

UNIVERSITY OF THE WITWATERSRAND

MASTERS DISSERTATION

---

**Automated Quantitative Discrimination of  
Parkinson's Disease Stages using Signal  
Processing and Machine Learning**

---

*Author:*

Nabeel SEEDAT- 719484

*A Dissertation submitted in fulfillment of the requirements  
for the degree of Master of Science*

*in the*

Signal Processing Research Group  
School of Electrical and Information Engineering

January 2019



The financial assistance of the **National Research Foundation (NRF)** towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the **NRF**.

# Declaration

I, **Nabeel Seedat**, declare that this Dissertation titled, ”**Automated Quantitative Discrimination of Parkinson’s Disease Stages using Signal Processing and Machine Learning**” and the work presented in it are my own. I confirm that:

- ◊ This work was done wholly or mainly while in candidature for a research degree at this University.
- ◊ Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, has been clearly stated.
- ◊ Where I have consulted the published work of others, this is always clearly attributed.
- ◊ Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work.
- ◊ I have acknowledged all main sources of help.

Signed:

---

Date:

---

---

UNIVERSITY OF THE WITWATERSRAND

# *Abstract*

School of Electrical and Information Engineering

Master of Science

## **Automated Quantitative Discrimination of Parkinson's Disease Stages using Signal Processing and Machine Learning**

by Nabeel SEEDAT

Supervisor: PROF. VERED AHARONSON

Current clinical methods that determine Parkinson's Disease (PD) stages are mostly qualitative. The quantitative methods necessitate expensive equipment and/or cumbersome wearable devices, which limits their usability. This research presents a quantitative discrimination of PD stages using kinematic signals obtained from low-cost walker mounted sensors. Signal processing, machine learning and statistical methods are applied to extract and select features pertaining to PD patients' gait at the different stages of the disease. The research re-uses accelerometer, force sensors and distance encoder signals acquired in an experiment of a movement disorders clinic. The study consists of five key areas. (1) Signal pre-processing where signal denoising is applied and a novel footfall detection algorithm is proposed (2) Feature extraction which produces different categories of features. (3) Feature selection using both machine learning and statistical methods, (4) Classification and regression machine learning paradigms using clinical labels, where several machine learning methods are compared (5) Statistical analysis and modelling of the probability distributions associated with PD feature manifestation. The results indicate that the different PD stages can be discriminated using a Random Forest classifier with a 93% accuracy. The majority of the features most relevant to this discrimination belong to the information theoretic and statistical feature sub-classes. Confidence intervals analysis validated the class separability and a generalized pareto distribution was indicated as the best fit distribution for PD features. These findings may provide an insight into the disease progression. Additionally, a novel footfall detection algorithm, which has higher accuracy when compared to methods from literature, could be useful in other gait analysis studies. The research indicated the feasibility of signal processing and machine learning tools to accurately classify PD stages and implies the potential of affordable, simple walker-mounted sensors to aid medical practitioners in a quantitative assessment of PD stages.

## ***Dedication***

*I dedicate this dissertation to my family who have always supported and motivated me. In particular, my parents Mohamed Ameen and Yasmin Seedat for their love, support and encouragement through all my endeavours.*

# Acknowledgements

First and foremost I praise the Almighty for the blessing in this opportunity and for granting me the ability to carry out this work.

The support and assistance of the following persons in this research is acknowledged:

- Prof. Vered Aharonson, for her guidance over the course of this research. I am grateful for the endless encouragement, guidance and for the direction, critical comments and reviews towards this research. Her invaluable support coupled with her passion to achieve the best has helped me to grow and develop over the course of this work.
- Prof. Anton van Wyk, Mr Andre McDonald and Dr Gilad Yahalom for their most helpful advice and thoughtful review of the dissertation.
- The patient's and staff at the Sheba Medical Center who were involved as part of the initial data collection.
- My research group partner, Mr David Beder with whom I worked together on the data pre-processing.
- Everyone who took the time to read the work and provided feedback.
- The National Research Foundation (NRF) for the financial support towards this research. The financial support was in the form of a NRF Master's Block Grant (UID: 111755)

# Contents

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>i</b>
<b>Dedication</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>Abbreviations</b>	<b>xiii</b>
<b>Publications</b>	<b>xiv</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Background and Context . . . . .	1
1.2 Problem Statement . . . . .	3
1.3 Hypothesis . . . . .	3
1.4 Research Objectives . . . . .	4
1.5 Research Impact . . . . .	5
1.6 Organization of Dissertation . . . . .	6
<b>2 LITERATURE REVIEW</b>	<b>7</b>
2.1 Parkinson’s Disease . . . . .	7
2.2 Gait Assessment . . . . .	9
2.2.1 Human Gait . . . . .	9
2.2.2 Gait rationale for PD assessment . . . . .	9
2.3 PD Gait Assessment . . . . .	10
2.3.1 Clinical Assessment . . . . .	11
2.3.2 Laboratory systems . . . . .	12
2.3.3 Instrumented walkway system . . . . .	12
2.3.4 Wearables . . . . .	13
2.3.5 Instrumented Walker . . . . .	14
2.4 Analysis approach . . . . .	14
2.4.1 Signal Pre-processing . . . . .	15

2.4.2	Feature Derivation . . . . .	16
2.4.3	Feature Selection . . . . .	18
2.4.4	Machine Learning . . . . .	19
2.4.4.1	Artificial Neural Networks . . . . .	20
2.4.4.2	Support Vector Machines (SVMs) . . . . .	24
2.4.4.3	Naive Bayes . . . . .	25
2.4.4.4	Random Forests . . . . .	25
2.4.4.5	Boosted Forest . . . . .	27
2.5	Regression . . . . .	28
2.6	Statistical Analysis . . . . .	33
<b>3</b>	<b>METHODOLOGY</b>	<b>35</b>
3.1	Dataset . . . . .	35
3.1.1	Data . . . . .	35
3.1.2	Instrumentation . . . . .	36
3.1.3	Test Protocol . . . . .	37
3.1.4	Dataset Structure . . . . .	37
3.2	Processing Methodology . . . . .	38
3.2.1	High-Level Breakdown . . . . .	38
3.3	Data Integrity and Organisation Check . . . . .	38
3.4	Signal Pre-Processing . . . . .	39
3.4.1	Velocity Signal Production . . . . .	39
3.4.2	Interpolation . . . . .	39
3.4.3	Signals Denoising . . . . .	40
3.4.4	Accelerometer Rotation . . . . .	41
3.4.5	Segmentation of Movement Phases . . . . .	42
3.4.6	Footfall Detection . . . . .	44
3.5	Feature Derivation . . . . .	47
3.6	Feature Selection . . . . .	53
3.6.1	Principal Components Analysis (PCA) . . . . .	53
3.6.2	Selection of features based on relevance to clinical usage . . . . .	54
3.6.3	Feature Selection using ANOVA . . . . .	55
3.7	Feature Space Analysis . . . . .	55
3.7.1	Normal Distribution Testing . . . . .	56
3.7.2	Distribution Fitting . . . . .	56
3.8	Machine Learning . . . . .	57
3.8.1	Classification . . . . .	58
3.8.1.1	Random Forest Implementation & Testing . . . . .	58
3.8.2	Regression . . . . .	61
3.8.2.1	Stepwise Linear Regression . . . . .	61
3.8.2.2	Robust Linear Regression . . . . .	62
3.8.2.3	Support Vector Regression . . . . .	63
3.8.2.4	Random Forest Regression/ Regression Forests . . . . .	63
3.8.2.5	Boosted Trees Regression . . . . .	64

