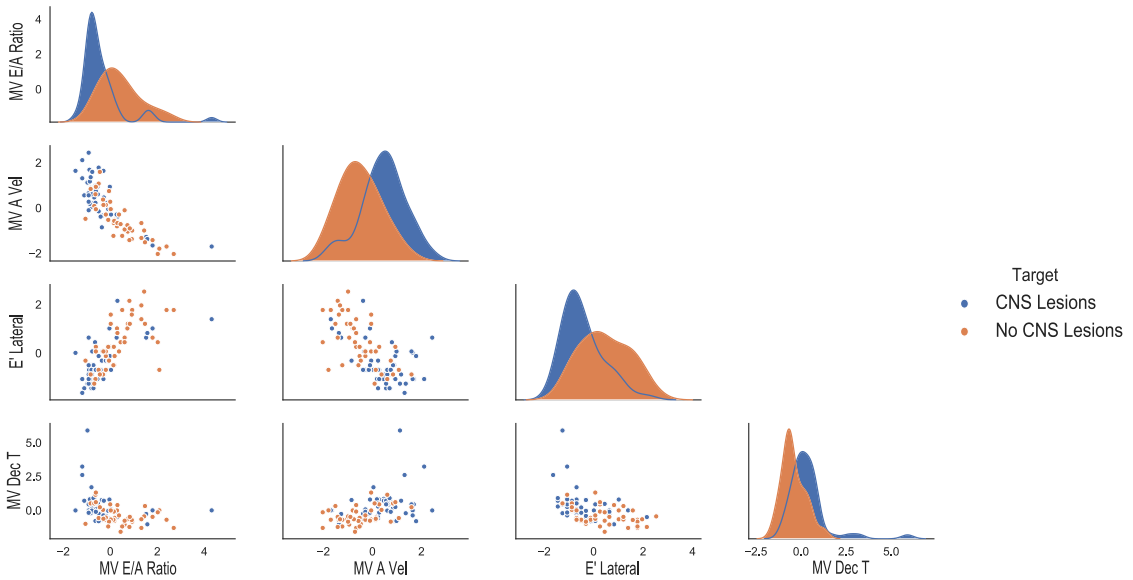


Figure 29: Pair Plot of MV E/A Ratio, MV A Vel, E' Lateral and MV Dec T. All data are standardized and dimensionless.

Principal components analysis was applied to the Echocardiogram features, the resulting PCs are summarized in the following Figure 30.

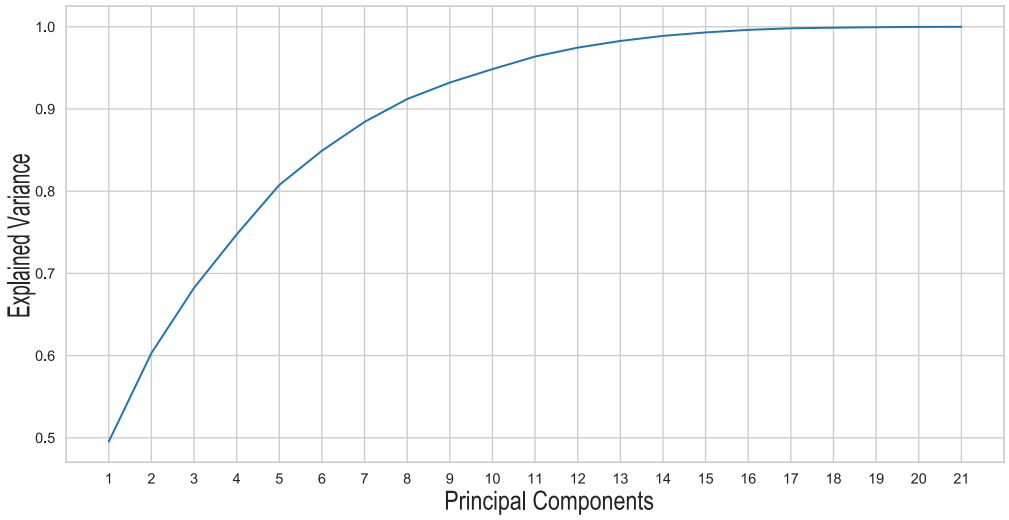


Figure 30: Principal components and the respective cumulative sum of the variance.

The first eleven principal components account for 96.38% of the total variance. These were used to

perform the clustering analysis.

To choose which method would be used to perform clustering analysis the classification performance of both K-means and hierarchical clustering was evaluated. These methods were implemented with the number of clusters pre-defined to two, the results are summarized in the following Table 79.

Table 81: Clustering performance for the FD without CNS lesions vs. FD with CNS lesions classification task based on the eleven Echocardiogram PCs with the number of clusters pre-defined to two.

	K-means Clustering			Hierarchical Clustering		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
11 PCs	68.82 %	63.64 %	76.32%	68.82 %	63.64 %	76.32%

From the analysis of Table 81 it can be concluded that both clustering methods achieve the same results. The clustering analysis was conducted using the K-means method (arbitrary choice). The following Figure 31 shows the elbow method based on the eleven Echocardiogram PCs.

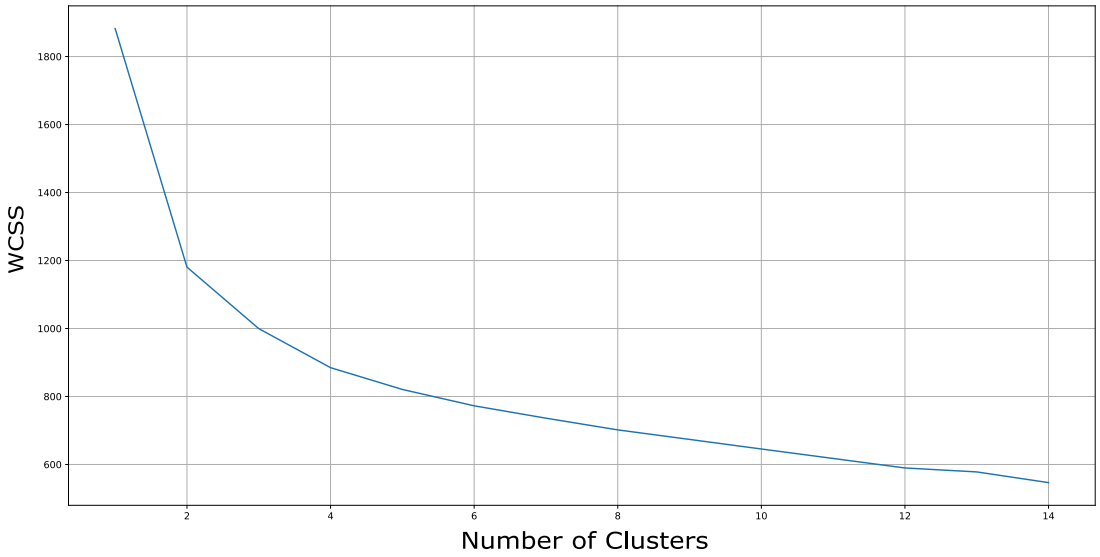


Figure 31: Elbow method obtained with the eleven Echocardiogram PCs.

With the assistance of the elbow method results (see Figure 31), the selected optimal number of clusters was four.

To better understand the upcoming clustering results it is important to briefly discuss the Echocardiogram features used to implement the clustering analysis, the following Table 82 summarizes according

to Caballero et al. (2015) and El Missiri et al. (2016) the normal ranges of the Echocardiogram features used to conduct the cluster analysis.

Table 82: Reference values for Echocardiogram measurements in healthy adults.

Feature	**Normal Range**
MV E/A Ratio	0.86 to 1.88
MV E Vel (cm/s)	0.61 to 0.93
MV A Vel (cm/s)	0.43 to 0.77 (must be smaller than MV E Vel)
MV Dec T (ms)	138.6 to 237.4
E' Lateral (cm/s)	9.5 to 17.5
E' Septal (cm/s)	7.3 to 13.3
E/E' Lateral	4.0 to 8.2
E/E' Medial	4.6 to 8.6
E/E' Septal	5.5 a 10.3
S' Septal (cm/s)	6.7 to 9.5
A' Septal (cm/s)	7.4 to 11.4
S' Lateral (cm/s)	7.4 to 12.2
LVPWd (mm)	7.65 to 10.07
IVSd (mm)	7.78 to 10.06
LVIDd (mm)	43.56 to 52.16
LADiam/SC (mm/m^2)	Not Defined
AoDiam (mm)	23.22 to 27.88
LVdMassInd(ASE) (g/m^2)	43 to 115
LADiam (mm)	23.74 to 30.44
LVdMass(ASE) (g)	95.78 to 192.81
LVIDd/SC (mm/m^2)	Not Defined

The following Table 83 summarizes the mean values of the Echocardiogram features for each cluster found by the K-means clustering based on the eleven Echocardiogram PCs with the number of clusters

set to four.

Table 83: Mean of the Echocardiogram features for each cluster found with K-means clustering method based on the eleven Echocardiogram PCs.

Feature	**Cluster 1** (16/14)	**Cluster 2** (5/19)	**Cluster 3** (20/6)	**Cluster 4** (3/10)
MV E/A Ratio	1.07	0.81	1.81	0.82
MV E Vel	0.83	0.67	0.87	0.69
MV A Vel	0.77	0.83	0.48	0.84
MV Dec T	217.46	240.28	191.13	281.92
E' Lateral	11.43	7.57	18.54	5.77
E' Septal	9.0	6.46	13.69	3.92
E/E' Lateral	7.48	9.35	4.76	13.0
E/E' Medial	8.56	10.15	6.02	15.85
E/E' Septal	9.64	10.73	6.5	18.69
S' Septal	8.11	6.5	8.46	5.38
A' Septal	10.1	9.42	7.49	7.38
S' Lateral	9.50	7.34	11.62	6.23
LVPWd	8.56	12.01	7.80	15.27
IVSd	8.99	13.95	8.04	17.92
LVIDd	43.69	44.99	46.02	41.09
LADiam/SC	19.41	21.98	18.79	24.9
AoDiam	30.79	34.23	27.60	34.81
LVdMassInd(ASE)	70.70	124.0	69.37	170.39
LADiam	33.65	39.29	31.96	41.68
LVdMass(ASE)	123.89	221.88	118.91	288.48
LVIDd/SC	25.16	25.21	26.86	24.53

The following conclusions can be drawn from the analysis of Table 83, the first cluster is composed of 16 FD patients without CNS lesions and 14 FD patients with CNS lesions. This cluster is characterized

by normal diastolic and systolic functions, the cardiac measurements are inside normal ranges except the AoDiam and LADiam that are slightly higher than normal. This cluster seems to represent patients with normal cardiac measurements and normal diastolic and systolic functions.

The second cluster is composed of 5 FD patients without CNS lesions and 19 FD patients with CNS lesions. The cluster is represented by an abnormal diastolic and systolic functions, the values of the E, A and S waves are outside the normal range, the A Wave is greater than the E wave and the deceleration time is slightly high which indicates grade 1 diastolic dysfunction. The majority of cardiac measurements also exceed the normal ranges. This cluster seems to describe cardiac features of patients with significant cardiac manifestations.

The third cluster is composed of 20 FD patients without CNS lesions and 6 FD patients with CNS lesions. This cluster represents normal diastolic and systolic functions along side normal cardiac measurements. The only measurement outside the normal range is the LADiam, only exceeding the range slightly. Except for the LADiam measure, this cluster represents normal cardiac measurements indicating good cardiac function.

The fourth cluster is composed of 3 FD patients without CNS lesions and 10 FD patients with CNS lesions. This cluster is characterized by diastolic and systolic abnormal functions, the A wave is greater than the E wave, the diastolic and systolic parameters are almost all out of the normal range indicating poor cardiac function. The cardiac measurements exceed the normal ranges by a large margin which also indicates serious cardiac manifestations. This cluster represents the worst cardiac behaviour among all clusters.

8.8 Fabry Disease Results Discussion

8.8.1 General discussion

The most frequent FD symptoms occur at a neurological level (Giugliani et al. (2016)) and recently there has been a growing hypothesis that gait is the final outcome of several CNS functions (Amboni

et al. (2013)), these findings led to the hypothesis analysed in this work that the CNS lesions from which FD patients suffer might be detectable in their gait patterns. To validate this hypothesis gait data were collected from 20 FD patients, the gait analysis process of FD patients followed a similar approach to the one conducted for the Parkinsonism patients' gait analysis. The main goal was to perform classification of FD patients without CNS lesions vs. FD patients with CNS lesions based on gait features.

All FD gait features were normalized according to the MR approach presented by Wahid et al. (Wahid et al. (2015)), after normalization four datasets including Raw All Strides, Normalized All Strides, Raw Mean Strides, and Normalized Mean Strides were available to tackle the classification task.

For each dataset, the best subset of features was selected using the gini, lasso, and backward step-wise feature selection methods alongside four SVM classifiers whose 10-fold stratified cross-validation performance served as a guideline (approach detailed in section 8.1). The feature selection achieved great results, the performance of the four SVM classifiers always increased when based on a subset of the original features. The SVM kernel that seems to achieve the best performance was the polynomial kernel (third degree), the feature selection method that finds the best subset of features based on the SVMs performances was the gini (mean accuracy: 89.10 ± 1.62), followed by the lasso method (mean accuracy: 86.88 ± 2.23) and finally, the backward step-wise method (mean accuracy: 86.38 ± 3.52) which achieved a slightly worse performance comparing to the lasso. These results are aligned with the Parkinsonism feature selection results which indicates that the feature selection approach is consistent and capable of obtaining good results for various types of gait patterns. These findings also further corroborate the importance of feature selection when performing classification tasks based on gait features.

The classifiers chosen for the classification of FD without CNS lesions vs. FD with CNS lesions were the ones that achieved the best performance when classifying controls vs. Parkinsonism patients. These were the SVMs (overall good performances), DBNs (when based on the Mean Strides datasets), the LSTMs (when based on the All Strides datasets) and the CNNs (when based on Raw All Strides dataset and the biometric data of each subject).

8.8.2 FD without CNS lesions vs. FD with CNS lesions

Classification performance of FD without CNS lesions vs. FD with CNS lesions was impressive, even more given the small sample size of 20 FD patients of which 13 have CNS lesions. Since this classification task presents a dominant class it is important to note that the base accuracy of a classifier that would always predict the same class, in this case always FD with CNS lesions, is 65%. All the developed classifiers achieved a performance above 65% accuracy, in fact, the worst overall classification performance was obtained by the DBN classifier based on three Normalized Mean Strides features selected by the lasso method, the classifier obtained an accuracy of 70.84% $\pm$ 4.16, a sensibility of 100% $\pm$ 0.0 and a specificity of 70.84% $\pm$ 4.16 (results are shown as mean of validation and test performance $\pm$ standard deviation), the three features were minimum toe clearance, CV of minimum toe clearance and CV of maximum toe clearance 2. The best overall classification performance was obtained by the CNN classifier based on the biometric data of each patient and on two Raw All Strides features selected by the step-wise method, the CNN classifier obtained an accuracy of 93.34% $\pm$ 6.66, a sensibility of 100% $\pm$ 0.0 and a specificity of 93.34% $\pm$ 6.66, the two Raw All Strides features were lift-off angle and stride length. This CNN classifier was extremely similar to the best classifier for the IPD vs. VaP classification task, both classifiers were CNN architectures based on the biometric data, lift-off angle and stride length, the only difference was that the IPD vs. VaP classifier also made use of peak swing. These findings indicate that lift-off angle and stride length can be powerful biomarkers for the gait analysis of disorders that affect the CNS, the results also further corroborate the proposed architecture that makes used of biometric data and gait data simultaneously as input. Contrarily to the controls vs. parkinsonism classification results this CNN classifier seems to have a higher sensitivity than specificity. This means that the classifier is incorrectly predicting that FD patients without CNS lesions suffer from CNS lesions. These results might occur because the dominant class is FD with CNS lesions, a higher sample size with an even number of patients from both classes could improve this performance.

To conclude, these are promising results that contribute to previous studies showing that gait impairment can be present in FD patients (Lohle et al. (2015)), and CNS lesions have an impact on gait patterns (Amboni et al. (2013)).

8.8.3 Clustering of FD patients

As stated in Chapter 2.3, the most common symptoms of FD are at the neurological level but the symptoms that cause the most number of premature deaths occur at the cardiac level (Mehta et al. (2009)), other studies have also connected the cardiac manifestations of FD patients to neurological events such as strokes (Rolfs et al. (2005), Giugliani et al. (2016)). It is then of extreme importance to analyse cardiac data of FD patients and investigate if there is a connection between neurological and cardiac symptoms. The hypothesis in question is if cardiac complications are connected with lesions in the CNS.