3.8.2.6	Artificial Neural Network (ANN) . . . . .	64
3.8.2.7	Model Testing . . . . .	64
3.9	Statistical Analysis . . . . .	66
3.9.1	Confidence Interval Calculation . . . . .	66
3.9.2	Feature Classification Performance Comparison . . . . .	67
<b>4</b>	<b>RESULTS</b>	<b>68</b>
4.1	Data Integrity and Organisation Check . . . . .	68
4.2	Signal Pre-processing . . . . .	69
4.2.1	Signal Denoising . . . . .	69
4.2.2	Segmentation of movement phases . . . . .	70
4.2.3	Footfall Detection Algorithm . . . . .	71
4.3	Feature Selection . . . . .	75
4.3.1	Feature Selection based on clinical usage . . . . .	75
4.3.2	PCA . . . . .	76
4.3.3	ANOVA feature selection . . . . .	77
4.4	Feature Space Analysis . . . . .	78
4.4.1	Normal Distribution Tests . . . . .	78
4.4.2	Distribution Fitting . . . . .	78
4.4.2.1	Entire Feature Space . . . . .	78
4.4.2.2	Distributions of H & Y Severity groups . . . . .	80
4.5	Machine Learning . . . . .	83
4.5.1	Classification: Random Forest . . . . .	84
4.5.1.1	Full Feature Vector . . . . .	84
4.5.1.2	Random Forest Selected Features . . . . .	85
4.5.1.3	Performance comparisons with the optimal number of trees . . . . .	86
4.5.2	Regression for H & Y Prediction . . . . .	90
4.5.2.1	Performance Metrics . . . . .	91
4.5.2.2	Relationship between feature values and H & Y scores . . . . .	92
4.6	Statistical Analysis using CI's and relation between CI's and selected features . . . . .	95
4.6.1	Confidence Intervals . . . . .	96
4.6.2	Feature Comparison . . . . .	99
<b>5</b>	<b>DISCUSSION AND CONCLUSIONS</b>	<b>102</b>
5.1	Key Findings . . . . .	102
5.1.1	Research Goal . . . . .	102
5.1.1.1	Extent of Discrimination between PD stages . . . . .	102
5.1.1.2	Best Feature Subset for Discrimination . . . . .	103
5.1.1.3	Finer H & Y scores using regression models . . . . .	105
5.1.1.4	Most significant features for PD discrimination . . . . .	107
5.1.2	Patterns in different PD severity groups . . . . .	108
5.1.2.1	Inter-Group variability and similarity patterns . . . . .	109



---

5.1.2.2	Statistical model: Most prevalent probability distribution . . . . .	110
5.1.2.3	Implication of the generalized pareto distribution . . . . .	111
5.1.2.4	CI Analysis and overlap validation . . . . .	112
5.1.3	Implications from signal pre-processing methods . . . . .	113
5.1.3.1	Accurate footfall detection using EMD and peak detection applied to the z-axis accelerometer . . . . .	113
5.1.4	Main Findings Summary . . . . .	114
5.1.5	Application of the findings . . . . .	114
5.1.6	Limitations and Future Work . . . . .	115
5.2	Conclusion . . . . .	115
<b>A</b>	<b>Ethics Approval</b>	<b>117</b>
<b>B</b>	<b>Additional Results</b>	<b>120</b>

## List of Figures

2.1	Comparison of a healthy vs PD affected brain from [1]. The red indicates higher levels of dopamine, in a brain with normal function.	7
2.2	Typical manifestation of a patient suffering from PD from [2] . . . .	8
2.3	Illustration of the human gait cycle and the stance and swing phases from [3] . . . . .	10
2.4	Illustration of the procedure of the TUG walking protocol from [3] .	12
2.5	Diagram of the PAMMII exo-body instrumented walker from [4] . .	15
2.6	Generalized structure of a 3-Layer ANN (adapted from [5] and [6]) .	21
2.7	Example of the criteria based decision tree split . . . . .	26
3.1	Labelled diagram of the PAMMII system from [4] . . . . .	37
3.2	Research methodology block diagram . . . . .	38
3.3	Graphical explanation of the different aspects of movement phase segmentation . . . . .	43
3.4	Ideal, "clean" accelerometer signal where the peaks (stars) represent footfalls from [7] . . . . .	45
4.1	Example of a velocity signal before (red line) and after the application of a moving average filter (blue line). . . . .	70
4.2	Movement Phases Algorithm Results where red lines are predicted movement periods and black lines are predicted stationary or turning periods . . . . .	70
4.3	Example of the First, Four IMFs which show a decreasing frequency	71
4.4	Example of the Footfall detection algorithm: The predicted footfalls are represented by circles and the validated video-captured footfalls ranges are plotted as "x"s when a red "x" is a start timestamp of the range and oranges "x" the end of the range . . . . .	72
4.5	Novel Accelerometer Algorithm: Percentage Accuracy of footfall detection with reconstruction from different IMF's. Average accuracy indicates the mean accuracy of the four trials for the specific IMF . . . . .	73
4.6	Novel Accelerometer Algorithm: Comparison of number of actual footfalls (segmented from the video) to the predicted ones: the blue bar represents the number of true measured footfalls, the green bar represents the number of predicted footfalls, the red bar represents the number of correctly predicted footfalls . . . . .	73
4.7	Force Sensor Algorithm: Percentage Accuracy of footfall detection when reconstructing from different IMFs. Average accuracy indicates the mean accuracy of the four trials for the specific IMF . . .	74

4.8	Force Sensor Algorithm: Comparison of number of actual footfalls (segmented from the video) to the predicted ones: the blue bar represents the number of true measured footfalls, the green bar represents the number of predicted footfalls, the red bar represents the number of correctly predicted footfalls . . . . .	74
4.9	The ANOVA selected features grouped by feature category for the 3m and 10m tests . . . . .	77
4.10	Number of features per each best fit (based on BIC) distribution for the 3m test . . . . .	79
4.11	Number of features per each best fit (based on BIC) distribution for the 10m test . . . . .	79
4.12	The probability distribution of the Mean Step Velocity feature. The distribution for each different PD stage is shown in different colours as per the figure legend. . . . .	82
4.13	The probability distribution of the Mean Step Length feature. The distribution for each different PD stage is shown in different colours as per the figure legend. . . . .	82
4.14	The probability distribution of the Mean Turn Velocity feature. The distribution for each different PD stage is shown in different colours as per the figure legend. . . . .	83
4.15	The probability distribution of the MI between z acceleration - RMS acceleration feature. The distribution for each different PD stage is shown in different colours as per the figure legend. . . . .	83
4.16	Discrimination error (in percentage) between PD stages, for 1-1000 trees, for the 3m test . . . . .	84
4.17	Discrimination error (in percentage) between PD stages, for 1-1000 trees, for the 10m test . . . . .	85
4.18	Random Forest Percentage Accuracy of discrimination between different PD stages for different feature selection subsets, for the 3m test . . . . .	87
4.19	Random Forest Percentage Accuracy of discrimination between different PD stages for different feature selection subsets, for the 10m test . . . . .	88
4.20	Comparison of the Random Forest Percentage Accuracy of discrimination between different PD stages for 3m and 10m Tests . . . . .	88
4.21	Execution time vs. 100 trees Classifier accuracy for the 5 feature subsets . . . . .	89
4.22	Execution time vs. 1-100 trees Classifier accuracy for the 5 feature subsets . . . . .	89
4.23	Confusion Matrix for the 3m test using ANOVA selected features . . . . .	90
4.24	Confusion Matrix for the 10m test using ANOVA selected features . . . . .	90
4.25	H & Y vs Mean Step Velocity. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound . . . . .	93
4.26	H & Y vs Mean Step Length. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound . . . . .	93