To assess this hypothesis, cardiac data was collected from two distinct cardiac exams, the Holter and the Echocardiogram. The cardiac features of each exam were statistically analysed among the two groups (FD without CNS lesions and FD with CNS lesions) using the Mann-Whitney test, a p-value less than 0.05 was considered statistically significant. The cardiac features that presented the most significant differences were selected for the clustering analysis. From all the 27 Holter cardiac features only 13 were considered statistically significant, 5 of which presented a p-value inferior to 0.001. The selected 13 features resulted in 7 PCs capturing 96.28% of the total variance. The best clustering method based on the clustering classification with the number of clusters pre-defined to two was the hierarchical clustering, this method achieved an accuracy of 66.32%. The hierarchical cluster analysis was conducted based on the 7 PCs resulting in five clusters. The two clusters that displayed the least cardiac complications combined are composed of 36 FD patients without CNS lesions (≈61%) and of 23 FD patients with CNS lesions (≈39%), the two clusters that displayed the worst cardiac complications combined are composed of only 1 FD patient without CNS lesions (≈9%) and of 10 FD patients with CNS lesions (≈91%), finally, the intermediate cluster between the two best clusters and the two worst clusters in terms of cardiac complications is composed of 8 FD patients without CNS lesions (≈32%) and of 17 FD patients with CNS lesions (≈68%).

From all the 22 Echocardiogram features, 21 were considered statistically significant, 12 of which presented a p-value inferior to 0.001. The selected 12 features resulted in 11 PCs capturing 96.38% of the total variance. The K-means and hierarchical clustering methods achieved the same clustering classification with the number of clusters pre-defined to two, both methods achieved an accuracy of 68.82%.

The K-means (arbitrary choice) cluster analysis was conducted based on the 11 PCs resulting in four clusters. The two clusters that displayed the least cardiac complications combined are composed of 36 FD patients without CNS lesions (≈64%) and of 20 FD patients with CNS lesions (≈36%), the two clusters that displayed the worst cardiac complications combined are composed of 8 FD patients without CNS lesions (≈22%) and of 10 FD patients with CNS lesions (≈78%).

The clustering results based on the Holter and Echocardiogram cardiac features are very similar, both analyses indicate that cardiac manifestations are associated with increased lesions in the CNS, the clusters characterized by the lowest cardiac manifestations are mainly composed of FD patients without CNS lesions while the clusters describing major cardiac complications are mainly composed of FD patients with CNS lesions, these findings are in accordance with previous studies showing the importance of cardiac manifestations and their connection with sudden neurological manifestation such as strokes and mini-strokes (Brito et al. (2018); Giugliani et al. (2016); Rolfs et al. (2005)).

To conclude, the clustering results indicate that there is a possible correlation between cardiac events and lesions in the CNS. From the statistical analysis, the Echocardiogram features seem to be the most significant for the prediction of CNS lesions, however, the clustering results based on the Holter and Echocardiogram features were similar. It is then important to conduct future analysis where clustering is performed based on the Holter and Echocardiogram features simultaneously, choosing only the most significant features (p-value < 0.001) might lead to further conclusions about the connection between FD cardiac manifestations and CNS lesions.

Chapter 9

Results Parkinsonism and Fabry Disease

In this Chapter, the link between Parkinsonism and FD is analysed. To better understand the connection between these two disorders the machine learning classifiers that achieved the best performance metrics when differentiating between controls and Parkinsonism patients are used to make predictions on the gait data of each FD patient.

To investigate the connection between these two disorders the best machine learning classifiers that were trained to differentiate between normal gait patterns and parkinsonian gait patterns will be used to predict if FD patients display normal gait or parkinsonian gait. The selected machine learning algorithms were the SVMs and the LSTMs. The SVMs were able to achieve the best overall results in all datasets and the LSTMs were the best time-series classifiers when differentiating between controls and Parkinsonism patients.

The data used to train the classifiers was the four gait datasets composed of the 15 control subjects and 29 parkinsonism (15 IPD + 14 VaP) patients. The classifiers were used to make predictions on the four FD gait datasets composed of the 20 FD patients, from which 13 patients have been diagnosed with lesions in the CNS. Furthermore, 3 of these 13 patients also suffer from Parkinsonism. For simplification purposes, the 3 patients that suffer from FD and Parkinsonism simultaneously are labeled and addressed as P1, P2, and P3 in the following analysis.

In the following confusion matrices analysis, only the 3 patients that suffer from FD and Parkinsonism simultaneously are considered as displaying parkinsonian gait patterns, all other FD patients are considered as displaying normal gait patterns.

9.1 Performance evaluation for the SVM classifiers

9.1.1 Performance with Raw All Strides Dataset

The SVM classifier was implemented with the same 6 gait features and the same hyperparameters that achieved the best performance results when differentiating between control subjects and Parkinsonism patients based on the Raw All Strides dataset. The 6 gait features were cadence, loading, speed, peak swing, strike angle, and maximum heel clearance. The hyperparameters were the sigmoid kernel function, a c value of 1.1, and a γ value of 0.1.

After implementation and training, the SVM classifier was used to classify the FD patients' gait patterns as normal or parkinsonian gait patterns. A patients' gait pattern is considered parkinsonian if more than 50% of the strides are classified as belonging to parkinsonian gait patterns. The results are summarized in the following Table 84.

Table 84: Confusion matrix of the SVM classifier for the task of classifying the Raw All Strides gait patterns of FD patients as normal gait or parkinsonian gait.

		Predicted	
		Normal Gait	Parkinsonian Gait
Actual	Normal Gait	15	2
	Parkinsonian Gait	1	2

From the analysis of Table 84 it can be concluded that the classifier predicted that four FD patients display parkinsonian gait patterns. From the four predicted patients, two (P1 and P2) suffer from Parkinsonism and FD simultaneously, the other two patients only suffer from FD. However, these two patients that only suffer from FD have been diagnosed with lesions in the CNS. All other patients, including P3, were classified as displaying normal gait patterns.

9.1.2 Performance with Normalized All Strides Dataset

The SVM classifier was implemented with the same 4 gait features and the same hyperparameters that achieved the best performance results when differentiating between control subjects and Parkinsonism patients based on the Normalized All Strides dataset. The 4 gait features were speed, stride length, strike angle, and maximum heel clearance. The hyperparameters were the sigmoid kernel function, a c value of 0.9, and a γ value of 0.2.

After implementation and training, the SVM classifier was used to classify the FD patients' gait patterns as normal or parkinsonian gait. A patients' gait pattern is considered parkinsonian if more than 50% of the strides are classified as belonging to parkinsonian gait patterns. The results are summarized in the following Table 85.

Table 85: Confusion matrix of the SVM classifier for the task of classifying the Normalized All Strides gait patterns of FD patients as normal gait or parkinsonian gait.

		Predicted	
		Normal Gait	Parkinsonian Gait
Actual	Normal Gait	17	0
	Parkinsonian Gait	1	2

From the analysis of Table 85 it can be concluded that the classifier predicted that P1 and P2 displayed parkinsonian gait patterns. All the other FD patients, including P3, were classified as displaying normal gait patterns.

9.1.3 Performance with Raw Mean Strides Dataset

The SVM classifier was implemented with the same 3 gait features and the same hyperparameters that achieved the best performance results when differentiating between control subjects and Parkinsonism patients based on the Mean Strides dataset. The 3 gait features were speed, peak swing and CV of strike

angle. The parameters were the RBF kernel function, a c value of 0.8, and a γ value of 0.04.

After implementation and training, the SVM classifier was used to classify the FD patients' gait patterns as normal or parkinsonian gait. The results are summarized in the following Table 86.

Table 86: Confusion matrix of the SVM classifier for the task of classifying the Raw Mean Strides gait patterns of FD patients as normal gait or parkinsonian gait.

		Predicted	
		Normal Gait	Parkinsonian Gait
Actual	Normal Gait	16	1
	Parkinsonian Gait	1	2

From the analysis of Table 86 it can be concluded that the classifier predicted that three FD patients displayed parkinsonian gait patterns. Two of these patients are P1 and P2, the other patient only suffers from FD and has been diagnosed with lesions in the CNS. All other FD patients, including P3, were classified as displaying normal gait patterns.

9.1.4 Performance with Normalized Mean Strides Dataset

The SVM classifier was implemented with the same 5 gait features and the same hyperparameters that achieved the best performance results when differentiating between control subjects and Parkinsonism patients based on the Mean Normalized Strides dataset. The 5 gait features were CV of strike angle, speed, CV of maximum toe clearance 2, CV of lift-off angle, and CV of swing. The hyperparameters were the RBF kernel function, a c value of 0.9, and a γ value of 0.03.

After implementation and training, the SVM classifier was used to classify the FD patients gait patterns as control or parkinsonian gait. The results are summarized in the following Table 87.

Table 87: Confusion matrix of the SVM classifier for the task of classifying the Normalized Mean Strides gait patterns of FD patients as normal gait or parkinsonian gait.

		Predicted	
		Normal Gait	Parkinsonian Gait
Actual	Normal Gait	16	1
	Parkinsonian Gait	1	2

From the analysis of Table 87 it can be concluded that these results are very similar to the results achieved with the Mean Raw Strides dataset.

9.2 Performance evaluation for the LSTM classifiers

9.2.1 Performance with Raw All Strides Dataset

The LSTM classifier was implemented with the same 8 gait features and the same hyperparameters that achieved the best performance results when differentiating between control subjects and Parkinsonism patients based on the Raw All Strides dataset. The 8 gait features were speed, peak swing, stride length, strike angle, foot flat, lift-off angle, loading, and cycle duration. The hyperparameters were 22/14 hidden LSTM cells, 22/16 hidden MLP neurons, a learning rate of 0.002, a dropout rate of 0.2, a batch size of 4, and 30 epochs.

After implementation and training, the LSTM classifier was used to classify the FD patients' gait patterns as normal or parkinsonian gait. The results are summarized in the following Table 88.

Table 88: Confusion matrix of the LSTM classifier for the task of classifying the Raw All Strides gait patterns of FD patients as normal gait or parkinsonian gait.

		Predicted	
		Normal Gait	Parkinsonian Gait
Actual	Normal Gait	13	4
	Parkinsonian Gait	1	2

From the analysis of Table 88 it can be concluded that the LSTM classifier predicted that six FD patients displayed parkinsonian gait patterns, two of these six patients are P1 and P2, the other four are FD patients that have been diagnosed with lesions in the CNS. All other FD patients, including P3, were classified as displaying normal gait patterns.

9.2.2 Performance with Normalized All Strides Dataset

The LSTM classifier was implemented with the same 3 gait features and the same hyperparameters that achieved the best performance results when differentiating between control subjects and Parkinsonism patients based on the All Normalized Strides dataset. The 3 gait features were speed, strike angle, and stride length. The hyperparameters were 22/14 hidden LSTM cells, 22/16 hidden MLP neurons, a learning rate of 0.002, a dropout rate of 0.2, a batch size of 4, and 30 epochs.

After implementation and training, the LSTM classifier was used to classify the FD patients gait patterns as normal or parkinsonian gait. The results are summarized in the following Table 89.

Table 89: Confusion matrix of the LSTM classifier for the task of classifying the Raw All Strides gait patterns of FD patients as normal gait or parkinsonian gait.

		Predicted	
		Normal Gait	Parkinsonian Gait
Actual	Normal Gait	13	4
	Parkinsonian Gait	1	2

From the analysis of Table 89 it can be concluded that the results for the All Normalized Strides dataset are similar to the results achieved with the All Raw Strides dataset. Six patients (all with CNS lesions), including P1 and P2, were classified has displaying parkinsonian gait patterns. All other patients, including P3, were classified as displaying normal gait patterns.

9.3 Parkinsonism and Fabry Disease Results Discussion

Recently the connection between FD and Parkinsonism has been investigated and assumptions have been drawn that the higher prevalence of strokes in FD patients might lead to the development of Parkinsonism (Lohle et al. (2015), Wise et al. (2018)). To further investigate this hypothesis and the connection between FD and Parkinsonism, the SVM and LSTM classifiers that achieved the best classification performance when classifying controls vs. Parkinsonism were used to make predictions on the gait patterns of FD patients. The goal was to investigate if FD gait patterns are classified as normal or parkinsonian. The gait features and hyperparameters used to develop the classifiers were the ones that achieved the best performance results when classifying controls vs. Parkinsonism patients. To perform this analysis the classifiers were trained using the four gait datasets composed of 15 controls and 29 Parkinsonism patients (15 IPD + 14 VaP). After training, the four FD gait datasets composed of 20 FD patients, from which 3 suffer from FD and IPD simultaneously, were used to perform the classification. To simplify the description of the results let's considerer that P1, P2, and P3 denote the 3 patients that suffer from FD and Parkinsonism simultaneously.

The results show that all the classifiers classified patients P1 and P2 as displaying parkinsonian gait patterns, the patient P3 was not classified has displaying parkinsonian gait patterns by any of the classifiers, furthermore, the only classifier that only classified patients P1 and P2 as displaying parkinsonian gait patterns is the SVM based on four All Normalized Strides features. All other classifiers predicted that FD patients displayed parkinsonian gait patterns, more specifically, the classifiers that predicted the most FD patients as displaying parkinsonian gait patterns were the LSTM classifiers. With both Raw and Normalized gait datasets the LSTM classifiers predicted that six FD patients suffer from parkinsonian gait patterns,

two of which were patients P1 and P2, the other four FD patients suffer from CNS lesions, in fact, all FD patients that have been classified as displaying parkinsonian gait patterns suffer from CNS lesions. These are interesting results, even more given the fact that in this work the SVM and LSTM classifiers have shown very good consistent results when classifying controls vs. Parkinsonism patients.

To conclude, these findings indicate that FD patients suffering from CNS lesions display gait impairments that are classified as parkinsonian by very reliable classifiers. The patients here classified has displaying parkinsonian gait patterns should be kept under diagnostic so that their condition is constantly evaluated. These results indicate that there is a possibility that the CNS lesions caused by FD might lead to the development of Parkinsonism. The next step should be to investigate if the gait patterns of FD patients can be distinguished from the gait patterns of control subjects, this analysis is performed in the following Chapter 10.

Chapter 10

Results Controls vs. Fabry Disease

The gait dataset for the analysis of control subjects vs. FD patients only became available in the later stages of this work. This dataset is composed of the arithmetic mean of each of the 17 described gait variables (the collected gait variables are described in section 4.2). The aim of this analysis is twofold, firstly, to use an MR gait normalization approach that accounts for subject age, height, weight, gender, speed, and stride length to identify differences in gait features between FD patients and healthy subjects, and secondly, to evaluate the effectiveness of different machine learning methods in classifying between controls and FD patients' gait before and after applying MR gait normalization. The hypothesis is that it will be possible to distinguish FD patients from healthy controls using machine learning approaches based on gait data, and that MR gait normalization will be able to improve the classification performance.

Since the gait dataset was only available in the later stages of this work the implemented classification approach differs from the previous approaches used in this work. For this analysis, a feature extraction approach using PCA was implemented instead of the previously used feature selection approach. The performance of the classifiers was also evaluated using only 10-fold stratified cross-validation. This different approach had to be implemented due to time limitations.

10.1 Gait data normalization

The gait normalization was applied using the previous detailed MR normalization approach (see Section 4.4 and Section 6.3).

Differences in the biometric properties including age, height, weight, gender, speed and stride length of controls vs. FD patients were accessed using the Mann-Whitney U-test. A p-value less than 0.05 was considered statistically significant. The results are described in the following Table 90.

Table 90: Differences between physical properties, speed and stride length of controls and FD patients.

Biometric Property	**Controls vs. FD patients**
Age (years)	$p = 0.384$
Weight (kg)	$p = 0.219$
Height (m)	$\mathbf{p = 0.008}$
Gender	$p = 0.676$
Speed	$p = 0.902$
Stride Length	$p = 0.668$

From the analysis of Table 90 it can be concluded that the groups are well matched, the only significant difference is in terms of height. Variance inflation factors for the independent gait variables were calculated to determine the severity of multicollinearity among the physical properties, speed and stride length. The VIFs for all combinations of independent variables are summarized in Table 91.

Table 91: Variance inflation factors for age, height, weight, sex, speed and stride length.

	Age	Height	Weight	sex	Speed	Stride Length
	2.45	3.85	1.94	1.27	4.13	8.21
VIF	1.89	2.88	1.92	1.26	1.26	—
	2.24	3.45	1.91	1.26	—	2.51

A VIF value is greater than 5 (8.21 for stride length) when considering all six independent variables. Then, speed and stride length independent variables are never considered simultaneously. Physical characteristics combined with speed or stride length were considered in the development of the MR models since their VIFs were less than 5 (Belsley (1991)). For each gait variable, the best regression model was

selected based on the adjusted R^2 and the Akaike information criterion (AIC) values. The developed MR models are summarized in the following Table 92.

Table 92: Resulting multiple linear regression models for the gait variables. The adjusted R^2 and Akaike information criterion (AIC) are shown. The independent variables are age (A), height (H), speed (S), sex (G), weight (W) and stride length (SL).

Gait variable	**Multiple Linear Regression Model**							**AIC**	**Ajusted R^2**
Spatial-Temporal Variables									
Cycle Duration	$= 1.04$	$-0.00095 \cdot A$	$+0.277 \cdot H$	$-0.308 \cdot S$				-109.19	0.615
Cadence	$= 112.22$	$+0.130 \cdot A$	$-29.0 \cdot H$	$+34.96 \cdot S$				203.12	0.680
Stance	$= 62.44$			$-3.35 \cdot S$	$+1.41 \cdot G$			132.17	0.188
Swing	$= 37.60$			$-3.35 \cdot S$	$-1.41 \cdot G$			132.17	0.188
Loading	$= 22.37$			$+3.32 \cdot S$	$-2.30 \cdot G$	$-0.154 \cdot W$		154.73	0.315
Foot Flat	$= 78.31$		$-8.95 \cdot H$	$-16.59 \cdot S$		$+0.194 \cdot W$		176.89	0.543
Pushing	$= 17.23$	$-0.036 \cdot A$		$+13.25 \cdot S$				168.30	0.469
Double Support	$= 25.57$			$-6.41 \cdot S$	$+1.81 \cdot G$			176.12	0.121
Stride Length	$= 0.24$	$-0.0018 \cdot A$	$+0.361 \cdot H$	$+0.455 \cdot S$				-105.97	0.864
Peak Swing	$= 231.52$		$-41.22 \cdot H$	$+129.90 \cdot S$		$+0.833 \cdot W$		305.71	0.526
Foot Clearance Variables									
Strike Angle	$= 19.24$				$-3.03 \cdot G$	$-0.165 \cdot W$	$+16.51 \cdot SL$	167.54	0.542
Lift-Off Angle	$= -46.51$	$+0.218 \cdot A$					$-0.137 \cdot SL$	206.11	0.660
MaxHC	$= 0.101$		$+0.17 \cdot H$		$-0.039 \cdot G$	$-0.00091 \cdot W$		-137.25	0.474
MaxTC1	$= 0.136$				$-0.022 \cdot G$	$-0.00055 \cdot W$		-143.17	0.078
MinTC	$= -0.015$	$+0.00049 \cdot A$	$+0.0121 \cdot H$					-184.17	0.309
MaxTC2	$= 0.095$	$-0.00070 \cdot A$			$-0.033 \cdot G$		$+0.110 \cdot SL$	-142.60	0.602

MaxHC: Maximum Heel Clearance; MaxTC1: Maximum Toe Clearance 1; MinTC: Minimum Toe Clearance; MaxTC2: Maximum Toe Clearance 2.

The Spearman's rank-order correlation coefficients (ρ) were computed to assess the influence of the independent variables on the spatial-temporal and foot clearance gait variables before and after MR normalization. The following Table 93 shows the results of the Spearman's rank-order correlations coefficients test before (raw) and after MR normalization for the control subjects.

Table 93: Spearman's correlation coefficients for the Controls gait data before (raw) and after MR normalization.

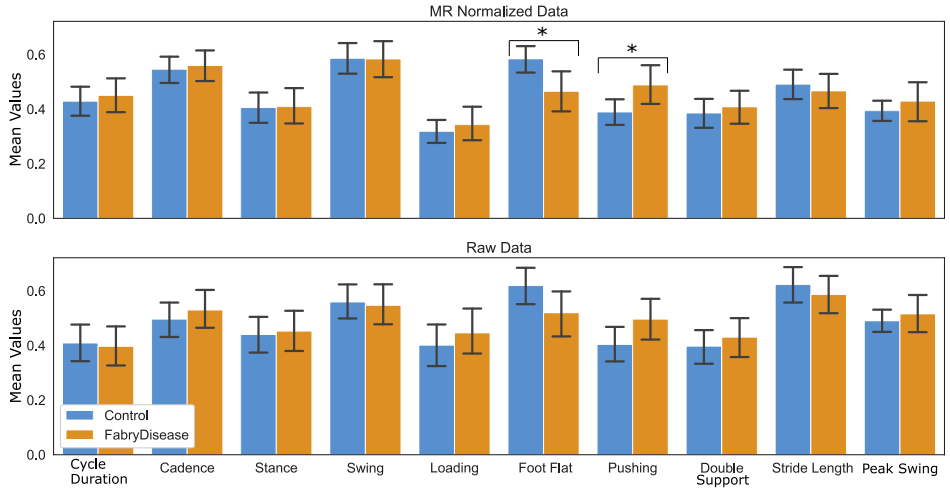
Correlations	Age		Weight		Height		Speed		Stride Length		Gender	
	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm
Spatial-Temporal Variables												
Cycle Duration	−0.28	−0.05	0.28	0.23	0.46	0.26	−0.36	−0.02	0.18	0.39	−0.29	−0.11
Cadence	0.3	0.11	−0.3	−0.27	−0.45	−0.27	0.35	0.02	−0.18	−0.41	0.28	0.1
Stance	0.13	0.0	0.01	0.14	−0.08	0.18	−0.27	−0.01	−0.29	−0.01	0.37	0.0
Swing	−0.13	−0.03	−0.01	−0.15	0.08	−0.18	0.27	−0.01	0.29	0.01	−0.37	−0.01
Loading	−0.05	0.04	−0.38	−0.0	−0.09	0.01	0.12	0.0	0.08	0.03	−0.26	0.01
Foot Flat	0.42	0.1	0.34	0.04	−0.16	−0.11	−0.64	−0.25	−0.58	−0.29	0.03	0.17
Pushing	−0.44	−0.01	−0.06	−0.12	0.3	−0.05	0.68	0.14	0.62	0.11	0.15	0.23
Double Support	0.07	−0.04	0.12	0.2	0.0	0.15	−0.23	−0.01	−0.23	−0.03	0.25	−0.01
Stride Length	−0.63	−0.07	0.23	0.21	0.68	0.24	0.82	0.28	1.0	0.56	−0.11	−0.02
Peak Swing	−0.03	0.13	0.21	0.01	0.25	0.05	0.69	0.14	0.57	0.16	−0.02	−0.07
Foot Clearance Variables												
Strike Angle	−0.44	−0.05	−0.1	0.06	0.36	0.07	0.49	0.06	0.66	0.18	−0.33	0.01
Lift-Off Angle	0.7	−0.03	−0.08	−0.05	−0.54	0.08	−0.65	0.13	−0.77	0.15	0.12	−0.06
Maximum Heel Clearance	−0.38	−0.22	0.22	0.1	0.59	0.18	0.27	0.12	0.52	0.27	−0.56	−0.02
Maximum Toe Clearance 1	0.3	0.36	−0.01	0.14	−0.07	−0.11	−0.33	−0.27	−0.2	−0.18	−0.32	0.01
Minimum Toe Clearance	0.66	0.07	0.02	−0.05	−0.34	−0.17	−0.42	−0.23	−0.47	−0.17	0.09	0.03
Maximum Toe Clearance 2	−0.62	0.02	0.08	−0.19	0.59	−0.08	0.54	0.05	0.77	0.11	−0.45	−0.02

The following Table 94 shows the results of the Spearman's rank order correlations coefficients test before (raw) and after MR normalization for the FD patients.

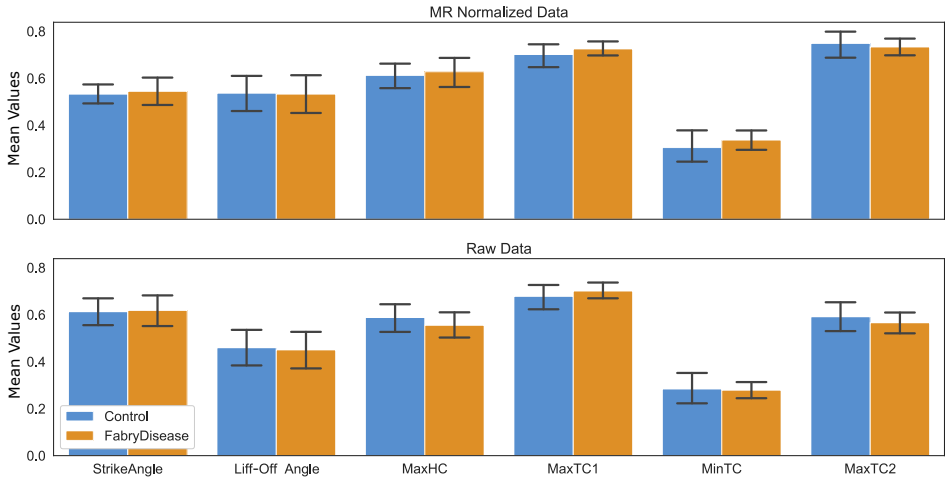
Table 94: Spearman's correlation coefficients for the FD patients gait data before (raw) and after MR normalization.

Correlations	Age		Weight		Height		Speed		Stride Length		sex	
	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm
Spatial-Temporal Variables												
Cycle Duration	0.12	0.06	0.38	0.25	0.32	0.24	−0.63	−0.24	−0.08	0.27	−0.46	−0.37
Cadence	−0.16	−0.17	−0.43	−0.28	−0.4	−0.34	0.63	0.24	0.08	−0.26	0.43	0.34
Stance	0.36	0.11	0.19	0.29	−0.05	0.26	−0.66	−0.42	−0.39	−0.07	−0.15	−0.48
Swing	−0.36	−0.13	−0.19	−0.29	0.05	−0.25	0.66	0.43	0.39	0.08	0.15	0.48
Loading	−0.37	−0.19	0.05	0.37	0.11	−0.04	0.53	0.27	0.56	0.28	−0.29	0.0
Foot Flat	0.4	0.07	0.17	−0.16	−0.05	0.12	−0.63	−0.15	−0.6	−0.22	0.04	0.06
Pushing	−0.32	0.24	−0.22	−0.18	0.01	−0.29	0.5	−0.21	0.45	−0.16	0.15	0.17
Double Support	0.46	0.25	0.17	0.17	−0.15	0.11	−0.75	−0.53	−0.49	−0.2	−0.02	−0.26
Stride Length	−0.7	−0.01	0.01	0.23	0.61	0.38	0.79	0.12	1.0	0.54	−0.25	−0.3
Peak Swing	−0.61	−0.51	−0.07	−0.02	0.17	0.12	0.78	0.52	0.57	0.39	−0.05	−0.13
Foot Clearance Variables												
Strike Angle	−0.63	−0.39	0.08	0.41	0.36	0.0	0.62	0.25	0.73	0.19	−0.29	0.07
Lift-Off Angle	0.7	−0.04	−0.22	0.22	−0.43	0.17	−0.6	0.04	−0.77	0.25	0.26	−0.34
Maximum Heel Clearance	−0.51	−0.34	0.38	0.27	0.48	−0.07	0.33	0.3	0.46	0.22	−0.47	0.19
Maximum Toe Clearance 1	−0.05	0.02	0.07	0.28	0.12	−0.11	0.2	0.12	0.11	−0.07	−0.28	0.23
Minimum Toe Clearance	0.6	−0.16	−0.29	−0.21	−0.28	0.11	−0.31	0.26	−0.45	0.13	0.04	−0.01
Maximum Toe Clearance 2	−0.67	0.05	0.13	0.02	0.55	−0.25	0.53	−0.09	0.69	−0.19	−0.34	0.4

The following Figure 32 shows a comparison between the mean value of raw and MR normalized gait features of FD patients and control subjects.



(a) Controls vs. FD patients spatial-temporal gait variables.



(b) Controls vs. FD patients foot clearance gait variables.

Figure 32: Comparison between the mean value of gait features in FD patients and controls. Data are shown for the MR normalized gait and the raw gait data. Significant differences in gait features between FD patients and controls are indicated with one asterisk (*p $<$:05). Whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot.

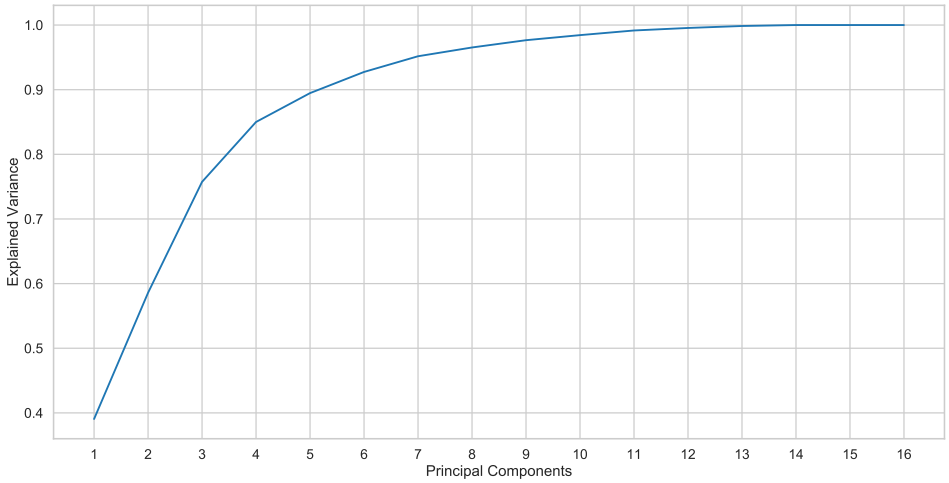
Normalization using the MR approach (see Table 92) was able to reduce the great majority of correlations between the independent variables and gait variables, however, few strong correlations remained

after normalization (Table 93 and Table 94). For both groups, controls and FD, all raw gait variables were strongly correlated with at least one of the independent variables: age, weight, height, speed, sex, stride length and speed. Regarding to the control subjects (Table 93), after normalization all gait variables were weakly correlated except the cycle duration which was mildly correlated with stride length ($\rho = 0.39$) and maximum toe clearance 1 which was mildly correlated with age ($\rho = -0.38$). Concerning to the FD patients (Table 94), after normalization stance and swing were still strongly correlated with speed ($\rho = -0.42$, $\rho = 0.43$) and sex ($\rho = -0.48$, $\rho = 0.48$), double support was strongly correlated with speed ($\rho = -0.53$), peak swing was strongly correlated with age ($\rho = -0.51$), speed ($\rho = 0.52$) and stride length ($\rho = 0.39$), and finally, strike angle was mildly correlated with weight ($\rho = 0.41$).

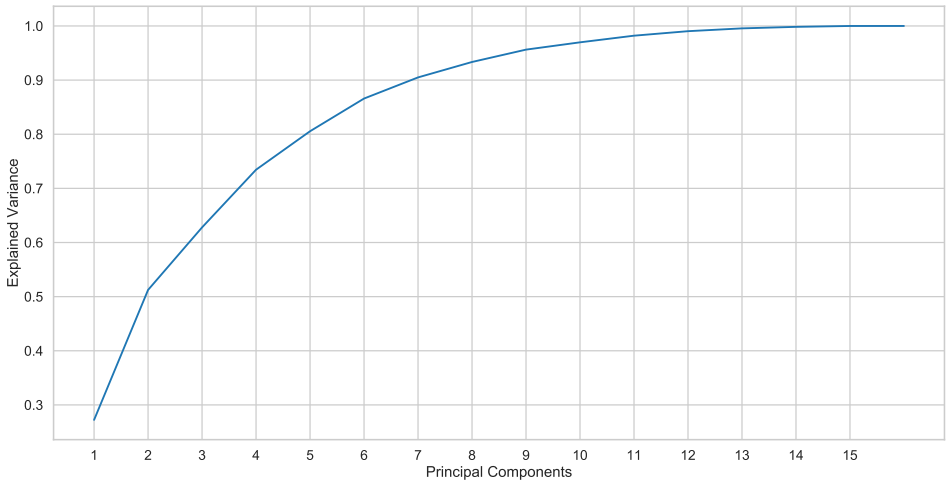
From the analysis of Figure 32 it can be concluded that using raw data, no statistically significant differences were found. After normalization using MR approach, significant differences between controls and FD patients were observed in foot flat (mean difference: 0.11, 95%CI: [0.10;0.14], $p = .011$) and pushing (mean difference: 0.10, 95%CI: [0.08,0.12], $p = .019$), with FD presenting lower percentages in foot flat and higher in pushing. Additional, before normalization the mean value of minimum toe clearance of FD patients was slightly smaller than controls, however after normalization, the mean value of minimum toe clearance of FD patients was greater than controls ($p = 0.051$).