---

4.27	H & Y vs Mean Encoder Difference. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound . . . . .	94
4.28	H & Y vs Y accel Q1. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound . . . . .	94
4.29	H & Y vs Mutual Information z accel-RMS accel. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound . . . . .	95
4.30	CI Plot for Mean Step Velocity, which shows no group overlap of greater than 50% and either a 95% or 99% significance of difference between H & Y group pairs . . . . .	96
4.31	CI Plot of Mean Encoder Difference, which shows no group overlap of greater than 50% and either a 95% or 99% significance of difference between H & Y group pairs . . . . .	97
4.32	CI Plot of Mean Step Length, which shows greater than 50% overlap for H & Y 0 and 1 . . . . .	98
4.33	CI Plot of Turn Velocity, which shows greater than 50% overlap for H & Y 2.5 and 3 . . . . .	98
4.34	Percentage of overlapping features by category between the ANOVA features and CI Features . . . . .	100
4.35	Percentage of overlapping features by category between the ANOVA features and Random Forest Selected Features . . . . .	100

## List of Tables

3.1	Spatio-Temporal Features . . . . .	49
3.2	Signal Statistic Features . . . . .	50
3.3	Frequency Domain Features . . . . .	51
3.4	Information Theoretic Features . . . . .	52
3.5	Regression Performance Metrics . . . . .	65
4.1	Dataset Integrity Analysis comparing the number of subjects excluded and remaining for each H & Y group . . . . .	69
4.2	Predicted Footfalls Accuracy for the two Algorithms. Each column contains the accuracy (in percentage) obtained for the different IMF reconstructions. The four rows are the 4 walking tests. The fifth row is the average across tests for each IMF. . . . .	75
4.3	The 3 Best Distributions for the different H & Y Groups in the 3m test. The percentage of features out of the total number of features is depicted below the distribution name . . . . .	80
4.4	The 3 Best Distributions for the different H & Y Groups in the 10m test. The percentage of features out of the total number of features is depicted below the distribution name . . . . .	81
4.5	Random Forest Selected Features which are the features used by the random forest in discriminating between different PD stages . . . . .	86
4.6	Accuracy (Acc) and Execution Time Comparison using the optimal 100 tree Random Forest . . . . .	87
4.7	Normalized Regression Performance Metrics for the 3m test, where the columns are the different regression models and the rows are the different performance metrics . . . . .	91
4.8	Normalized Regression Performance Metrics for the 10m Test, where the columns are the different regression models and the rows are the different performance metrics . . . . .	92
4.9	Percentage of features overlapping between the significant CI features and the ANOVA and random forest selected features . . . . .	99
4.10	Common features between the features subsets which allowed for the highest accuracy of discrimination between PD stages . . . . .	101
B.1	ANOVA Significant Features - 3m Test . . . . .	121
B.2	ANOVA Significant Features - 10m Test . . . . .	122
B.3	Features Similar between the significant CI features and the ANOVA features . . . . .	123
B.4	Features Similar between the significant CI features and the random forest selected features . . . . .	124

## Abbreviations

<b>AD</b>	<b>A</b> nderson <b>D</b> arling
<b>ANN</b>	<b>A</b> rtificial <b>N</b> eural <b>N</b> etwork
<b>ANOVA</b>	<b>A</b> nalysis <b>O</b> f <b>V</b> ariance
<b>CI</b>	<b>C</b> onfidence <b>I</b> nterval
<b>EMD</b>	<b>E</b> mpirical <b>M</b> ode <b>D</b> ecomposition
<b>FoG</b>	<b>F</b> reezing <b>o</b> f <b>G</b> ait
<b>H &amp; Y</b>	<b>H</b> oehn and <b>Y</b> har
<b>IMF</b>	<b>I</b> ntrinsic <b>M</b> ode <b>F</b> unction
<b>KS</b>	<b>K</b> olmogorov <b>S</b> mirnov
<b>PAMM</b>	<b>P</b> ersonal <b>A</b> id for <b>M</b> obility and <b>H</b> ealth <b>M</b> onitoring
<b>PD</b>	<b>P</b> arkinson's <b>D</b> isease
<b>PCA</b>	<b>P</b> rincipal <b>C</b> omponent <b>A</b> nalysis
<b>PC</b>	<b>P</b> rincipal <b>C</b> omponent
<b>RF</b>	<b>R</b> andom <b>F</b> orest
<b>SVM</b>	<b>S</b> upport <b>V</b> ector <b>M</b> achine
<b>TUG</b>	<b>T</b> imed <b>U</b> p and <b>G</b> o
<b>UPDRS</b>	<b>U</b> nified <b>P</b> arkinson's <b>D</b> isease <b>R</b> ating <b>S</b> cale

# Publications

## Overlapping Peer-reviewed papers directly related to this Dissertation chapters

---

- [3,4,5] **N.Seedat**, D.Beder, V.Aharonson and S.Dubowsky, "A Comparison of Footfall Detection Algorithms from Walker Mounted Sensors Data", *IEEE 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pp. 1-4, 2018.

# CHAPTER 1

## INTRODUCTION

### 1.1 Background and Context

Parkinson's Disease (PD) is one of the most common neuro-degenerative diseases [8–10]. It is caused by the progressive loss of dopaminergic neurons in the substantia nigra of the brain [11]. Studies indicate that globally, around 7-10 million people are affected by PD [10, 12]. A major issue for PD patients is the motor impairments that reduce their quality of life by restricting functional independence, as well as being a major cause of both morbidity and mortality, as a result of falls [13].

The diagnosis of PD is largely clinical and qualitative and is based on the presence of clinical signs such as resting tremor, bradykinesia, rigidity and postural instability [4]. Two scales are commonly used for PD diagnosis and severity evaluation based on these clinical signs. (1) The Unified Parkinson's Disease Rating Scale (UPDRS) includes an assessment of symptoms, as well as historical data obtained from the patient or caregiver [11]. (2) Hoehn and Yahr (H & Y) scale has 5 stages which reflect disease severity and the subsequent impairment to patient gait and postural stability [14]. While these scales are validated and have a standardized assessment procedure, the evaluation of the symptoms that provide the scores in the scales are performed by clinicians and therefore entails a subjective component in the scores given [15]. This means that patients could either be mis-diagnosed or un-diagnosed and possibly mistreated. Moreover, since expert neurologists and/or other healthcare professionals are needed to carry out these assessments, they are



not accessible for many patients, who cost- and distance-wise cannot reach the clinics. This need for professional labour also burdens the healthcare system. Hence, automated and quantifiable assessments may assist clinicians as a diagnostic tool, as well as for monitoring disease progression and therapeutic efficacy.

Whilst PD manifests with many physical and mental symptoms, many of these symptoms pertain to gait. Gait is the locomotive movement of human limbs and characterises the cyclical manner in which a person walks [16, 17]. Motor disorders such as PD, result in impaired gait, which significantly limits a person's functionality and independence and can cause instability and fall hazards. Thus, the assessment of gait is an important component in the management and long-term monitoring of patients suffering from PD and other movement disorders [18, 19].

As a result, research and technology aims to provide a quantitative evaluation of PD patient's gait-based symptoms, in an automated, professional labour-saving manner [11]. A typical quantitative method is the Timed-Up-and-Go (TUG) test, where a patient sitting on a chair is requested to stand up, walk straight forward, turn around, walk back and sit on the chair, and the task completion time is logged. This method, however, only provides a single, time-related measure and furthermore, still necessitates human observation and manual handling of a stopwatch or button, which limits its accuracy [4]. Quantitative sensor-based methods have therefore been suggested to solve these limitations through assessment of PD gait using kinematic measurements.

Examples of sensor-based applications include devices strapped onto the subject's body [11] or fitted to the subject's shoes [20]. These applications typically use accelerometers to acquire kinematic patterns of the walking [13]. A drawback of both method types is that they are complex and time-consuming to use for patients or caregivers. Additionally, these devices are often cumbersome and uncomfortable to wear, thereby harming the user experience, especially for motor impaired persons [21]. Moreover, in more severe cases of PD, the patient would

require a walking aid such as a walker or cane, which further limits the usability of wearables for PD progression monitoring [4, 22].