10.2 Performance evaluation for the classifiers

Before the implementation of the machine learning classifiers, the dimensionality of the datasets was reduced using PCA. Using PCA to reduce data dimensionality results in reduction of computational cost and may increase the performance of the machine learning classifiers (Abdi and Williams (2010), Han et al. (2011)). For each dataset, the PCs were selected based on a threshold of around 95% for the total explained variance. The following Figure 33 shows the PCs extracted for the raw and normalized gait datasets.



(a) Raw Principal Components.



(b) Normalized Principal Components.

Figure 33: Principal components and the respective cumulative sum of the variance.

The raw gait data was reduced to the first seven PCs, these explained 95.17% of the total variance, the normalized gait data was reduced to the first nine PCs, these explained 95.63% of the total variance.

Four classifiers were implemented for the differentiation task of controls vs. FD patients based on the extracted PCs. The implemented classifiers were SVMs, RFs, MLPS, and DBNs, these machine learning

classifiers were implemented using: 1) PCs extracted from the raw spatial-temporal and foot clearance gait variables; 2) PCs extracted from the MR normalized spatial-temporal and foot clearance gait variables.

The hyperparameters of each classifier were optimized using a randomized search method and 10-fold stratified cross-validation. For the SVM classifiers, different configurations between the regularization c, Kernel functions and Kernel functions specific hyperparameters (e.g. degree of the polynomial kernel function) were evaluated. For the RF classifiers different configurations between the number of trees, the functions to measure the quality of a split (Gini or Entropy), the minimum number of samples required to split an internal node, the minimum number of samples required to be a leaf node and the maximum number of features to consider at each node split were evaluated. Finally, for the DBN and MLP classifiers, different configurations between the number of hidden neurons, learning rates, dropout rates, batch sizes, and epochs were evaluated. Both classifiers were implemented with two hidden layers. The best classification results are summarized in the following Table 95. The hyperparameters that achieved the best classification results are described in the Appendix.

Table 95: Classification performance measures obtained with each classifier based on the principal components derived from the raw and MR normalized gait variables. Values in parentheses indicate standard deviation of the performance. All performance results are in percentage.

Classifier		Controls vs. FD patients			
		Raw Dataset		MR Normalized Dataset	
		Mean	95% CI	Mean	95% CI
MLP	Accuracy	68.87 (± 16.08)	(56.75:80.99)	72.20 (± 15.11)	(60.81:83.59)
	Sensitivity	70.0 (± 19.05)	(55.62:84.38)	76.83 (± 20.20)	(61.60:92.06)
	Specificity	71.33 (± 19.5)	(56.63:86.04)	72.33 (± 16.33)	(60.02:84.65)
DBN	Accuracy	61.55 (± 11.27)	(53.05:70.04)	65.12 (± 13.87)	(54.66:75.57)
	Specificity	62.0 (± 14.47)	(51.09:72.91)	64.33 (± 17.05)	(51.48:77.19)
	Sensitivity	64.83 (± 15.75)	(52.96:76.71)	67.33 (± 14.95)	(56.06:78.60)
SVM	Accuracy	**71.96 (± 12.57)**	**(62.48:81.44)**	**78.21 (± 9.63)**	**(70.95:85.47)**
	Sensitivity	**75.83 (± 18.05)**	**(62.22:89.44)**	**76.89 (± 12.57)**	**(67.35:86.31)**
	Specificity	**73.0 (± 16.41)**	**(60.62:85.38)**	**88.50 (± 14.67)**	**(77.44:99.56)**
RF	Accuracy	59.52 (± 14.86)	(48.32:70.73)	76.37 (± 10.78)	(68.24:84.50)
	Sensitivity	58.50 (± 24.16)	(40.28:76.72)	80.50 (± 18.23)	(66.76:94.24)
	Specificity	70.50 (± 21.86)	(54.02:86.98)	84.67 (± 16.91)	(72.14:97.19)

Classification accuracy of controls vs. FD patients using ML methods was the lowest when using the PCs extracted from the raw gait variables, and the highest when based on the PCs extracted from the MR normalized gait variables (see Table 95). The SVM classifier performed the best, it was able to yield an accuracy of 78.21% (76.89% sensitivity and 88.50% specificity) when predicting controls vs. FD patients based on the PCs extracted from the MR normalized gait. The MLP, DBN and RF classifiers based on the PCs extracted from MR normalized gait achieved an accuracy of 72.20%, 65.12% and 76.37%, respectively.

Peak classification performances were observed when using the PCs extracted from gait data normalized by the MR approach.

10.3 Controls vs. Fabry Disease patients Results Discussion

The developed work was twofold, firstly it aimed to investigate the capabilities of MR normalization approach for de-correlation between spatial-temporal and foot clearance gait variables, physical properties, speed and stride length. Secondly, it evaluated the performance of ML classifiers based on raw and MR normalized gait variables when distinguishing FD patients and control subjects.

After gait data normalization using MR approach weak to moderate correlations ($0.00 < |\rho| < 0.49$) were observed between physical characteristics, speed, stride length, and gait features, while for raw gait data these correlations were weak to strong ($0.01 < |\rho| < 0.82$). Strong correlations between most of spatial-temporal gait variables and speed were observed, especially in FD patients. After MR normalization all correlations decreased to moderate values ($0.00 < |\rho| < 0.54$). Speed has shown a profound role in predicting spatial-temporal gait variables, all the MR models make use of speed for spatial-temporal prediction (see Table 92). As stated previously, this goes in line with the hypothesis that normal ranges for gait variables should be defined with reference to the speed of walking (Kirtley et al. (1985)), speed has also been correlated with most spatial-temporal gait variables previously (Bejek et al. (2006); Mikos et al. (2018); Wahid et al. (2016, 2015)). Another study also concluded that differences in gait parameters between healthy subjects and osteoarthritis patients decrease when walking speed is accounted for in the gait analysis (Zeni Jr and Higginson (2009)). Additionally, in this analysis stride length was included as an independent variable for the development of the MR models. The results show that foot clearance measurements from FD patients and controls were strongly correlated with stride length, the variables most correlated with stride length were strike angle ($\rho = 0.73$ and $\rho = 0.66$), lift-off angle ($\rho = -0.77$ and $\rho = -0.77$) and maximum toe clearance 2 ($\rho = 0.69$ and $\rho = 0.77$). After MR normalization, using stride length as an independent variable, the correlations between stride length, strike angle, lift-off angle and maximum toe clearance 2 were all weakened ($|\rho| < 0.25$ and $|\rho| < 0.18$). In fact, the regression models for strike angle, lift-off angle and maximum toe clearance 2 explained 54.2% to 66.0% in the observed variance using stride length as an independent variable (Table 92). Stride length seems to be the strongest predictor for foot clearance variables, suggesting that future gait evaluation should account for the effect of stride length on those variables. In this work, the developed MR models best able to

predict foot clearance patterns had stride length as an independent variable. Overall the MR normalization was able to significantly reduce the correlations between physical properties, speed, stride length and gait variables. These MR normalization results are comparable to previous works and further corroborate the favorable outcomes of using MR normalization (Mikos et al. (2018); Wahid et al. (2016, 2015)).

The interpretation of Figure 32 indicates that the MR normalization was able to uncover significant differences in the foot flat and pushing gait variables. These uncovered differences suggest that MR normalization would improve the performance classification of FD gait patterns. This hypothesis is corroborated by the performance of the developed classifiers. All machine learning classifiers performed better based on the MR normalized variables, SVM was the best classifier achieving an accuracy of 78.2%, a sensitivity of 76.9% and a specificity of 88.5%, the accuracy of the SVM classifier increased 6.3% when MR normalization is applied before classification (see Table 95). The SVM classifier achieved the most consistent and best performances among all classifiers both based on raw and normalized gait, these results are in accordance with previous studies that used gait variables to classify gait patterns of various neurological disorders. A recent study used various machine learning classifiers to differentiate between Parkinson's disease and Alzheimer's disease gait patterns, the best performance was achieved by the SVM classifier (92.6% accuracy) (Aich et al. (2018)). Another recent study compared the performance of SVM, MLP, RF, and KNN when differentiating between controls and neurodegenerative disorders based on gait rhythm signals, the best classifier also was the SVM with an accuracy of 96.83% (Xia et al. (2015)).

The results also show the RF classifier has an impressive performance increase when based on the MR normalized gait variables, the RF classifier accuracy increased from 59.5% when using raw gait data to 76.4% when using the MR normalized gait data, it seems that the RF classifier benefits the most from the MR normalization approach. These findings are also alined with a previous study where an RF classifier went from being the worst classifier when classifying PD gait patterns based on raw gait variables to the best classifier when based on MR normalized gait variables, the accuracy increased from 75.0% to 92.6% (Wahid et al. (2015)).

Finally and most importantly, based on a thorough search of the relevant literature, this is the first work that includes an FD group and analysis the capabilities of different machine learning classifiers for the prediction of FD when based on gait data. The results obtained in this work are extremely promising for the diagnosis and evaluation of FD.

Chapter 11

Conclusion and future work

11.1 General conclusions

Recently, machine learning algorithms associated with powerful biomarkers have been presented as an effective support system able to assist doctors in the diagnosis and evaluation of multiple diseases such as IPD, AD, among others (Aich et al. (2018); Costa et al. (2016); Wahid et al. (2015); Xia et al. (2015)). Inspired by these previous studies, the major goal of this dissertation was to evaluate the effectiveness of machine learning approaches based on gait patterns for the diagnosis and evaluation of VaP, IPD, and FD. Various parallel goals where also evaluated in the path of this major goal, these include, the evaluation of various feature selection methods when applied to gait variables, the evaluation of a MR normalization strategy for the de-correlation of physical properties, speed, stride length and gait variables, the identification of subgroups of Parkinsonism patients based on gait patterns, the identification of subgroups of FD patients based on cardiac data and finally, the connection between Parkinsonism and FD patients' gait patterns. This is the first work that investigates the effectiveness of machine learning methods for the diagnosis and understanding of FD and is also the first work that investigates the effectiveness of machine learning methods for the differentiation of IPD and VaP patients based on gait data.

All the developed classifiers based on gait patterns presented extremely good results. The feature selection approach was able to increase the performance of all classifiers. Furthermore, gait normalization

using an MR approach proved to be a very successful pre-classification data processing step, in particular when differentiating between controls and FD patients. However, the superior approach seems to be the proposed architecture based on a CNN that processes gait data and an MLP that processes the physical properties, when this architecture was used it was able to achieve the best classification performance results. These results indicate that physical properties should not be discarded when performing gait analysis, physical properties might uncover specific disease-related gait features that are not present in raw gait data.

The powerful classifiers developed for the classification of controls and Parkinsonism patients were used to classify the gait patterns of FD patients, the classifiers predicted that some FD patients with CNS lesions displayed gait patterns similar to those of patients that suffer from Parkinsonism. These results indicate that the neurological lesions of FD patients might lead to gait impairments severe enough to be considered as parkinsonian gait, these are interesting results that further corroborate the recent connection between FD and Parkinsonism.

The clustering analysis of VaP and IPD patients proved to be an excellent tool to understand the underlying gait patterns of these disorders, the clustering results indicate that the most impacted gait patterns were displayed by VaP patients, these patients display slow, flat-footed and shuffling gait patterns. Clustering analysis was also applied to understand the cardiac manifestations of FD patients, the results indicate that FD patients with severe cardiac manifestations also suffered from severe CNS manifestations, the cardiac features of FD patients seem to be powerful biomarkers for the identification of neurological manifestations. These results also corroborate previous studies indicating that the FD neurological manifestations could be a result of the cardiac manifestations. However, to better understand this connection further research is necessary.

11.2 Limitations

There are various limitations that should be addressed to correctly interpret the results of this dissertation. Firstly, the developed MR models were based on a fairly small number of control subjects ($n = 15$

and $n = 34$), this number is comparable to the number of subjects used in previous works (Mikos et al. (2018); Wahid et al. (2016, 2015)), however, some weak correlations were still present after normalization which may be related to the restricted number of controls or it might be related to the fact that there are other factors besides speed, stride length and physical properties that impact gait variables, an example is cognition (Amboni et al. (2013); Hollman et al. (2007)). A larger number of control subjects aligned with other independent variables like cognition might improve MR models and further reduce the correlations. Focusing now on the classification tasks, although the number of subjects is comparable or higher with respect to previous works (Tahir and Manap (2012), Wahid et al. (2015), Tien et al. (2010), Md. Tahir and Manap (2012), Xia et al. (2015)), it is not exhaustive for the assumption that the selected gait features could be the definitive ones for the classification of the tackled disorders. To validate the best set of gait variables found by the feature selection methods further research with a higher number of subjects is required.

11.3 Future work

For future research, a larger group of control subjects as well as their cognitive measurements should be collected to further evaluate the capabilities of MR normalization for the de-correlation of gait variables and the ability to increase the classification performance. It would also be very interesting to compare the performance results of the MR normalization approach and the proposed CNN and MLP architecture. The results related to FD should also be further investigated, it would be extremely interesting to analyse the results of a machine learning classifier that differentiates FD and Parkinsonism patients based on gait patterns, the results of this classification task can further contribute to the recently uncovered connection between Parkinsonism and FD. The cardiac data of FD patients could also be used alongside gait patterns for the classification of FD patients, this work has shown that cardiac features are extremely significant when comparing FD patients without CNS lesions and FD patients with CNS lesions so, it can be hypothesized that a classifier architecture that makes use of cardiac and gait data would be very successful and could further assist in the diagnosis and evaluation of FD.

Appendix

Table 96: Performance of the best SVM Classifier developed with the 7 Raw All Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	93.24%	87.77%	94.83%
Validation Set	93.42%	92.79%	95.85%
Test Set	99.44%	100%	99.28%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	3.0	0.0001	––

Table 97: Performance of the best SVM Classifier developed with the 8 Raw All Strides features selected with the Backward Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 8 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.28%	83.0%	95.15%
Validation Set	95.18%	93.3%	97.57%
Test Set	94.26%	96.15%	93.81%
Test Set Individual Performance	87.5%	100%	83.33%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	0.9	0.1	––

Table 98: Performance and parameters of the best SVM Classifier developed with the 7 Raw All Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	85.69%	84.51%	87.99%
Validation Set	68.65%	67.6%	65.62%
Test Set	56.47%	69.79%	44.05%
Test Set Individual Performance	66.67%	66.67%	66.67%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.5	0.02	––

Table 99: Performance of the best SVM Classifier developed with the 3 Raw All Strides features selected with the Backward Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	69.1%	65.42%	96.48%
Validation Set	64.66%	60.05%	49.74%
Test Set	72.48%	71.32%	77.91%
Test Set Individual Performance	66.67%	60.0%	100.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	1.7	0.2	4

Table 100: Performance and parameters of the best SVM Classifier developed with the 2 Normalized All Strides features selected with the Lasso method the Controls vs. Parkinsonism differentiation task.

Lasso Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.29%	80.75%	99.8%
Validation Set	92.52%	87.94%	98.73%
Test Set	99.44%	100%	99.28%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.1	4.73	––

Table 101: Performance of the best SVM Classifier developed with the 3 Normalized All Strides features selected with the Backward Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.06%	80.90%	94.14%
Validation Set	92.48%	88.72%	95.66%
Test Set	99.26%	100%	99.04%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	1.2	0.2	––

Table 102: Performance of the best SVM Classifier developed with the 2 Normalized All Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	79.63%	77.67%	84.0%
Validation Set	78.18%	75.98%	76.2%
Test Set	87.68%	85.71%	92.36%
Test Set Individual Performance	83.33%	75.0%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	50	0.02	––

Table 103: Performance of the best SVM Classifier developed with the 6 Normalized All Strides features selected with the Backward Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	79.84%	78.93%	81.73%
Validation Set	77.17%	75.86%	74.37%
Test Set	87.27%	89.07%	84.09%
Test Set Individual Performance	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	5.0	0.02	––

Table 104: Performance and parameters of the best SVM Classifier developed with the 5 Raw Mean Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.67%	84.73%	95.67%
Validation Set	92.17%	91.67%	96.67%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.5	0.1	––

Table 105: Performance of the best SVM Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	93.54%	83.91%	100%
Validation Set	95.5%	91.67%	100%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.8	0.04	––

Table 106: Performance and parameters of the best SVM Classifier developed with the 5 Raw Mean Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	84.06%	82.52%	86.13%
Validation Set	84.17%	91.67%	91.67%
Test Set	33.33%	40.0%	0.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	2.5	0.04	––

Table 107: Performance of the best SVM Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.69%	91.73%	98.18%
Validation Set	89.17%	91.67%	95.0%
Test Set	66.67%	60.0%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	1	0.014	––

Table 108: Performance and parameters of the best SVM Classifier developed with the 5 Normalized Mean Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.9%	82.55%	100%
Validation Set	92.17%	86.67%	100.0%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.5	0.1	––

Table 109: Performance of the best SVM Classifier developed with the 5 Normalized Mean Strides features selected with the Backward Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.27%	86.24%	96.08%
Validation Set	93.5%	91.67%	97.5%
Test Set	100%	100%	100%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	0.9	0.03	––

Table 110: Performance and parameters of the best SVM Classifier developed with the 5 Normalized Mean Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	82.61%	81.93%	93.36%
Validation Set	82.5%	86.67%	85.0%
Test Set	50.0%	50.0%	50.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	30	0.002	––

Table 111: Performance of the best SVM Classifier developed with the 4 Normalized Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	93.71%	95.09%	93.14%
Validation Set	82.5%	91.67%	80.0%
Test Set	50.0%	50.0%	50.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	1	0.014	––

Table 112: Performance and parameters of the best MLP Classifier developed with the 6 Raw All Strides features selected with the Lasso method for the CTR vs. PD differentiation task.

Lasso Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	99.47%	99.8%	99.65%
Validation Set	91.81%	90.53%	95.31%
Test Set	94.81%	89.6%	96.39%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.002	0.3	20/16	16	30

Table 113: Performance and parameters of the best MLP Classifier developed with the 8 Raw All Strides features selected with the Step Wise method for the CTR vs. PD differentiation task.

Step Wise Method - 8 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	98.51%	96.65%	99.07%
Validation Set	86.96%	83.45%	94.46%
Test Set	93.33%	98.92%	92.17%
Test Set Individual Performance	87.5%	100%	83.33%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.2	18/12	32	26

Table 114: Performance and parameters of the best MLP Classifier developed with the 7 Raw All Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 7 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	92.51%	93.26%	91.46%
Validation Set	61.02%	62.66%	64.17%
Test Set	57.49%	72.48%	45.35%
Test Set Individual Performance	66.67%	66.67%	66.67%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.001	0.2	20/16	64	26

Table 115: Performance and parameters of the best MLP Classifier developed with the 3 Raw All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	75.09%	74.82%	64.02%
Validation Set	50.34%	59.35%	54.07%
Test Set	75.98%	77.01%	73.38%
Test Set Individual Performance	83.33%	75.0%	100.0%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.0015	0.2	12/8	32	26

Table 116: Performance and parameters of the best MLP Classifier developed with the 4 Normalized All Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 4 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	96.37%	86.67%	99.68%
Validation Set	90.39%	87.69%	96.56%
Test Set	97.96%	93.28%	99.51%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.002	0.3	24/26	64	26

Table 117: Performance and parameters of the best MLP Classifier developed with the 3 Normalized All Strides features selected with the Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	95.56%	85.09%	99.45%
Validation Set	85.43%	80.75%	94.84%
Test Set	98.89%	100%	98.57%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	Dropout Rate	Neurons	Batch Size	Epochs
0.001	0.1	16/14	64	46

Table 118: Performance and parameters of the best MLP Classifier developed with the 2 Normalized All Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	83.11%	84.14%	81.92%
Validation Set	66.69%	74.12%	66.14%
Test Set	91.99%	93.75%	89.07%
Test Set Individual Performance	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.1	16/12	64	26

Table 119: Performance and parameters of the best MLP Classifier developed with the 6 Normalized All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	95.76%	95.34%	96.4%
Validation Set	66.11%	78.61%	46.9%
Test Set	72.9%	98.87%	58.06%
Test Set Individual Performance	83.33%	100%	75.0%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.3	20/14	64	36

Table 120: Performance and parameters of the best MLP Classifier developed with the 5 Raw Mean Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.76%	86.53%	100%
Validation Set	92.17%	91.67%	96.67%
Test Set	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.2	38/14	2	36

Table 121: Performance and parameters of the best MLP Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.16%	85.22%	100%
Validation Set	93.5%	88.33%	100%
Test Set	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.001	0.2	26/14	6	66

Table 122: Performance and parameters of the best MLP Classifier developed with the 5 Raw Mean Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	88.42%	84.75%	93.16%
Validation Set	74.17%	86.67%	76.67%
Test Set	66.67%	100%	60.0%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.001	0.1	24/26	6	16

Table 123: Performance and parameters of the best MLP Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.64%	91.53%	98.03%
Validation Set	79.17%	91.67%	86.67%
Test Set	66.67%	60.0%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.002	0.1	18/16	2	16

Table 124: Performance and parameters of the best MLP Classifier developed with the 5 Normalized Mean Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.91%	82.73%	100%
Validation Set	93.5%	80.0%	100%
Test Set	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.001	0.2	32/16	6	36

Table 125: Performance and parameters of the best MLP Classifier developed with the 5 Normalized Mean Strides features selected with the Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	96.72%	91.28%	100%
Validation Set	90.83%	85.56%	100%
Test Set	100%	100%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.002	0.3	18/10	2	56

Table 126: Performance and parameters of the best MLP Classifier developed with the 4 Normalized Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.44%	86.1%	87.25%
Validation Set	72.5%	81.67%	70.0%
Test Set	66.67%	100%	60.0%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.2	22/14	2	36

Table 127: Performance and parameters of the best MLP Classifier developed with the 5 Normalized Mean Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	80.18%	80.15%	81.5%
Validation Set	71.67%	75.0%	80.0%
Test Set	83.33%	75.0%	100%

MLP Classifier Hyperparameters

Learning Rate	**Dropout Rate**	**Neurons**	**Batch Size**	**Epochs**
0.0015	0.2	24/14	4	26

Table 128: Performance and parameters of the best DBN Classifier developed with the 7 Raw All Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 7 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	90.86%	75.56%	97.13%
Validation Set	89.40%	86.05%	96.77%
Test Set	99.91%	100%	99.76%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.006	0.002	0.1	14/12	64	10	110

Table 129: Performance and parameters of the best DBN Classifier developed with the 8 Raw All Strides features selected with the Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 8 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	86.87%	66.03%	98.9%
Validation Set	84.84%	72.94%	99.65%
Test Set	97.59%	99.14%	97.17%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.005	0.0015	0.2	16/16	32	10	100

Table 130: Performance and parameters of the best DBN Classifier developed with the 4 Raw All Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	61.16%	64.87%	55.03%
Validation Set	59.26%	62.39%	53.52%
Test Set	76.18%	72.45%	100%
Test Set Individual Performance	66.67%	60.0%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.007	0.0015	0.2	10/12	32	10	100

Table 131: Performance and parameters of the best DBN Classifier developed with the 3 Raw All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	68.92%	69.60%	67.87%
Validation Set	65.97%	69.31%	59.72%
Test Set	80.59%	76.25%	100%
Test Set Individual Performance	83.33%	75.0%	100.0%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0075	0.0015	0.0	16/16	32	10	100

Table 132: Performance and parameters of the best DBN Classifier developed with the 2 Normalized All Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 2 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	93.98%	80.52%	99.37%
Validation Set	93.87%	89.03%	99.87%
Test Set	96.30%	86.39%	100%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.001	0.0015	0.2	20/16	32	8	100

Table 133: Performance and parameters of the best DBN Classifier developed with the 3 Normalized All Strides features selected with the Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	91.97%	77.78%	97.41%
Validation Set	91.45%	83.62%	97.93%
Test Set	95.74%	84.67%	100%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.002	0.0015	0.1	18/16	32	8	100

Table 134: Performance and parameters of the best DBN Classifier developed with the 2 Normalized All Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 2 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	79.92%	78.83%	82.54%
Validation Set	76.46%	66.26%	88.57%
Test Set	88.30%	84.83%	97.71%
Test Set Individual Performance	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.0025	0.002	0.2	16/16	32	8	120

Table 135: Performance and parameters of the best DBN Classifier developed with the 6 Normalized All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 6 Selected Features

	Accuracy	Sensitivity	Specificity
Training Set	70.96%	74.12%	66.71%
Validation Set	72.02%	64.12%	79.35%
Test Set	86.04%	84.75%	89.04%
Test Set Individual Performance	83.33%	75.0%	100%

DBN Classifier Hyperparameters

LR S1	LR S2	DR	Neurons	Batch Size	Epochs S1	Epochs S2
0.002	0.0015	0.1	24/16	32	9	120

Table 136: Performance and parameters of the best DBN Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.67%	80.83%	99.52%
Validation Set	95.5%	91.67%	100%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.0015	0.0	18/16	2	10	90

Table 137: Performance and parameters of the best DBN Classifier developed with the 3 Raw Mean Strides features selected with the Step Wise method for the CTR vs. PD differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.67%	80.83%	99.52%
Validation Set	95.5%	91.67%	100%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.0015	0.0	18/16	2	10	90

Table 138: Performance and parameters of the best DBN Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	68.70%	77.90%	68.49%
Validation Set	75.0%	90.0%	80.0%
Test Set	66.67%	60.0%	100%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.007	0.0015	0.2	24/24	2	7	100

Table 139: Performance and parameters of the best DBN Classifier developed with the 13 Raw Mean Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 13 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	65.37%	66.15%	65.31%
Validation Set	74.17%	75.0%	80.0%
Test Set	83.33%	100%	75.0%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.0015	0.1	14/16	4	10	70

Table 140: Performance and parameters of the best DBN Classifier developed with the 5 Normalized Mean Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	93.25%	83.4%	100%
Validation Set	92.17%	86.67%	100%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.0015	0.0	24/16	2	10	120

Table 141: Performance and parameters of the best DBN Classifier developed with the 5 Normalized Mean Strides features selected with the Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	95.50%	91.67%	100%
Validation Set	90.15%	80.18%	98.75%
Test Set	100%	100%	100%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.0075	0.002	0.1	18/24	2	8	100

Table 142: Performance and parameters of the best DBN Classifier developed with the 5 Normalized Mean Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	68.48%	81.61%	66.83%
Validation Set	75.0%	100%	75.0%
Test Set	66.67%	66.67%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.0075	0.002	0.2	24/18	2	8	120

Table 143: Performance and parameters of the best DBN Classifier developed with the 4 Normalized Mean Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	70.18%	74.85%	69.11%
Validation Set	74.17%	91.67%	80.0%
Test Set	66.67%	66.67%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.0075	0.0015	0.2	16/16	2	10	100

Table 144: Performance and parameters of the best LSTM Classifier developed with the 6 Raw All Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	88.95%	78.70%	97.60%
Validation Set	93.50%	90.0%	96.67%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	24/16	18/14	2	30

Table 145: Performance and parameters of the best LSTM Classifier developed with the 7 Raw All Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	89.55%	80.49%	97.29%
Validation Set	93.50%	90.0%	100%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	38/16	24/18	2	30

Table 146: Performance and parameters of the best LSTM Classifier developed with the 7 Raw All Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	65.52%	66.34%	75.52%
Validation Set	64.17%	75.0%	80.0%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	64/32	32/16	2	30

Table 147: Performance and parameters of the best LSTM Classifier developed with the 3 Raw All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	61.34%	62.50%	64.37%
Validation Set	66.67%	75.0%	66.67%
Test Set	83.33%	75.0%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	24/16	18/14	2	30

Table 148: Performance and parameters of the best LSTM Classifier developed with the 2 Normalized All Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	89.23%	79.95%	95.40%
Validation Set	90.17%	80.0%	96.67%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.2	22/14	22/16	4	20

Table 149: Performance and parameters of the best LSTM Classifier developed with the 4 Normalized All Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	82.39%	79.74%	86.92%
Validation Set	90.17%	80.0%	96.67%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	22/14	22/16	4	30

Table 150: Performance and parameters of the best LSTM Classifier developed with the 2 Normalized All Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	76.87%	73.14%	84.18%
Validation Set	82.50%	81.67%	80.0%
Test Set	83.33%	75.0%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.2	22/14	22/16	4	20

Table 151: Performance and parameters of the best LSTM Classifier developed with the 6 Normalized All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	74.39%	74.27%	76.98%
Validation Set	77.50%	81.67%	85.0%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	64/32	32/16	2	20

Table 152: Performance and parameters of the best CNN Classifier developed with the 6 Raw All Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	93.19%	86.17%	97.79%
Validation Set	92.17%	81.67%	96.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	26/14	5	22/16	4	40

Table 153: Performance and parameters of the best CNN Classifier developed with the 8 Raw All Strides features selected with the Step Wise method for the Controls vs. Parkinsonism differentiation task.

Step Wise Method - 8 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.91%	85.22%	97.60%
Validation Set	92.17%	91.67%	96.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	32/14	5	22/16	2	20

Table 154: Performance and parameters of the best CNN Classifier developed with the 7 Raw All Strides features selected with the Lasso method for the IPD vs. VaP differentiation task.

Lasso Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.04%	81.51%	92.11%
Validation Set	75.0%	70.0%	80.0%
Test Set	83.33%	100%	75.0%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	28/14	5	22/16	4	40

Table 155: Performance and parameters of the best CNN Classifier developed with the 3 Raw All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	65.80%	65.13%	70.48%
Validation Set	64.17%	70.0%	65.0%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	22/14	3	22/16	2	30

Table 156: Performance and parameters of the best CNN Classifier developed with the 2 Normalized All Strides features selected with the Lasso method for the Controls vs. Parkinsonism differentiation task.

Lasso Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.43%	83.70%	95.78%
Validation Set	86.83%	85.0%	96.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.003	0.2	22/14	5	22/16	4	20

Table 157: Performance and parameters of the best CNN Classifier developed with the 4 Normalized All Strides features selected with the Gini method for the Controls vs. Parkinsonism differentiation task.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	90.17%	85.0%	100%
Validation Set	89.54%	88.15%	97.55%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	20/16	5	18/14	2	30

Table 158: Performance and parameters of the best CNN Classifier developed with the 2 Normalized All Strides features selected with the Gini method for the IPD vs. VaP differentiation task.

Gini Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	79.70%	75.20%	85.42%
Validation Set	77.50%	81.67%	75.0%
Test Set	83.33%	75.0%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.2	22/14	5	22/16	4	20

Table 159: Performance and parameters of the best CNN Classifier developed with the 6 Normalized All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task.

Step Wise Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	80.18%	80.40%	82.63%
Validation Set	82.50%	81.67%	80.0%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.0005	0.2	42/18	7	24/18	2	20

Table 160: Performance and parameters of the best CNN Classifier developed with the 7 Raw All Strides features selected with the Lasso method for the IPD vs. VaP differentiation task. The features were collected when the patients were on and off medication.

Lasso Method - 7 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	97.14%	97.27%	97.50%
Validation Set	94.15%	96.67%	96.67%
Test Set	83.33%	100%	75.0%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	48/24	5	40/22	4	30

Table 161: Performance and parameters of the best CNN Classifier developed with the 4 Raw All Strides features selected with the Gini method for the IPD vs. VaP differentiation task. The features were collected when the patients were on and off medication.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	90.38%	90.45%	90.92%
Validation Set	84.17%	91.67%	91.67%
Test Set	100%	100%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.3	20/16	5	20/16	4	24

Table 162: Performance and parameters of the best CNN Classifier developed with all 17 Raw All Strides gait features for the IPD vs. VaP differentiation task. The features were collected when the patients were on and off medication.

All 17 Gait Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	97.14%	97.33%	97.42%
Validation Set	84.17%	91.67%	91.67%
Test Set	83.33%	100%	75.0%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	48/24	5	40/22	4	30

Table 163: Performance and parameters of the best CNN Classifier developed with the 3 Normalized All Strides features selected with the Lasso method for the IPD vs. VaP differentiation task. The features were collected when the patients were on and off medication.

Lasso Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	85.99%	83.16%	89.09%
Validation Set	82.50%	91.67%	90.0%
Test Set	83.33%	100%	75.0%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.3	20/16	5	20/16	4	30

Table 164: Performance and parameters of the best CNN Classifier developed with the 6 Normalized All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task. The features were collected when the patients were on and off medication.