A solution to these limitations is to mount the sensors on a walking aid. The Personal Aid for Mobility and Health Monitoring II (PAMMII) is an example to this concept. This exo-body walking frame is fitted with sensors to monitor patient gait and support walking concurrently. The device has shown preliminary results of high discrimination power between PD and control patients [4].

This research aims to assess the potential of kinematic signals acquired from the PAMMII system to provide a quantitative discrimination between the different PD stages, as pertains to gait. This five-class, gait features-based discrimination of PD stages, seeks to enhance the scope of previous studies which discriminated between PD patients and controls. The low-cost nature of the system entails noisy sensors that are not attached directly to the subject's body, as in the wearables paradigm, but rather are mounted on a walker. These properties present a signal processing challenge which needs to be addressed especially when a multiclass discrimination is desired. If proven accurate and efficient, the signal processing and machine learning methods used have the potential to allow doctors, technicians or nurses to use this quantitative and objective system for PD patients monitoring, providing patient-centred care with minimal cost and effort.

## **1.2 Problem Statement**

To what extent can kinematic signals acquired from an instrumented walker sensors provide an automated, quantitative and reliable discrimination between PD progression stages?

## **1.3 Hypothesis**

Signal processing and machine learning techniques applied to kinematic signals acquired from low-cost and noisy sensors mounted on an exo-body walker can

provide a quantitative measure of PD progression and discriminate between clinical sub-types at various stages of the disease.

## 1.4 Research Objectives

The objective of this research is to discriminate between different PD stages, as pertains to gait, using kinematic signals and applying signal processing and machine learning techniques.

The following sub-objectives will feed into the broader objective:

1. Investigate the applicability of different signal pre-processing techniques for human motion signals acquired by low-cost and noisy sensors. Signal pre-processing is a necessary compensation when dealing with such sensors to ensure reliable feature derivation.
2. Develop a method for footfall detection (a basic element of gait) from signals acquired by exo-body, walker mounted sensors.
3. Extract, evaluate and compare different categories of gait features. This serves to evaluate which categories of features are most relevant to the discrimination between the different PD stages.
4. Implement and evaluate different feature selection methods for the machine learning algorithms.
5. Develop a quantitative understanding of features relevant to the discrimination between different PD stages, pertaining to gait.
6. Explore the viability of a finer PD progression scale, pertaining to gait.
7. Derive a statistical model to fit PD gait features and assess how the parameters of this statistical model change for the different disease stages.
8. Apply statistical testing to validate and gain further quantitative insight into the similarity and differences between the different PD stages.

## 1.5 Research Impact

This research aims to yield a reliable discrimination between PD stages that can enable an automated, user-friendly PD assessment which can be performed within everyday activities of patients and improve patient care. The simple usage and low cost of this assessment may offer a potential of affordable PD monitoring to developing countries and under-resourced healthcare settings.

The information derived from the features that provide this discrimination can integrate with clinical knowledge of PD and offer a quantitative insight into different patterns of the gait involved in the disease progression. Moreover, the PAMMII device can be used as a walking aid during gait evaluation, enabling patients with severe gait impairment to participate in studies of gait and posture without the risk of fall.

The engineering impact is three-fold. Firstly, the noisy and low-cost kinematic sensors and their placement on an exo-body device present a signal processing challenge. This is notably the case when these settings are compared to previous studies, where good quality, expensive sensors were attached to the patient's body, thus producing more reliable readings, while reducing affordability and user comfort. The signal processing and data mining tools and algorithms used to address this challenge could be generalized and used in other kinematic signals research and applications. Secondly, the insight gained into which features provide the best discrimination between PD stages may yield valuable information regarding the limitations and advantages of both the instrumentation and the categories of the extracted features. Thirdly, statistical analysis and modelling of these features at different PD stages may provide an insight into the features and patterns of gait disorders. Both feature types analysis and feature modelling in the context of PD stage discrimination were not performed in previous studies.

The research provides quantitative measures for PD stages, which may contribute to broader PD research. The multi-disciplinary, broad PD research strives for better understanding of the disease symptoms and its progressions and for searches for

better ways to evaluate existing and new treatments' efficacy. The research forms part of a research collaboration between the University of the Witwatersrand, the Massachusetts Institute of Technology (MIT- USA) and Halmstad University (Sweden) concerning patient-centred monitoring technologies for PD.

## **1.6 Organization of Dissertation**

The rest of the dissertation is organized as follows.

Chapter 2 reviews studies related to the thesis research. A review is presented with regard to PD and current clinical and quantitative diagnostic approaches. Thereafter, a review is presented of signal processing, feature extraction, feature selection and machine learning methods widely used in literature for PD vs control discrimination, as well as for other gait assessment applications.

Chapter 3 presents the methodology and technical details used in this research. The chapter describes the dataset used as well as the signal pre-processing, feature derivation, feature selection, machine learning, and statistical analysis techniques employed.

Chapter 4 conveys the key results of each processing sub-section as outlined by the methodology in Chapter 3.

Chapter 5 discusses and analyses the key findings of the research. Additionally, applications of the findings, their limitations and future recommendations, as well as concluding remarks are conveyed.

# CHAPTER 2

## LITERATURE REVIEW

### 2.1 Parkinson's Disease

Parkinson's Disease (PD) is the second most common neuro-degenerative disease after Alzheimer's disease [16, 17], affecting approximately 7-10 million people globally according to the Parkinson's Disease Foundation [12] and is most prevalent in patients over the age of 50 [23].

PD is a progressive degenerative disorder which affects the nervous system [16, 23]. It is manifested by the degeneration of dopamine producing neurons in the substantia nigra of the brain [16, 17]. Dopamine is the neurotransmitter responsible for controlling smooth and co-ordinated muscle function [17] and the reduced dopamine impacts the basal ganglia (responsible for motor control) which can be seen in Figure 2.1.

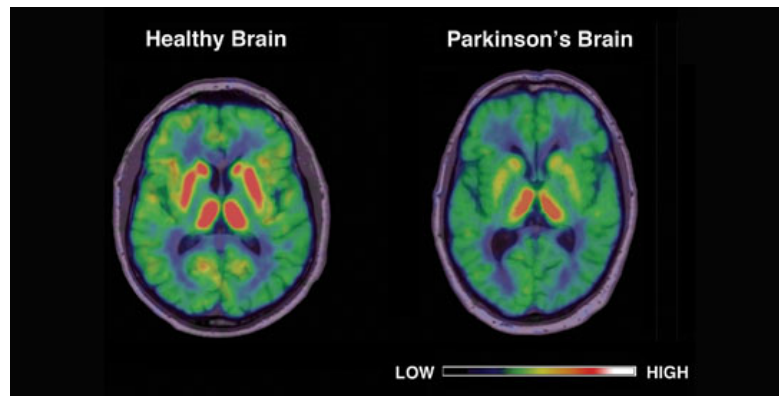


FIGURE 2.1: Comparison of a healthy vs PD affected brain from [1]. The red indicates higher levels of dopamine, in a brain with normal function.

The reduced levels of dopamine in PD patients, when compared to non-PD patients as shown in Figure 2.1 can affect the motor performance of the PD patient [16]. Moreover, the motor disturbance can compromise autonomy and fluidity of patient movement [16]. In PD this is manifested by impaired balance/postural instability, reduced co-ordination (ataxia), slowed voluntary movements (bradykinesia), rigidity, resting tremor which can spread into the forearm, elbow and upper arm [16, 17, 24].

Additionally, some PD patients experience Freezing of Gait (FOG), which is a period of hesitation to move despite the intention. The typical PD patient will display a stooped posture, slow gait, reduced arm swings and shortened steps [25] and has appearance similar to Figure 2.2.