Step Wise Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.08%	92.17%	90.87%
Validation Set	87.50%	91.67%	95.0%
Test Set	83.33%	75.0%	100%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.0015	0.2	22/12	5	22/12	2	30

Table 165: Performance and parameters of the best CNN Classifier developed with all 17 Normalized All Strides gait features for the IPD vs. VaP differentiation task. The features were collected when the patients were on and off medication.

All 17 Gait Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	97.12%	96.36%	98.17%
Validation Set	77.50%	86.67%	90.0%
Test Set	83.33%	100%	75.0%

CNN Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	48/24	5	40/22	4	30

Table 166: Performance and parameters of the best CNN/MLP Classifier developed with the 4 Raw All Strides features selected with the Gini method for the IPD vs. VaP differentiation task. The classifier was developed with the gait features collected while the patients are on and off medication and the biometric data of each patient.

Gini Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	88.93%	87.28%	91.57%
Validation Set	87.50%	91.67%	95.0%
Test Set	100%	100%	100%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.002	0.3	20/16	5	16/8	20/16	4	24

Table 167: Performance and parameters of the best CNN/MLP Classifier developed with the 3 Raw All Strides features selected with the Step Wise method for the IPD vs. VaP differentiation task. The classifier was developed with the gait features collected while the patients are on and off medication and the biometric data of each patient.

Step Wise Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.09%	86.39%	86.61%
Validation Set	84.17%	91.67%	91.67%
Test Set	100%	100%	100%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.001	0.2	22/14	5	16/8	22/14	4	30

Table 168: Performance and parameters of the best CNN/MLP Classifier developed with all 17 Raw All Strides gait features and the biometric data of each patient for the IPD vs. VaP differentiation task. The classifier was developed with the gait features collected while the patients are on and off medication.

All 17 Gait Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	85.57%	84.88%	87.71%
Validation Set	81.67%	90.0%	86.67%
Test Set	83.33%	75.0%	100%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.001	0.2	32/16	7	16/8	32/16	4	30

Table 169: Performance of the best SVM Classifier developed with the 3 Raw All Strides features selected with the Gini method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	90.80%	96.11%	89.08%
Validation Set	86.56%	69.76%	82.73%
Test Set	73.23%	100%	72.8%
Test Set Individual Performance	66.37%	100%	66.37%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	1.2	0.17	5

Table 170: Performance of the best SVM Classifier developed with the 2 Raw All Strides features selected with the Backward Step Wise method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Step Wise Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.45%	93.21%	90.87%
Validation Set	81.60%	69.44%	82.11%
Test Set	72.44%	100%	72.22%
Test Set Individual Performance	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	25	0.2	3

Table 171: Performance and parameters of the best SVM Classifier developed with the 4 Normalized All Strides features selected with the Lasso method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Lasso Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.78%	100%	89.28%
Validation Set	83.78%	89.19%	83.46%
Test Set	74.02%	100%	73.39%
Test Set Individual Performance	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	1.1	0.2	5

Table 172: Performance of the best SVM Classifier developed with the 6 Normalized All Strides features selected with the Backward Step Wise method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Step Wise Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.20%	99.89%	92.23%
Validation Set	85.63%	77.11%	86.66%
Test Set	79.53%	100%	77.78%
Test Set Individual Performance	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	70	0.1	5

Table 173: Performance of the best SVM Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	86.95%	80.39%	91.98%
Validation Set	86.67%	90.0%	96.67%
Test Set	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
RBF	4.0	0.1	––

Table 174: Performance of the best SVM Classifier developed with the 2 Raw Mean Strides features selected with the Backward Step Wise method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Step Wise Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	77.77%	100.0%	74.53%
Validation Set	86.67%	100%	86.67%
Test Set	66.67%	100%	66.67%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Polynomial	0.5	0.17	5

Table 175: Performance and parameters of the best SVM Classifier developed with the 3 Normalized Mean Strides features selected with the Lasso method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Lasso Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	92.12%	86.48%	96.33%
Validation Set	90.0%	90.0%	100.0%
Test Set	33.33%	0%	50.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Sigmoid	2.5	0.2	––

Table 176: Performance of the best SVM Classifier developed with the 5 Normalized Mean Strides features selected with the Backward Step Wise method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Step Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	94.79%	87.62%	100.0%
Validation Set	90.0%	95.0%	95.0%
Test Set	33.33%	0%	50.0%

SVM Classifier Hyperparameters

Kernel Function	**C**	**Gamma**	**Degree**
Linear	3.0	––	––

Table 177: Performance of the best DBN Classifier developed with the 3 Raw Mean Strides features selected with the Gini method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	76.67%	100%	76.67%
Validation Set	65.43%	95.56%	68.10%
Test Set	66.67%	100%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.002	0.2	14/12	2	10	70

Table 178: Performance of the best DBN Classifier developed with the 2 Raw Mean Strides features selected with the Backward Step Wise method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Step Wise Method - 2 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	81.67%	100.0%	81.67%
Validation Set	64.76%	90.10%	71.43%
Test Set	66.67%	100%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.007	0.0015	0.1	24/24	4	7	100

Table 179: Performance of the best DBN Classifier developed with the 6 Normalized Mean Strides features selected with the Gini method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 6 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	67.62%	98.33%	68.33%
Validation Set	75.0%	100%	75.0%
Test Set	66.67%	100%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.001	0.2	16/22	2	10	150

Table 180: Performance of the best DBN Classifier developed with the 5 Normalized Mean Strides features selected with the Backward Step Wise method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Step Wise Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	67.62%	98.33%	68.33%
Validation Set	75.0%	100%	75.0%
Test Set	66.67%	100%	66.67%

DBN Classifier Hyperparameters

LR S1	**LR S2**	**DR**	**Neurons**	**Batch Size**	**Epochs S1**	**Epochs S2**
0.006	0.002	0.2	24/18	4	10	100

Table 181: Performance and parameters of the best LSTM Classifier developed with the 5 Raw All Strides features selected with the Lasso method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	77.64%	85.92%	80.18%
Validation Set	66.67%	80.0%	86.67%
Test Set	100%	100%	100%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	22/14	22/16	4	30

Table 182: Performance of the best LSTM Classifier developed with the 3 Raw All Strides features selected with the Gini method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	71.93%	80.10%	80.41%
Validation Set	76.67%	90.0%	86.67%
Test Set	66.67%	100%	66.67%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	32/16	24/18	2	30

Table 183: Performance and parameters of the best LSTM Classifier developed with the 4 Normalized All Strides features selected with the Lasso method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Lasso Method - 4 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	75.77%	93.75%	75.30%
Validation Set	85.0%	100%	85.0%
Test Set	66.67%	100%	66.67%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.002	0.2	22/14	22/16	4	30

Table 184: Performance of the best LSTM Classifier developed with the 5 Normalized All Strides features selected with the Gini method for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	79.90%	83.45%	82.35%
Validation Set	85.0%	95.0%	90.0%
Test Set	66.67%	50.0%	100.0%

LSTM Classifier Hyperparameters

LR	**DR**	**LSTM Neurons**	**MLP Neurons**	**Batch Size**	**Epochs**
0.001	0.2	32/16	24/18	2	30

Table 185: Performance and parameters of the best CNN/MLP Classifier developed with the 5 Raw All Strides features selected with the Lasso method and biometric data of each patient for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Lasso Method - 5 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	91.58%	85.90%	95.36%
Validation Set	81.67%	90.0%	91.67%
Test Set	100%	100%	100%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.001	0.2	28/16	5	16/8	28/16	4	30

Table 186: Best hyperparameter configuration for each classifier for the differentiation of controls vs. FD patients based on raw and MR normalized gait data.

Classifier		**Controls vs. FD Patients**	
		Raw Dataset	**Normalized Dataset**
MLP	**Hidden Neurons**	20/22	20/22
	Dropout Rate	0.1	0.1
	Learning Rate	0.0015	0.0015
	Batch Size	6	6
	Epochs	48	48
DBN	**Hidden Neurons**	18/12	18/12
	Dropout Rate	0.1	0.1
	Learning Rate S1/S2	0.001/0.001	0.001/0.001
	Batch Size	4	4
	Epochs S1/S2	110/11	110/11
SVM	**c**	0.9	0.55
	γ	0.3	0.047
	kernel	RBF	RBF
	δ	0	0
RF	**Split Criteria**	Gini	Gini
	Max Features	4	4
	Number of Trees	200	140
	Min Samples Split	6	3

S1: Stage 1 (unsupervised learning stage); S2: Stage 2 (fine tuning learning stage); Max Features: The number of features to consider when looking for the best split; Min Samples Split: The minimum number of samples required to split an internal node.

Table 187: Performance of the best CNN/MLP Classifier developed with the 3 Raw All Strides features selected with the Gini method and biometric data of each patient for the FD with CNS lesions vs. FD without CNS lesions differentiation task.

Gini Method - 3 Selected Features

	Accuracy	**Sensitivity**	**Specificity**
Training Set	88.77%	83.48%	92.27%
Validation Set	85.0%	90.0%	95.0%
Test Set	66.67%	100%	66.67%

CNN/MLP Classifier Hyperparameters

LR	**DR**	**CNN Neurons**	**Kernel**	**MLP Bio**	**Final MLP**	**Batch Size**	**Epochs**
0.001	0.2	22/12	5	16/8	22/12	4	30

Table 188: Holter cardiac features significance level according to Mann Whitney U-test when comparing FD patients without CNS lesions and FD patients with CNS lesions.

Feature	**Mann Whitney U Test p-value**
HR Max	< 0.001
Unique APCs	< 0.001
Total SVE Beats	< 0.001
QT Mean	< 0.001
Unique PVC	< 0.001
Total VE Beats	0.0017
QTc > 450	0.0025
QT Min	0.0027
QT Max	0.0036
QTc Mean	0.0047
Total Heart Beats	0.014
HR Mean	0.0152
Unique VEs	0.0153
ASDNN 5	0.0666
QTc Max	0.0986
MaxSTCa3	0.1508
MaxSTCa2	0.2007
MinSTCa1	0.2452
QTc Min	0.2476
MaxSTCa1	0.2535
RMSSD	0.2906
SDNN	0.3140
SDANN 5	0.3167
HR Min	0.3545
MinSTCa3	0.3771
MinSTCa2	0.3971
Longest R-R	0.4404

Table 189: Significance level of Echocardiogram features according to Mann Whitney U-test when comparing FD patients without CNS lesions and FD patients with CNS lesions

Feature	**Mann Whitney U Test p-value**
MV E/A Ratio	< 0.001
MV A Vel	< 0.001
E' Lateral	< 0.001
MV Dec T	< 0.001
E' Septal	< 0.001
E/E' Lateral	< 0.001
E/E' Medial	< 0.001

LVPWd	< 0.001
E/E' Septal	< 0.001
IVSd	< 0.001
LVIDd	< 0.001
LADiam/SC	< 0.001
AoDiam	0.0029
S' Lateral	0.0063
MVEVel	0.0072
LVdMassInd(ASE)	0.0099
LADiam	0.0121
S' Septal	0.0136
A' Septal	0.0181
LVdMass(ASE)	0.0199
LVIDd/SC	0.0315
A' Lateral	0.0545

Bibliography

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

Aich, S., Choi, K.-W., Pradhan, P. M., Park, J., and Kim, H.-C. (2018). A performance comparison based on machine learning approaches to distinguish parkinson's disease from alzheimer disease using spatiotemporal gait signals. *Advanced Science Letters*, 24(3):2058–2062.

Alam, M. N., Garg, A., Munia, T. T. K., Fazel-Rezai, R., and Tavakolian, K. (2017). Vertical ground reaction force marker for Parkinson's disease. *PLOS ONE*, 12(5):e0175951.

Alcock, L., Galna, B., Perkins, R., Lord, S., and Rochester, L. (2018). Step length determines minimum toe clearance in older adults and people with Parkinson's disease. *Journal of Biomechanics*, 71:30–36.

Ali, K. and Morris, H. R. (2015). Parkinson's disease: chameleons and mimics. *Practical Neurology*, 15(1):14–25.

Allen, N. E., Schwarzel, A. K., and Canning, C. G. (2013). Recurrent falls in Parkinson's disease: a systematic review. *Parkinson's disease*, 2013:906274.

Alpaydin, E. (2010). *Introduction to machine learning*. MIT Press.

Alsheikh, M. A., Selim, A., Niyato, D., Doyle, L., Lin, S., and Tan, H.-P. (2016). Deep Activity Recognition Models with Triaxial Accelerometers.

Amboni, M., Barone, P., and Hausdorff, J. M. (2013). Cognitive contributions to gait and falls: evidence and implications. *Movement disorders : official journal of the Movement Disorder Society*, 28(11):1520–33.

Ascherio, A. and Schwarzschild, M. A. (2016). The epidemiology of Parkinson's disease: risk factors and prevention. *The Lancet Neurology*, 15(12):1257–1272.

Association, A. H. (2019a). Echocardiogram (echo) | american heart association.

Association, A. H. (2019b). Electrocardiogram (ecg or ekg) | american heart association.

Association, A. H. (2019c). Holter monitor | american heart association.

Baldereschi, M., Di Carlo, A., Rocca, W. A., Vanni, P., Maggi, S., Perissinotto, E., Grigoletto, F., Amaducci, L., and Inzitari, D. (2000). Parkinson's disease and parkinsonism in a longitudinal study: two-fold higher incidence in men. ILSA Working Group. Italian Longitudinal Study on Aging. *Neurology*, 55(9):1358–63.

Bäzner, H., Oster, M., Daffertshofer, M., and Hennerici, M. (2000). Assessment of gait in subcortical vascular encephalopathy by computerized analysis: a cross-sectional and longitudinal study. *Journal of neurology*, 247(11):841–9.

Bejek, Z., Paróczai, R., Illyés, Á., and Kiss, R. M. (2006). The influence of walking speed on gait parameters in healthy people and in patients with osteoarthritis. *Knee Surgery, Sports Traumatology, Arthroscopy*, 14(7):612–622.

Belsley, D. A. (1991). A Guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, 4(1):33–50.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166.

Benito-León, J., Bermejo-Pareja, F., Morales-González, J. M., Porta-Etessam, J., Trincado, R., Vega, S., Louis, E. D., and Neurological Disorders in Central Spain (NEDICES) Study Group (2004). Incidence of Parkinson disease and parkinsonism in three elderly populations of central Spain. *Neurology*, 62(5):734–41.

Berg, D., Lang, A. E., Postuma, R. B., Maetzler, W., Deuschl, G., Gasser, T., Siderowf, A., Schapira, A. H., Oertel, W., Obeso, J. A., Olanow, C. W., Poewe, W., and Stern, M. (2013). Changing the research criteria for the diagnosis of Parkinson's disease: obstacles and opportunities. *The Lancet Neurology*, 12(5):514–524.

Birkmayer, W. and Hornykiewicz, O. (1961). The L-3,4-dioxyphenylalanine (DOPA)-effect in Parkinson-akinesia. *Wiener klinische Wochenschrift*, 73:787–8.

Bolsover, F. E., Murphy, E., Cipolotti, L., Werring, D. J., and Lachmann, R. H. (2014). Cognitive dysfunction and depression in fabry disease: a systematic review. *Journal of inherited metabolic disease*, 37(2):177–187.

Borsini, W., Giuliacci, G., Torricelli, F., Pelo, E., Martinelli, F., and Scordo, M. R. (2002). Anderson-Fabry disease with cerebrovascular complications in two Italian families. *Neurological Sciences*, 23(2):49–53.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, New York, New York, USA. ACM Press.

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brito, D., Cardim, N., Lopes, L. R., Belo, A., Mimoso, J., Gonçalves, L., Madeira, H., and Portuguese Registry of Hypertrophic Cardiomyopathy (PRo-HCM) Investigators (2018). Awareness of Fabry disease in cardiology: A gap to be filled. *Revista Portuguesa de Cardiologia*, 37(6):457–466.

Buechner, S., De Cristofaro, M. T. R., Ramat, S., and Borsini, W. (2006). Parkinsonism and Anderson Fabry's disease: A case report. *Movement Disorders*, 21(1):103–107.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference - A Practical InformationTheoretic Approach*. Springer-Verlag New York, 2 edition.