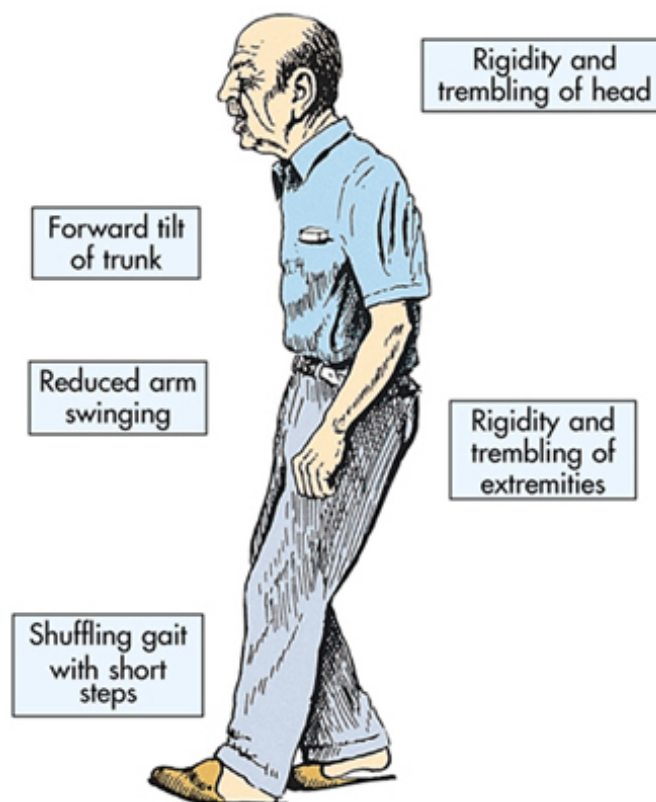


FIGURE 2.2: Typical manifestation of a patient suffering from PD from [2]

Whilst, PD is not a fatal disease in and of itself, the complications are debilitating and thus, impact the patient's quality of life. The clinical features and disease course vary among different individuals [26]. The heterogeneity means that assessment of the disease is a challenging problem, as each patient can have different combinations at different severities of manifestation [15].

Among the aforementioned disease symptoms, many are symptoms that pertain to gait. Although many other physical and mental symptoms hinder the patients' quality of life, reduced mobility is among the issues that most limit the patient's functionality and independence.

## **2.2 Gait Assessment**

### **2.2.1 Human Gait**

Gait is the cyclical manner in which a person walks [16, 17]. The gait cycle shown in Figure 2.3 is the time between 2 successive and repetitive events of the same footfall (ipsilateral movement). Gait is defined from initial contact where the foot touches the ground, ending with the initial contact of the other foot (called terminal contact) [16]. i.e. a singular gait cycle consists of left and right steps. Typically the gait cycle is divided into two phases namely the stance and swing phases [16, 17]. The stance phase is at the beginning of the gait cycle. It refers to when one or both feet are on the ground and is the basis of initial and terminal contact [16]. The swing phase is between initial and terminal contact, whereby the reference foot is not on the ground and is moving forward [16]. In some cases, the patient cannot lift their foot and thus, the swing phase is then defined as forward motion of the foot [16].

### **2.2.2 Gait rationale for PD assessment**

Gait is an important parameter of human movement that is impaired by motor diseases such as PD. The assessment of gait is key to evaluate the quality of



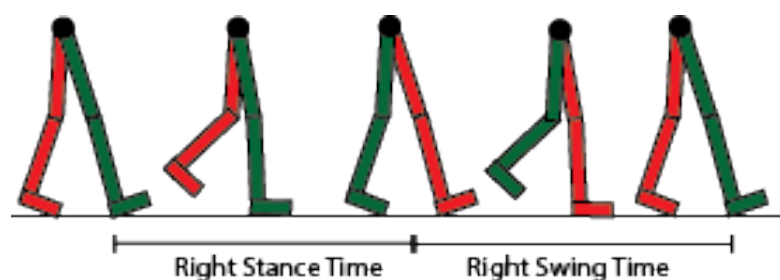


FIGURE 2.3: Illustration of the human gait cycle and the stance and swing phases from [3]

movement and can assist in patient rehabilitation and management [19]. Gait can be impacted by many factors including age, trauma and disease, which can affect the locomotor system controlled by the central nervous system (CNS) [19].

This research deals with the quantitative analysis of gait deterioration related to PD stage severity. Previous works presented the following findings of gait decline.

- Stride length and speed are reduced with PD [27].
- PD patients increase frequency of steps to compensate for the reduced stride length [28].
- Increase in stride time due to decreased mobility [19].
- Stride time variability increases with PD, due to PD affecting stability and rhythmic movement [29].

### 2.3 PD Gait Assessment

PD gait is assessed clinically by healthcare practitioners. In recent years, other quantitative solutions have been proposed including: laboratory systems, instrumented walkway systems, wearables and instrumented walkers. This section will outline these existing solutions for PD assessment.

### 2.3.1 Clinical Assessment

The two main clinical PD assessment scales are the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr (H & Y) scale. UPDRS assesses motor and gait symptoms such as rigidity, tremor, stability and patient posture, as well as non-motor symptoms of PD [30]. It includes 4 scales of assessment [15]. Firstly, the non-motor symptoms such as behaviour are evaluated. Secondly, the clinician performs a mobility assessment. This involves both clinical questions and a historical evaluation of the disease. Thirdly, a motor examination of the levels of performance is carried out. Fourthly, a historical assessment of prior motor complications is assessed. Each of the scales is rated between 0 and 4. A rating of 4 indicating severe PD and 0 indicating non-PD. This scale involves patient history, hence introducing subjectivity and variability. The variability can be introduced both by the doctor's assessment, as well as patient recollection.

The H & Y scale reflects disease severity by means of involvement of one or two sides of the body and the severity of postural instability. The H & Y scale has a PD 'score' between 0-5 assigned by a clinician. The scale is as follows [15]:

- 0- Healthy, non-PD patient.
- 1- Unilateral disease. The patient exhibits minor or no disability.
- 2- Bilateral disease without impairment of balance.
- 3- Bilateral disease. The patient exhibits mild to moderate disability and instability.
- 4- Severe disease, however can still walk unassisted.
- 5- Bed ridden or confined to a wheelchair without assistance.

In addition to the UPDRS and H & Y scores, the Timed-Up-and-Go (TUG) protocol is often used by clinicians to measure balance, fall risk and general mobility

[31, 32]. The aim of the TUG test as seen in Figure 2.4, is to measure how long the subject takes to complete the walking protocol. The time is then related to the stage of disease [31]. The test involves the patient standing up from a chair, walking 3m, turning around and walking 3m back to the starting point, thereafter sitting down [31, 32]. The test is regarded as a gold standard for early assessment.



FIGURE 2.4: Illustration of the procedure of the TUG walking protocol from [3]

### 2.3.2 Laboratory systems

Optical motion capture is used in laboratory-based gait assessment, both in terms of biomechanics and kinetics, for diagnostic, sports and rehabilitation applications [23]. Optical measurements are taken of the subjects while they walk under controlled conditions, using external body markers [16, 23]. With the aid of cameras, the real-time motion of the subject is assessed to provide highly accurate, quantitative gait models. The Vicon and CODA capture systems are the two devices largely used in research settings to compute spatio-temporal parameters based on the position of the markers [16, 23]. Despite the high accuracy, these devices are limited by the high expense, complex setup, as well as needing a dedicated testing environment.

### 2.3.3 Instrumented walkway system

Instrumented walkway systems have also been explored as a means of quantitative PD gait assessment [33]. These systems aim to mimic the TUG protocol, whilst simultaneously performing quantitative measurements. Both the Gaitrite and Zeno

walkway systems, measure the pressure exerted by the patients foot as they walk [16, 33]. The premise is that foot pressure varies through the gait cycle and hence can be used to quantify gait events, as well as patient stability [16]. Whilst, these systems do have an advantage of high accuracy and lower cost, the downside is the length constraint and dedicated environment.

### **2.3.4 Wearables**

Wearable devices have been proposed to address the limitations of the aforementioned systems. These wearable devices are used for ambulatory gait monitoring. Typically accelerometers, gyroscopes, force sensors, magnetometers etc are used for this purpose [16]. The caveat to these approaches is that gait is not measured directly but as a function of accelerations, angular velocities and forces. Hence, signal processing is needed to obtain clinical gait information and spatio-temporal parameters of gait. Three wearable devices appear in literature [33]. Firstly, the Ambulatory Parkinson's Disease Monitoring System (APDM) [33], is a watch-based wearable which includes an accelerometer, gyroscope and magnetometer. The system carries out signal processing to obtain velocity, cadence, stride length, stride time, as well as trunk sway. APDM has been used in a TUG protocol and has been proven to be sensitive in discriminating PD vs non-PD. This being correlated with the UPDRS score, with a p-value of 0.006.

The second device is the Physilog device [33] which uses gyroscopes attached to the body in order to obtain spatio-temporal gait parameters such as stride length, time, and speed, as well as measurements of tremor. The algorithms have shown 96% sensitivity for PD.

The third device is the Axivity (AX3) [33] which is a three-axis accelerometer mounted on the patient for continuous monitoring, in order to assess PD in free-living conditions. The same spatio-temporal parameters were extracted along with accelerometer magnitudes. The study concluded that PD patients at risk of falling

had a less variable walking pattern to non-fallers and that generally PD patients took more steps over shorter increments than non-PD patients.

Whilst wearables seemingly solve important problems, according to [4], a significant drawback of both systems is the lack of time to attach the sensors to the patient within busy hospital settings. Additionally, these devices are often uncomfortable to wear thereby harming patient usability. Linked to this, in more severe cases of PD the patient may need a walking aid, and thus the aforementioned analysis tools are inappropriate monitoring solutions [4].

### **2.3.5 Instrumented Walker**

Instrumented walkers aim to solve the aforementioned limitations by mounting the sensors onto a walking aid. This allows patient gait to be measured using the mounted kinematic sensors, whilst supporting the patient's walking. The challenge of such an exo-body device, is that determining gait parameters is challenging compared to systems where the sensors are attached to and move with the subject's body.

One such exo-body walker is the Personal Aid for Mobility and Health Monitoring II (PAMMII), shown in Figure 2.5. It is a commercially available, non-motorized, exo-body walker, fitted with low-cost sensors to capture kinematic signals as the subject uses the walker [4]. The PAMMII device has successfully been shown to have the capability to differentiate between PD and control subjects with sensitivity and specificity scores of 91% and 95% respectively for the TUG walking protocol [4].

## **2.4 Analysis approach**

Typically in literature the analysis of any signal processing and data mining research consists of the following stages post-data acquisition: Signal Pre-Processing, Feature Derivation, Feature Selection and Classification/Machine Learning.

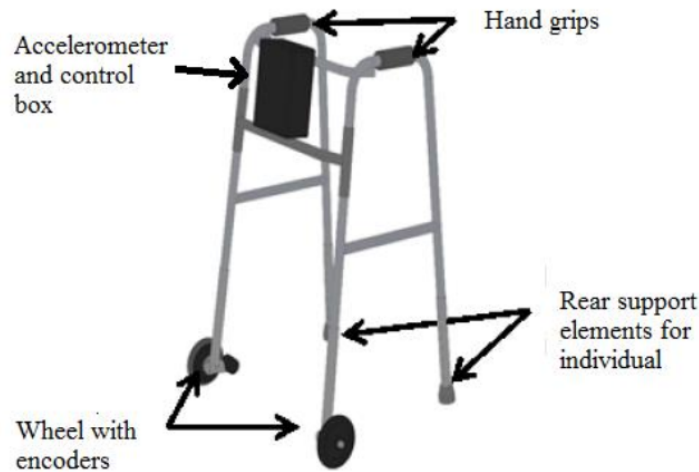


FIGURE 2.5: Diagram of the PAMMII exo-body instrumented walker from [4]

### 2.4.1 Signal Pre-processing

Signal pre-processing is an important part in signal analysis, especially when low-cost systems and noisy sensors are concerned. The reason is that the measurement errors introduced by low-cost sensors, unless compensated for, can impact the features derived and thereby on overall performance [4]. Among the vast literature on these issues, this review selected several that were used in similar applications to this research.

Signal pre-processing of noisy signals is often done in the digital domain. Commonly, low-pass filters are applied to attenuate high frequency components [4]. Moreover, signal pre-processing can involve more traditional frequency domain techniques such as Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters. These types of filters were explored for PD gait assessment by [11] and [34]. In particular, [34] filtered raw accelerometer data for gait extraction, using a high-pass IIR filter with a cut-off of 1 Hz to mitigate body movement noise and a low-pass IIR filter with a 3 Hz cut-off based on the frequency associated with bradykinesia and dyskinesia. The second common technique uses the wavelet transform and pre-processes the signal through setting of a wavelet co-efficient. The wavelet-based denoising method has been used by [4] and [35]. Additionally, 4th order, low-pass, butterworth filters have been used to mitigate

high frequency noise in the accelerometer data with a cut-off frequency in the range of 2-3 Hz [4, 36].

It is also important to apply rotational correction when working with accelerometer signals, in order to align the accelerometer signal with the direction of movement [4, 37, 38]. Another important reason for rotation is that an accelerometer measures Earth's gravitational acceleration in addition to acceleration of motion. In order, to isolate motion-based acceleration, the gravity component must be removed. Thus, once the accelerometer measurement has been rotated, [38] subtracted the z-component of acceleration (which is gravity) of  $9.81m/s^2$  in order that only motion measurements remain.

Two additional findings from literature relevant to the study with regard to signal pre-processing are presented. Firstly, [39] used an instrumented rollator walker to determine footfalls. It was found that heel contact begins when the force sensor value on the same side as the stride increases, whilst the opposite force sensor value decreases. The peaks of the difference between the force sensor values were proposed as being representative of footfalls [39].

Secondly, [40] successfully used Empirical Mode Decomposition (EMD) for pre-processing of environmental non-stationary signals. The accelerometer signal was decomposed into oscillatory components, called intrinsic mode functions (IMFs). The IMFs were set adaptively to the actual signal, in contrast with analytical methods such as wavelet transforms. The IMFs were then analysed and the signal was reconstructed. It was found that the first IMF contains high frequency data associated with noise and can be discarded [40].

## 2.4.2 Feature Derivation

Features are instances or representations constructed from the original data/signal and are important to achieve membership to a particular class [35]. In the case of disease and progression identification certain features might provide insight into

classifying levels of disease [41]. Moreover, the classification performance depends heavily on the features derived.

A variety of features are typically used in literature for gait assessment. The three most common are stride time, length and velocity [4, 38, 39, 42, 43]. However, there are numerous other features which have proven successful in similar studies. These are illustrated below:

- Accelerometer Signal Entropy [34, 43, 44]
- Accelerometer and force sensor Signal RMS [4, 34, 43, 44]
- Cross correlation between the accelerometer axes [4, 34]
- Cadence [38]
- Kurtosis [43]
- Turn time and turn velocity [4, 38]
- Harmonic ratio [43, 44]
- Centroid of PSD [4, 43, 44]
- Accelerometer signal bandwidth [43]
- Mean force [4]
- Standard deviation of step velocity, step time and step length [4, 43, 44]
- Statistical features such as the Inter-Quartile Range (IQR), median, range etc [45, 46]
- Information theory related features such as entropy, mutual information, correlations etc [45, 46]



### 2.4.3 Feature Selection

Feature derivation typically produces a large number of features, which results in a high dimensional feature space [47]. This high dimensional feature space may contain features not relevant for classification or may have redundant features [35]. Feature selection methods produce a low dimensional feature set from a high dimensional feature set [47]. This is called dimensionality reduction. It is also shown by [41], that dimensionality reduction yields better classifier performance. Moreover, [44] showed that feature selection can make the classifier faster, due to fewer features being analysed. Moreover, larger dimensions of features means that more data is needed to train the model.