Caballero, L., Kou, S., Dulgheru, R., Gonjilashvili, N., Athanassopoulos, G. D., Barone, D., Baroni, M., Cardim, N., Gomez de Diego, J. J., Oliva, M. J., Hagendorff, A., Hristova, K., Lopez, T., Magne, J., Martinez, C., de la Morena, G., Popescu, B. A., Penicka, M., Ozyigit, T., Rodrigo Carbonero, J. D., Salustri, A., Van De Veire, N., Von Bardeleben, R. S., Vinereanu, D., Voigt, J.-U., Zamorano, J. L., Bernard, A., Donal, E., Lang, R. M., Badano, L. P., and Lancellotti, P. (2015). Echocardiographic reference ranges for normal cardiac Doppler data: results from the NORRE Study. *European Heart Journal - Cardiovascular Imaging*, 16(9):1031–1041.

Cardim, N., Brito, D., Rocha Lopes, L., Freitas, A., Araújo, C., Belo, A., Gonçalves, L., Mimoso, J., Olivotto, I., Elliott, P., and Madeira, H. (2018). The Portuguese Registry of Hypertrophic Cardiomyopathy: Overall results. *Revista Portuguesa de Cardiologia (English Edition)*, 37(1):1–10.

Che, C., Xiao, C., Liang, J., Jin, B., Zho, J., and Wang, F. (2017). An RNN Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson's Disease. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 198–206. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Chen, G., Ward, B. D., Xie, C., Li, W., Wu, Z., Jones, J. L., Franczak, M., Antuono, P., and Li, S.-J. (2011). Classification of Alzheimer Disease, Mild Cognitive Impairment, and Normal Cognitive Status with Large-Scale Network Analysis Based on Resting-State Functional MR Imaging. *Radiology*, 259(1):213–221.

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., and Sun, J. (2016). Doctor AI: Predicting Clinical Events via Recurrent Neural Networks.

Colosimo, C., Morgante, L., Antonini, A., Barone, P., Avarello, T. P., Bottacchi, E., Cannas, A., Ceravolo, M. G., Ceravolo, R., Cicarelli, G., Gaglio, R. M., Giglia, L., Iemolo, F., Manfredi, M., Meco, G., Nicoletti, A., Pederzoli, M., Petrone, A., Pisani, A., Pontieri, F. E., Quatrale, R., Ramat, S., Scala, R., Volpe, G., Zappulla, S., Bentivoglio, A. R., Stocchi, F., Trianni, G., Del Dotto, P., Simoni, L., Marconi, R., and PRIAMO STUDY GROUP (2010). Non-motor symptoms in atypical and secondary parkinsonism: the PRIAMO study. *Journal of Neurology*, 257(1):5–14.

Costa, L., Gago, M. F., Yelshyna, D., Ferreira, J., David Silva, H., Rocha, L., Sousa, N., and Bicho, E. (2016). Application of Machine Learning in Postural Control Kinematics for the Diagnosis of Alzheimer's Disease. *Computational Intelligence and Neuroscience*, 2016(891253):1–15.

Critchley, M. (1929). Arteriosclerotic Parkinsonism. *Brain*, 52(1):23–83.

Dadashi, F., Mariani, B., Rochat, S., Büla, C. J., Santos-Eggimann, B., and Aminian, K. (2013). Gait and foot clearance parameters obtained using shoe-worn inertial sensors in a large-population sample of older adults. *Sensors (Basel, Switzerland)*, 14(1):443–57.

Daliri, M. R. (2012). Automatic diagnosis of neuro-degenerative diseases using gait dynamics. *Measurement*, 45(7):1729–1734.

de Veber, G. A., Schwarting, G. A., Kolodny, E. H., and Kowall, N. W. (1992). Fabry disease: Immunocytochemical characterization of neuronal involvement. *Annals of Neurology*, 31(4):409–415.

Demirkiran, M., Bozdemir, H., and Sarica, Y. (2001). Vascular parkinsonism: a distinct, heterogeneous clinical entity. *Acta neurologica Scandinavica*, 104(2):63–7.

Devore, J. L. (2012). *Probability and statistics for engineering and the sciences*. Brooks/Cole, Cengage Learning.

Deyan, G. (2019). How Many Webites Are There in 2019.

Dick, F. D., De Palma, G., Ahmadi, A., Scott, N. W., Prescott, G. J., Bennett, J., Semple, S., Dick, S., Counsell, C., Mozzoni, P., Haites, N., Wettinger, S. B., Mutti, A., Otelea, M., Seaton, A., Soderkvist, P., Felice, A., and study group, G. (2007). Environmental risk factors for Parkinson's disease and parkinsonism: the Geoparkinson study. *Occupational and Environmental Medicine*, 64(10):666–672.

Dorsey, E. R., Constantinescu, R., Thompson, J. P., Biglan, K. M., Holloway, R. G., Kieburtz, K., Marshall, F. J., Ravina, B. M., Schifitto, G., Siderowf, A., and Tanner, C. M. (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5):384–6.

Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G. B., and Collins, D. L. (2008). MRI-Based Automated Computer Classification of Probable AD Versus Normal Controls. *IEEE Transactions on Medical Imaging*, 27(4):509–520.

Duchesne, S., Rolland, Y., and Vérin, M. (2009). Automated Computer Differential Classification in Parkinsonian Syndromes via Pattern Analysis on MRI. *Academic Radiology*, 16(1):61–70.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *The Journal of Machine Learning Research*, 12:2121–2159.

Dutta, S., Chatterjee, A., and Munshi, S. (2009). An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification. *Expert Systems*, 26(2):202–217.

Ecker, C., Marquand, A., Mourao-Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., Maltezos, S., Murphy, C. M., Robertson, D., Williams, S. C., and Murphy, D. G. M. (2010). Describing the Brain in Autism in Five Dimensions–Magnetic Resonance Imaging-Assisted Diagnosis of Autism Spectrum Disorder Using a Multiparameter Classification Approach. *Journal of Neuroscience*, 30(32):10612–10623.

El Missiri, A. M., El Meniawy, K. A. L., Sakr, S. A. S., and Mohamed, A. S. E.-d. (2016). Normal reference values of echocardiographic measurements in young Egyptian adults. *The Egyptian Heart Journal*, 68(4):209–215.

Everitt, B., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*. Wiley.

Fernandes, C., Fonseca, L., Ferreira, F., Gago, M., Costa, L., Sousa, N., Ferreira, C., Gama, J., Erlhagen, W., and Bicho, E. (2018). Artificial Neural Networks Classification of Patients with Parkinsonism based on Gait. *IEEE Journal of Biomedical and Health Informatics*, pages 1–7.

Ferreira, F., Gago, M. F., Bicho, E., Carvalho, C., Mollaei, N., Rodrigues, L., Sousa, N., Rodrigues, P. P., Ferreira, C., and Gama, J. (2019). Gait stride-to-stride variability and foot clearance pattern analysis in idiopathic parkinson's disease and vascular parkinsonism. *Journal of biomechanics*.

Fischer, A. and Igel, C. (2012). An Introduction to Restricted Boltzmann Machines. pages 14–36. Springer, Berlin, Heidelberg.

Fitzgerald, P. M. and Jankovic, J. (1989). Lower body parkinsonism: Evidence for vascular etiology. *Movement Disorders*, 4(3):249–260.

Focke, N. K., Helms, G., Scheewe, S., Pantel, P. M., Bachmann, C. G., Dechent, P., Ebentheuer, J., Mohr, A., Paulus, W., and Trenkwalder, C. (2011). Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic parkinson syndrome and healthy controls. *Human Brain Mapping*, 32(11):1905–1915.

Foltynie, T., Barker, R., and Brayne, C. (2002). Vascular Parkinsonism: A Review of the Precision and Frequency of the Diagnosis. *Neuroepidemiology*, 21(1):1–7.

Foundation, N. F. D. (2019). How is fabry disease inherited?

Fox, K., Borer, J. S., Camm, A. J., Danchin, N., Ferrari, R., Lopez Sendon, J. L., Steg, P. G., Tardif, J.-C., Tavazzi, L., and Tendera, M. (2007). Resting Heart Rate in Cardiovascular Disease. *Journal of the American College of Cardiology*, 50(9):823–830.

GaitUp Switzerland (2018). Gait Analyser Description of Walking Performance. *Datasheet*.

Geron, A. (2015). *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*.

Ginsberg, L., Manara, R., Valentine, A., Kendall, B., and Burlina, A. (2006). Magnetic resonance imaging changes in Fabry disease. *Acta Paediatrica*, 95(0):57–62.

Giugliani, R., Niu, D. M., Ramaswami, U., West, M., Hughes, D., Kampmann, C., Pintos-Morell, G., Nicholls, K., Schenk, J. M., and Beck, M. (2016). A 15-year perspective of the fabry outcome survey. *Journal of Inborn Errors of Metabolism and Screening*, 4:1–12.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. In Gordon, G., Dunson, D., and Dudík, M., editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA. PMLR.

Goetz, C. G. (2011). The history of Parkinson's disease: early clinical descriptions and neurological therapies. *Cold Spring Harbor perspectives in medicine*, 1(1):a008862.

Golik, P., Doetsch, P., and Ney, H. (2013). Cross-Entropy vs. Squared Error Training: a Theoretical and Experimental Comparison.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Gupta, D. and Kuruvilla, A. (2011). Vascular parkinsonism: what makes it different? *Postgraduate Medical Journal*, 87(1034):829–836.

Hale, J. (2019). More Than 500 Hours Of Content Are Now Being Uploaded To YouTube Every Minute.

Hall, M. A. and Smith, L. A. (1998). Practical feature subset selection for machine learning.

Haller, S., Badoud, S., Nguyen, D., Garibotto, V., Lovblad, K., and Burkhard, P. (2012). Individual Detection of Patients with Parkinson Disease using Support Vector Machine Analysis of Diffusion Tensor Imaging Data: Initial Results. *American Journal of Neuroradiology*, 33(11):2123–2128.

Han, J., Kamber, M., and Pei, J. (2011). *Data Mining – Concepts & Techniques*. Elsevier.

Hastie, T. and Qian, J. (2014). Glmnet vignette. *Retrieve from http://www. web. stanford. edu/~ hastie/Papers/Glmnet_Vignette. pdf. Accessed September*, 20:42.

Hastie, T., Tibshirani, R., and Friedman, J. H. J. H. (2009). *The elements of statistical learning : data mining, inference, and prediction*.

Haykin, S. (2009). *Neural Networks and Learning Machines Third Edition*.

Hebb, D. (1949). *The organization of behavior : a neuropsychological theory*. New York: Wiley.

Heemels, M.-T. (2016). Neurodegenerative diseases. *Nature*, 539(7628):179–179.

Heiss, W. D. and Würker, M. (1999). Value of functional imaging in Parkinson's disease and related movement disorders. *Der Nervenarzt*, 70 Suppl 1:S2–10.

Hinton, G. E. (2012). A Practical Guide to Training Restricted Boltzmann Machines. pages 599–619. Springer, Berlin, Heidelberg.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–1554.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hoehn, M. M. and Yahr, M. D. (1967). Parkinsonism: onset, progression and mortality. *Neurology*, 17(5):427–42.

Hollman, J. H., Kovash, F. M., Kubik, J. J., and Linbo, R. A. (2007). Age-related differences in spatiotemporal markers of gait stability during dual task walking. *Gait & posture*, 26(1):113–119.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441.

Hua, J., Xiong, Z., Lowey, J., Suh, E., and Dougherty, E. R. (2004). Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*, 21(8):1509–1515.

Huang, Z., Jacewicz, M., and Pfeiffer, R. F. (2002). Anticardiolipin antibody in vascular parkinsonism. *Movement Disorders*, 17(5):992–997.

Hughes, A. J., Daniel, S. E., Kilford, L., and Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of neurology, neurosurgery, and psychiatry*, 55(3):181–4.

Inoue, T., Hattori, K., Ihara, K., Ishii, A., Nakamura, K., and Hirose, S. (2013). Newborn screening for Fabry disease in Japan: prevalence and genotypes of Fabry disease in a pilot study. *Journal of Human Genetics*, 58(8):548–552.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning. page 440.

Javed, F., Thomas, I., and Memedi, M. (2018). A comparison of feature selection methods when using motion sensors data: a case study in parkinson's disease. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5426–5429. IEEE.

Jellinger, K. A. (2003). Prevalence of cerebrovascular lesions in Parkinson's disease. A postmortem study. *Acta neuropathologica*, 105(5):415–9.

Johnson, J. N. and Ackerman, M. J. (2009). QTc: how long is too long? *British journal of sports medicine*, 43(9):657–62.

Joshi, S., Shenoy, D., G.G., V. S., Rrashmi, P., Venugopal, K., and Patnaik, L. (2010). Classification of Alzheimer's Disease and Parkinson's Disease by Using Machine Learning and Neural Network Methods. In *2010 Second International Conference on Machine Learning and Computing*, pages 218–222. IEEE.

Kalia, L. V. and Lang, A. E. (2015). Parkinson's disease. *The Lancet*, 386(9996):896–912.

Kampmann, C., Perrin, A., and Beck, M. (2015). Effectiveness of agalsidase alfa enzyme replacement in Fabry disease: cardiac outcomes after 10 years' treatment. *Orphanet Journal of Rare Diseases*, 10(1):125.

Kaye, E. M., Kolodny, E. H., Logigian, E. L., and Ullman, M. D. (1988). Nervous system involvement in Fabry's disease: Clinicopathological and biochemical correlation. *Annals of Neurology*, 23(5):505–509.

Khemphila, A. and Boonjing, V. (2012). Parkinsons Disease Classification using Neural Network and Feature Selection.

Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization.

Kirtley, C., Whittle, M. W., and Jefferson, R. (1985). Influence of walking speed on gait parameters. *Journal of biomedical engineering*, 7(4):282–288.

Klein, C. and Westenberger, A. (2012). Genetics of Parkinson's Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(1):a008888–a008888.

Kloppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., Fox, N. C., Jack, C. R., Ashburner, J., and Frackowiak, R. S. J. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–689.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

Kolodny, E., Fellgiebel, A., Hilz, M. J., Sims, K., Caruso, P., Phan, T. G., Politei, J., Manara, R., and Burlina, A. (2015). Cerebrovascular involvement in fabry disease: current status of knowledge. *Stroke*, 46(1):302–313.

Korten, A., Lodder, J., Vreeling, F., Boreas, A., van Raak, L., and Kessels, F. (2001). Stroke and idiopathic Parkinson's disease: does a shortage of dopamine offer protection against stroke? *Movement disorders : official journal of the Movement Disorder Society*, 16(1):119–23.

Kubota, K. J., Chen, J. A., and Little, M. A. (2016). Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement Disorders*, 31(9):1314–1326.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4):541–551.

Lehosit, J. B. and Leslie J., C. (2015). Early Parkinsonism: Distinguishing Idiopathic Parkinson's Disease from Other Syndromes. *NCBI*, 22(6):257–265.

Lev, J. (1949). The Point Biserial Coefficient of Correlation. *The Annals of Mathematical Statistics*, 20(1):125–126.

Liepelt-Scarfone, I., Gräber, S., Fruhmann Berger, M., Feseker, A., Baysal, G., Csoti, I., Godau, J., Gaenslen, A., Huber, H., Srulijes, K., Brockmann, K., and Berg, D. (2012). Cognitive Profiles in Parkinson's Disease and Their Relation to Dementia: A Data-Driven Approach. *International Journal of Alzheimer's Disease*, 2012:1–11.

Linhart, A., Kampmann, C., Zamorano, J. L., Sunder-Plassmann, G., Beck, M., Mehta, A., Elliott, P. M., and European FOS Investigators (2007). Cardiac manifestations of Anderson-Fabry disease: results from the international Fabry outcome survey. *European Heart Journal*, 28(10):1228–1235.

Lipton, Z. C., Kale, D. C., Elkan, C., and Wetzel, R. (2017). Learning to Diagnose with LSTM Recurrent Neural Networks.

Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., and Feng, D. (2014). Early diagnosis of Alzheimer's disease with deep learning. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 1015–1018. IEEE.

Lohle, M., Hughes, D., Milligan, A., Richfield, L., Reichmann, H., Mehta, A., and Schapira, A. H. V. (2015). Clinical prodromes of neurodegeneration in Anderson-Fabry disease. *Neurology*, 84(14):1454–1464.

Lord, S., Galna, B., and Rochester, L. (2013). Moving forward on gait measurement: Toward a more refined approach. *Movement Disorders*, 28(11):1534–1543.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees.