Feature selection falls into one of two categories: projection methods and criteria-based feature selection.

The most common projection method is called Principal Component Analysis (PCA). It has been used successfully for PD feature selection by [4], [48], [39], [49], [37] and [47]. PCA orthogonally transforms the input parameter/features in order to maximize the variability of the output data [39]. PCA maximizes variance of the feature space by mapping the feature space as a linear combination of the original features and projecting it to produce a new, smaller set of derived variables [4]. These variables are called principal components (PCs).

PCA involves the computation of the covariance matrix given by Equation 2.1.

$$\mathbf{X} = \mathbf{X}^T \mathbf{X} \quad (2.1)$$

Where:

$\mathbf{X}$  = matrix of features

The eigenvalues and eigenvectors of the covariance matrix are then computed. The PCs are obtained by projecting the features matrix through the matrix computation with the matrix  $\mathbf{W}$ , which represents the eigenvectors corresponding to the largest eigenvalues. This is given by Equation 2.2.

$$PC = \mathbf{X}^T \mathbf{W} \quad (2.2)$$

Where:

$\mathbf{X}$  = matrix of features

$\mathbf{W}$  = matrix of eigenvectors corresponding to the largest eigenvalues

The second technique of dimensionality reduction is criteria-based feature selection. According to [44], this method is preferred to projection methods as instead of creating new variables, the features remain the same which makes it easier for clinicians to analyse. There are three types of feature selection methods namely: filter, wrapper or embedded techniques [44].

Filter techniques consider the correlation between features when selecting features. However, since they ignore interaction with the classifier they are prone to lower accuracy results [44].

Wrapper techniques specify a variety of feature subsets based on clinical relevance, correlation etc. The specific classifier is then trained and evaluated with the different feature subsets and the accuracy is the evaluative method [44].

Lastly, embedded techniques are those that select optimal features within the construction of the classifier. However, the choice of classifier is limited as only certain classifiers can utilize this technique: decision trees, random forest, naïve bayes and SVMs [44].

#### **2.4.4 Machine Learning**

Once features have been selected the aim of the machine learning algorithm is to find patterns in the data. These can be framed as either a classification or

regression problems. Typically, these algorithms of classification or regression are divided into two types: supervised and un-supervised learning [41].

Supervised Learning [41]: Each object has a pre-assigned label and the label is used in the training stage to teach the algorithm to recognise a particular class of object. The aim is to map labelled features to output classes

Unsupervised Learning [41]: data is unlabelled and the algorithm aims to learn structure of the data. i.e. what characteristics make objects similar within groups and different across groups. The algorithm learns patterns on its own.

Supervised learning, however is the most prevalent in literature for PD classification. Typically, it is done for PD vs non-PD classification. [50] surveyed the most used data mining techniques in literature for classifying gait based PD vs non-PD. It was found that the following 5 frameworks were used extensively: Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests, Naïve Bayes and K-Means Clustering. The first four are supervised learning paradigms, whilst K-Means Clustering is an unsupervised learning paradigm.

#### **2.4.4.1 Artificial Neural Networks**

The following works used shallow ANN, multi-layer perceptron (MLP) architectures with a single hidden layer to classify between PD vs non-PD subjects.

- An ANN, MLP with one hidden layer using Levenberg Marquardt backpropagation showed a 95.63% accuracy using the four highest PCs [51].
- A similar ANN, MLP with one hidden layer using gradient descent backpropagation for extracted spatio-temporal parameters showed a 98% classification accuracy [48].
- A gradient descent backpropagation MLP, achieved a 91% accuracy on the training data set and 80% on validation data set [52].

The base element of a neural network is a neuron that receives inputs as features or from other preceding neurons [5, 53, 54]. The input to the neuron is processed by

an activation function, to produce the output [5, 55]. The strength of connections between the output of one neuron and the input of another neuron is specified by the weight of the connection [5, 56]. Neurons are arranged in layers to form a neural network [5, 55, 56]. Typically there is an input and output layer of the neural network, with hidden layers between the input and output as shown in Figure 2.6.

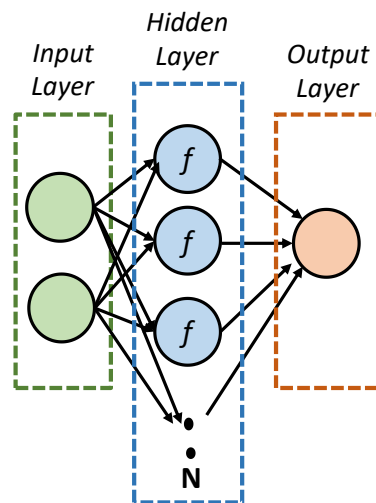


FIGURE 2.6: Generalized structure of a 3-Layer ANN (adapted from [5] and [6])

There exist two types of multi-layer ANN architectures: feedforward and recurrent architectures. Feedforward ANNs require the data to be passed from the input to the output, whereas a recurrent ANN also includes feedback to the layers [55]. Whilst, [57] states that recurrent ANNs are more powerful at solving systems with non-linearities, there is also the drawback of a much longer training time and the need for more training data. Thus, recurrent architectures are not used especially for extendibility and speed. A feedforward network, would thus consist of an input layer, a number of hidden layers and output layer. It is suggested by both [55] and [54] that the activation function should be a tan-sigmoidal activation function ( $f(x) = \tanh(x)$ ) to deal with non-linearity.

Supervised learning makes use of a feedforward and backpropagation process. The input data is propagated layer by layer through the network and the data is multiplied by the neuron weights to calculate the outputs [56].

The predicted and desired outputs are then compared and the cost function is computed, which is usually the Mean Squared Error (MSE), given by Equation 2.3.

$$MSE = \frac{1}{2}(y_d(k) - y(k))^2 \quad (2.3)$$

Where:

$y_d(k)$  = Desired output

$y(k)$  = Actual output

Since, the hidden layer neurons contribute to the error, the weights are re-adjusted using a backpropagation algorithm until the MSE is below a certain value [58]. Typically, gradient descent backpropagation or Levenberg Marquardt backpropagation is used. However, according to [59] these techniques are prone to result in overfitting due to the high epochs of training. Moreover, it results in increased bias and variance and consequently, poor generalization. It has been proposed that regularization be used as a method to avoid this, with early stopping and Bayesian Regularization being the two common techniques [59]. ANN's that make use of Bayesian Regularisation are called Bayesian Regularized Artificial Neural Networks (BRANNs) [60].

BRANNs perform this role by minimizing the linear combination of the squared errors and the weights [59, 60]. To understand this it would be useful to highlight the differences in the performance functions.