Malek, N., Lawton, M. A., Swallow, D. M. A., Grosset, K. A., Marrinan, S. L., Bajaj, N., Barker, R. A., Burn, D. J., Hardy, J., Morris, H. R., Williams, N. M., Wood, N., Ben-Shlomo, Y., Grosset, D. G., and PRoBaND Clinical Consortium, o. b. o. t. P. C. (2016). Vascular disease and vascular risk factors in relation to motor features and cognition in early Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, 31(10):1518–1526.

Manap, H. H., Md Tahir, N., and Yassin, A. I. M. (2011). Statistical analysis of parkinson disease gait classification using Artificial Neural Network. In *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 060–065. IEEE.

Marras, C. and Lang, A. (2013). Parkinson's disease subtypes: lost in translation? *Journal of Neurology, Neurosurgery & Psychiatry*, 84(4):409–415.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

Md. Tahir, N. and Manap, H. H. (2012). Parkinson Disease Gait Classification based on Machine Learning Approach. *Journal of Applied Sciences*, 12(2):180–185.

Mehta, A., Beck, M., Elliott, P., Giugliani, R., Linhart, A., Sunder-Plassmann, G., Schiffmann, R., Barbey, F., Ries, M., Clarke, J., and Fabry Outcome Survey investigators (2009). Enzyme replacement therapy with agalsidase alfa in patients with Fabry's disease: an analysis of registry data. *The Lancet*, 374(9706):1986–1996.

Mehta, A., Ricci, R., Widmer, U., Dehout, F., Garcia de Lorenzo, A., Kampmann, C., Linhart, A., Sunder-Plassmann, G., Ries, M., and Beck, M. (2004). Fabry disease defined: baseline clinical manifestations of 366 patients in the Fabry Outcome Survey. *European Journal of Clinical Investigation*, 34(3):236–242.

Miguel-Puga, A., Villafuerte, G., Salas-Pacheco, J., and Arias-Carrión, O. (2017). Therapeutic Interventions for Vascular Parkinsonism: A Systematic Review and Meta-analysis. *Frontiers in Neurology*, 8:481.

Mikos, V., Yen, S.-C., Tay, A., Heng, C.-H., Chung, C. L. H., Liew, S. H. X., Tan, D. M. L., and Au, W. L. (2018). Regression analysis of gait parameters and mobility measures in a healthy cohort for subject-specific normative values. *PloS one*, 13(6):e0199215.

Minsky, M. (1951). A neural-analogue calculator based upon a probability model of reinforcement.

Minsky, M. and Papert, S. (1969). *Perceptrons; an introduction to computational geometry*. MIT Press.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2012). *Foundations of machine learning*.

Mu, J., Chaudhuri, K. R., Bielza, C., de Pedro-Cuesta, J., Larrañaga, P., and Martinez-Martin, P. (2017). Parkinson's Disease Subtypes Identified from Cluster Analysis of Motor and Non-motor Symptoms. *Frontiers in aging neuroscience*, 9:301.

Murphy, K. P. (2012). Machine learning: a probabilistic perspective.

Nancy Jane, Y., Khanna Nehemiah, H., and Arputharaj, K. (2016). A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease. *Journal of Biomedical Informatics*, 60:169–176.

Neumann, J., Bras, J., Deas, E., O'Sullivan, S. S., Parkkinen, L., Lachmann, R. H., Li, A., Holton, J., Guerreiro, R., Paudel, R., Segarane, B., Singleton, A., Lees, A., Hardy, J., Houlden, H., Revesz, T., and Wood, N. W. (2009). Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. *Brain*, 132(7):1783–1794.

Nguyen, T. L. and Thomas, L. (2010). Supraventricular Ectopic Activity: When Excessive it is not all Benign! *Journal of atrial fibrillation*, 3(2):307.

Nuytemans, K., Theuns, J., Cruts, M., and Van Broeckhoven, C. (2010). Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Human Mutation*, 31(7):763–780.

Oertel, W. H. (2017). Recent advances in treating Parkinson's disease. *F1000Research*, 6:260.

Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., and Acharya, U. R. (2018). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, pages 1–7.

Ondo, W. G., Chan, L. L., and Levy, J. K. (2002). Vascular parkinsonism: clinical correlates predicting motor improvement after lumbar puncture. *Movement disorders : official journal of the Movement Disorder Society*, 17(1):91–7.

Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience & Biobehavioral Reviews*, 36(4):1140–1152.

Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Pereira, C. R., Pereira, D. R., Rosa, G. H., Albuquerque, V. H., Weber, S. A., Hook, C., and Papa, J. P. (2018). Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. *Artificial Intelligence in Medicine*, 87:67–77.

Perera, T. and Thevathasan, W. (2014). An Introduction to Parkinson's Disease. Technical report, Bionics Institute of Australia.

Perumal, S. V. and Sankar, R. (2016). Gait and tremor assessment for patients with Parkinson's disease using wearable sensors. *ICT Express*, 2(4):168–174.

Pilotto, A., Turrone, R., Liepelt-Scarfone, I., Bianchi, M., Poli, L., Borroni, B., Alberici, A., Premi, E., Formenti, A., Bigni, B., Cosseddu, M., Cottini, E., Berg, D., and Padovani, A. (2016). Vascular Risk Factors and Cognition in Parkinson's Disease. *Journal of Alzheimer's Disease*, 51(2):563–570.

Plant, C., Teipel, S. J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A. W., Hampel, H., and Ewers, M. (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *NeuroImage*, 50(1):162–174.

Post, B., Speelman, J. D., Haan, R. J., and study group, C. (2008). Clinical heterogeneity in newly diagnosed Parkinson's disease. *Journal of Neurology*, 255(5):716–722.

Pringsheim, T., Jette, N., Frolkis, A., and Steeves, T. D. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders*, 29(13):1583–1590.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151.

Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., and Ng, A. Y. (2017). Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks.

Rampello, L., Alvano, A., Battaglia, G., Raffaele, R., Vecchio, I., and Malaguarnera, M. (2005). Different clinical and evolutional patterns in late idiopathic and vascular parkinsonism. *Journal of Neurology*, 252(9):1045–1049.

Rees, D. G. (1989). *Essential Statistics*. Springer US, Boston, MA.

Rektor, I., Bohnen, N. I., Korczyn, A. D., Gryb, V., Kumar, H., Kramberger, M. G., de Leeuw, F.-E., Pirtošek, Z., Rektorová, I., Schlesinger, I., Slawek, J., Valkovič, P., and Veselý, B. (2018). An updated diagnostic approach to subtype definition of vascular parkinsonism – Recommendations from an expert working group. *Parkinsonism & Related Disorders*, 49:9–16.

Rivera Gallego, A., López Rodríguez, M., Barbado Hernández, F. J., Barba Romero, M. A., García de Lorenzo Y Mateos, A., Pintos Morelle, G., and Grupo Español de Estudio de Fabry Outcome Survey (2006). Fabry disease in Spain: first analysis of the response to enzyme replacement therapy. *Medicina clinica*, 127(13):481–4.

Rolfs, A., Böttcher, T., Zschiesche, M., Morris, P., Winchester, B., Bauer, P., Walter, U., Mix, E., Löhr, M., Harzer, K., Strauss, U., Pahnke, J., Grossmann, A., and Benecke, R. (2005). Prevalence of Fabry disease in patients with cryptogenic stroke: a prospective study. *The Lancet*, 366(9499):1794–1796.

Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton : (Project Para)*. Cornell Aeronautical Laboratory report. Cornell Aeronautical Laboratory, Buffalo, NY.

Rovini, E., Maremmani, C., Moschetti, A., Esposito, D., and Cavallo, F. (2018). Comparative Motor Preclinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches. *Annals of Biomedical Engineering*, 46(12):2057–2068.

Rudic, B., Schimpf, R., and Borggrefe, M. (2014). Short QT Syndrome – Review of Diagnosis and Treatment. *Arrhythmia & Electrophysiology Review*, 3(2):76.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., Arora, T., and Taheri, S. (2016). Sleep Quality Prediction From Wearable Data Using Deep Learning. *JMIR mHealth and uHealth*, 4(4).

Schuepbach, W., Rau, J., Knudsen, K., Volkmann, J., Krack, P., Timmermann, L., Hälbig, T., Hesekamp, H., Navarro, S., Meier, N., Falk, D., Mehdorn, M., Paschen, S., Maarouf, M., Barbe, M., Fink, G., Kupsch, A., Gruber, D., Schneider, G.-H., Seigneuret, E., Kistner, A., Chaynes, P., Ory-Magne, F., Brefel Courbon, C., Vesper, J., Schnitzler, A., Wojtecki, L., Houeto, J.-L., Bataille, B., Maltête, D., Damier, P.,

Raoul, S., Sixel-Doering, F., Hellwig, D., Gharabaghi, A., Krüger, R., Pinsker, M., Amtage, F., Régis, J.-M., Witjas, T., Thobois, S., Mertens, P., Kloss, M., Hartmann, A., Oertel, W., Post, B., Speelman, H., Agid, Y., Schade-Brittinger, C., and Deuschl, G. (2013). Neurostimulation for Parkinson's Disease with Early Motor Complications. *http://dx.doi.org/10.1056/NEJMoa1205158*.

Shah, J. S. and Elliott, P. M. (2005). Fabry disease and the heart: an overview of the natural history and the effect of enzyme replacement therapy. *Acta paediatrica (Oslo, Norway : 1992). Supplement*, 94(447):11–4; discussion 9–10.

Shahbakhi, M., Far, D. T., and Tahami, E. (2014). Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. *Journal of Biomedical Science and Engineering*, 07(04):147–156.

Sibon, I., Fenelon, G., Quinn, N. P., and Tison, F. (2004). Vascular parkinsonism. *Journal of Neurology*, 251(5):513–524.

Sidransky, E., Nalls, M., Aasly, J., Aharon-Peretz, J., Annesi, G., Barbosa, E., Bar-Shira, A., Berg, D., Bras, J., Brice, A., Chen, C.-M., Clark, L., Condroyer, C., De Marco, E., Dürr, A., Eblan, M., Fahn, S., Farrer, M., Fung, H.-C., Gan-Or, Z., Gasser, T., Gershoni-Baruch, R., Giladi, N., Griffith, A., Gurevich, T., Januario, C., Kropp, P., Lang, A., Lee-Chen, G.-J., Lesage, S., Marder, K., Mata, I., Mirelman, A., Mitsui, J., Mizuta, I., Nicoletti, G., Oliveira, C., Ottman, R., Orr-Urtreger, A., Pereira, L., Quattrone, A., Rogaeva, E., Rolfs, A., Rosenbaum, H., Rozenberg, R., Samii, A., Samaddar, T., Schulte, C., Sharma, M., Singleton, A., Spitz, M., Tan, E.-K., Tayebi, N., Toda, T., Troiano, A., Tsuji, S., Wittstock, M., Wolfsberg, T., Wu, Y.-R., Zabetian, C., Zhao, Y., and Ziegler, S. (2009). Multicenter Analysis of Glucocerebrosidase Mutations in Parkinson's Disease. *New England Journal of Medicine*, 361(17):1651–1661.

Sigmundsdottir, L., Tchan, M. C., Knopman, A. A., Menzies, G. C., Batchelor, J., and Sillence, D. O. (2014). Cognitive and psychological functioning in fabry disease. *Archives of clinical neuropsychology*, 29(7):642–650.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359.

Simpson, R., Langtree, J., and Mitchell, A. (2017). Ectopic Beats: How Many Count?

Spada, M., Pagliardini, S., Yasuda, M., Tukel, T., Thiagarajan, G., Sakuraba, H., Ponzone, A., and Desnick, R. J. (2006). High Incidence of Later-Onset Fabry Disease Revealed by Newborn Screening*. *The American Journal of Human Genetics*, 79(1):31–40.

Stutz, D. (2014). Understanding Convolutional Neural Networks. Technical report.

Sutskever, I. (2014). *Training Recurrent Neural Networks.* Library and Archives Canada.

Tahir, N. M. and Manap, H. H. (2012). Parkinson Disease gait classification based on machine learning approach. *Journal of Applied Sciences*, 12(2):180–185.

Thadhani, R. and Pastores, G. M. (2002). Advances in the management of Anderson–Fabry disease:enzyme replacement therapy. *Expert Opinion on Biological Therapy*, 2(3):325–333.

Thanvi, B., Lo, N., and Robinson, T. (2005). Vascular parkinsonism–an important cause of parkinsonism in older people. *Age and Ageing*, 34(2):114–119.

Tien, I., Glaser, S. D., and Aminoff, M. J. (2010). Characterization of gait abnormalities in parkinson's disease using a wireless inertial sensor system. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 3353–3356. IEEE.

Up, G. (2019). Gait up - make sense of motion | science.

Uribe, C., Segura, B., Baggio, H. C., Abos, A., Garcia-Diaz, A. I., Campabadal, A., Marti, M. J., Valldeoriola, F., Compta, Y., Tolosa, E., and Junque, C. (2018). Cortical atrophy patterns in early Parkinson's disease patients using hierarchical cluster analysis. *Parkinsonism & related disorders*, 50:3–9.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York, New York, NY.

Wahid, F., Begg, R., Lythgo, N., Hass, C. J., Halgamuge, S., and Ackland, D. C. (2016). A Multiple Regression Approach to Normalization of Spatiotemporal Gait Features. *Journal of Applied Biomechanics*, 32(2):128–139.

Wahid, F., Begg, R. K., Hass, C. J., Halgamuge, S., and Ackland, D. C. (2015). Classification of Parkinson's Disease Gait Using Spatial-Temporal Gait Features. *IEEE Journal of Biomedical and Health Informatics*, 19(6):1794–1802.

Werbos, P. (1974). Beyond regression : new tools for prediction and analysis in the behavioral sciences.

Wilcox, W. R., Oliveira, J. P., Hopkin, R. J., Ortiz, A., Banikazemi, M., Feldt-Rasmussen, U., Sims, K., Waldek, S., Pastores, G. M., Lee, P., Eng, C. M., Marodi, L., Stanford, K. E., Breunig, F., Wanner, C., Warnock, D. G., Lemay, R. M., Germain, D. P., and Fabry Registry (2008). Females with Fabry disease frequently have major organ involvement: Lessons from the Fabry Registry. *Molecular Genetics and Metabolism*, 93(2):112–128.

Winikates, J. and Jankovic, J. (1999). Clinical correlates of vascular parkinsonism. *Archives of neurology*, 56(1):98–102.

Wirdefeldt, K., Adami, H.-O., Cole, P., Trichopoulos, D., and Mandel, J. (2011). Epidemiology and etiology of Parkinson's disease: a review of the evidence. *European Journal of Epidemiology*, 26(S1):1–58.

Wise, A. H., Yang, A., Naik, H., Stauffer, C., Zeid, N., Liong, C., Balwani, M., Desnick, R. J., and Alcalay, R. N. (2018). Parkinson's disease prevalence in Fabry disease: A survey study. *Molecular Genetics and Metabolism Reports*, 14(September 2017):27–30.

Wroge, T. J., Ozkanca, Y., Demiroglu, C., Si, D., Atkins, D. C., and Ghomi, R. H. (2018). Parkinson's Disease Diagnosis Using Machine Learning and Voice. In *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–7. IEEE.

Wu, G., Yan, B., Wang, X., Feng, X., Zhang, A., Xu, X., and Dong, H. (2008). Decreased activities of lysosomal acid alpha-D-galactosidase A in the leukocytes of sporadic Parkinson's disease. *Journal of the Neurological Sciences*, 271(1-2):168–173.

Xia, Y., Gao, Q., and Ye, Q. (2015). Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models. *Biomedical Signal Processing and Control*, 18:254–262.

Zareapoor, M. and K. R, S. (2015). Feature Extraction or Feature Selection for Text Classification: A Case Study on Phishing Email Detection. *International Journal of Information Engineering and Electronic Business*, 7(2):60–65.

Zeiler, M. D. (2012). ADADELTA: An Adaptive Learning Rate Method.

Zeni Jr, J. A. and Higginson, J. S. (2009). Differences in gait parameters between healthy subjects and persons with moderate and severe knee osteoarthritis: a result of altered walking speed? *Clinical biomechanics*, 24(4):372–378.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 55(3):856–867.

Zhang, L., Wu, X., and Luo, D. (2015). Real-Time Activity Recognition on Smartphones Using Deep Neural Networks. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, pages 1236–1242. IEEE.

Zijlmans, J. C. M., Katzenschlager, R., Daniel, S. E., and Lees, A. J. L. (2004). The L-dopa response in vascular parkinsonism. *Journal of neurology, neurosurgery, and psychiatry*, 75(4):545–7.