Ordinary ANNs calculate optimal weights by minimizing the error function which compares the distance between the true and predicted values given by Equation 2.4 [60].

$$f(x) = E_D(D|w, M) = \frac{1}{N} \sum_{i=0}^n (pred - actual)^2 \quad (2.4)$$

Where:

- $E_D$  = Mean sum of square errors
- $D$  = data with input-output pairs
- $w$  = weights
- $M$  = neural network under question

Bayesian Regularization is based on Bayes Theorem hence a probabilistic aspect is added to the training, which involves a probability distribution of the weights [59, 60]. The difference given by Equation 2.5 to the above function is that a  $\beta$  value is added coupled with a function which aims to penalize the network for large weights [59, 60]. This penalizing ensures that the fit is smoother, rather than fitting to the noise of the dataset.

$$f(x) = \beta E_D(D|w, M) + \alpha E_W(w|M) = \frac{1}{N} \sum_{i=0}^n (pred - actual)^2 \quad (2.5)$$

Where:

- larger  $\beta$  = better error minimization
- larger  $\alpha$  = smoother network, by prioritizing weight minimization

The use of Bayes Rule changes the manner in which the weights are then updated [60]. The prior belief (posterior distribution) of the weights is then updated by Bayes Rule as shown in Equation 2.6.

$$P(w|D, \alpha, \beta, M) = \frac{P(D|w, \beta, M)P(w|\alpha, M)}{P(D|\alpha, \beta, M)} \quad (2.6)$$

The aim when updating the weights is to produce optimal weights that optimize the posterior probability of the weights.

#### 2.4.4.2 Support Vector Machines (SVMs)

SVMs are another widely used machine learning algorithm prevalent in literature. Aside from PD classification, SVMs have been successful in many other classification systems from speaker classification, to face detection to gene classification [34]. SVMs generate non-linear decision boundaries by mapping the feature space into a higher dimensional space using kernels [34]. The mapping allows the features to be linearly separable. The reason is that the linear hyperplanes in the higher dimensional space, maximize the separation between classes.

The kernel functions produce a function  $f(x)$  which maps input-output relationships [61]. The most used is epsilon-insensitive ( $\epsilon$ ) SVM. The goal of  $\epsilon$ -SVM is to find the function  $f(x)$ , whose output deviates from the true Y/observed Y ( $Y_o$ ), by no more than epsilon for each training point [61]. Thus, the goal is to minimize the error function.

The mapping is shown in Equation 2.7 below [61]:

$$\langle \phi(x), \phi(y) \rangle = K(x, y) \quad (2.7)$$

Where:

$\phi$  = mapping function

$K$  = Kernel

The most common kernels are Gaussian and polynomial kernels [61] given by Equations 2.8 and 2.9 respectively.

$$K(x, y) = \exp\left(-\frac{1}{2\sigma^2}\|x - y\|^2\right) \quad (2.8)$$

$$K(x, y) = (x^T y + 1)^d \quad (2.9)$$

- An SVM used spatio-temporal parameters to classify between PD and non-PD, with a 100% accuracy [48].

- An SVM was used with a different set of features to classify between PD and healthy patient with a 76% accuracy [62].

Whilst, SVMs are typically used as a binary classifier, they can be extended to multi-class classification [63].

The Random Forest and Naïve Bayes algorithms have also been used albeit not as prevalent as the aforementioned two, however the relevant literature is as follows.

#### **2.4.4.3 Naive Bayes**

The Naive Bayes approach assumes that the classification problem can be modelled probabilistically and that the data can be described or approximated by an a priori probability distribution [28]. This method was used to classify spatio-temporal parameters, achieving a 95% accuracy in classifying between PD patient and non-PD patients.

Although the aforementioned methods have proven successful in general classification tasks, as well as in PD related research, they lack in an important aspect: These methods cannot provide an insight into which features from their input was relevant to the classification, and to what extent. This information is important for the clinical interpretation of the classes.

#### **2.4.4.4 Random Forests**

Random Forests introduced by Leo Breiman [64] is a decision tree based ensemble which aims to produce a prediction function  $f(x)$  in order to predict  $Y$ , where  $x$  is a vector of random variables or features [64, 65]. The random forest ensemble constructs the predictor function  $f(x)$  as a function of the weak learners (individual decision trees)[65].

The root node of these weak learners or decision tree is exposed to the entire feature space, whilst the end nodes/terminal leaves represent the decisions/predictors.



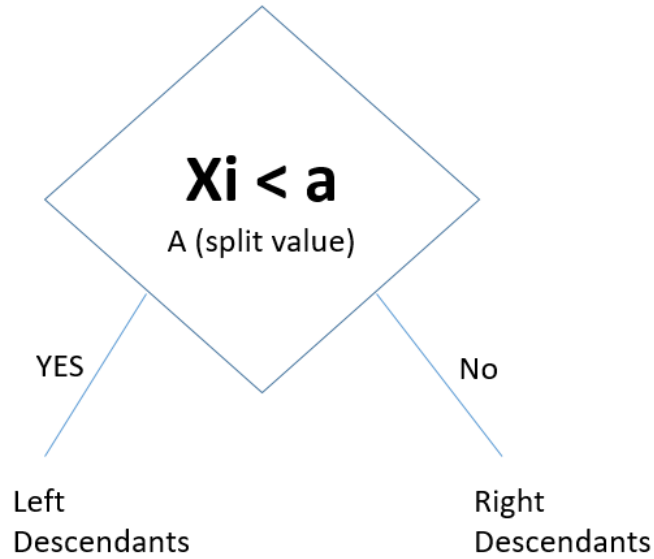


FIGURE 2.7: Example of the criteria based decision tree split

A sequence of splits is carried out in between, with predictors smaller than the criterion going left whilst those larger going right [64]. This can be seen in Figure 2.7[65]. The split considers every possible split of every feature and chooses the “best” based on the Gini Index given by Equation 2.10.

$$Gini(feature) = 1 - \sum_{i=1}^N (p_i)^2 \quad (2.10)$$

Where:

$N$  = Number of classes

$p_i$  = relative frequency of class  $i$

The Gini Index measures purity (i.e. homogeneity), thereby maximizing which features are most representative.

These weak learners are combined to form a strong learner. In classification, the function  $f(x)$  is the most frequently predicted class [65]. This is called majority voting and is given by Equation 2.11 below.

$$f(x) = \operatorname{argmax}\left(\sum_{i=1}^N h_i(x)\right) \quad (2.11)$$

Where:

$N$  = Number of trees

$h_i$  =  $i$ th decision tree in the forest

When a random forest is used in a regression problem, instead of majority voting of the decision trees which is used in classification, the outputs of each "weak learner" is averaged as given by Equation 2.12.

$$f(x) = \frac{1}{N} \sum_{i=1}^N h_i(x) \quad (2.12)$$

Where:

$N$  = Number of trees

$h(x)$  = decision tree

$f(x)$  = Predictive function

In an application of a Random Forest classifier, [66] achieved a 99% accuracy in discriminating PD from non-PD subjects. The benefit of Random Forest is that variables which contribute to decisions can be obtained unlike in ANN's or SVM's [64, 65].

#### 2.4.4.5 Boosted Forest

The usage of decision tree ensembles can be done in a boosting process rather than the random forest method which uses a bagging approach. This would result in a Boosted Forest [67, 68].

Boosted Forests produce a model that is additive rather than a majority voting model and the predictive function is given by Equation 2.13 (i.e. it is a sum of the weak learners) [67, 68].

$$g(x) = f_1(x) + f_2(x) \dots + f_n(x) \quad (2.13)$$

Where:

$N$  = Number of trees

$g(x)$  = predictive function

$f_n(x)$  = specific decision tree (i.e. weak learner)

Unlike random forests, where each tree is grown independently, rather in the boosting method it makes use of gradient boosting, which is similar to gradient descent optimization in that the training is iterative [67, 68]. Effectively what this does is fit a new decision tree to the residual of the previous iteration.

This iterative nature is explained by Equation 2.14

$$g_t(x) = \sum_{i=0}^{t-1} f_i(x) \quad (2.14)$$

Where:

$g_t(x)$  = function after t iterations

$f_i(x)$  = ith trained decision tree

The aim with gradient boosting is to minimize a loss function [67, 68], by iteratively moving  $g_t$  such that  $g_{t+1} = g_t + \eta f_t$ .

## 2.5 Regression

Regression makes use of the same models as classification. One added aspect is measuring the performance of regression models. In classification, the performance is measured by accuracy of classification. Whereas in regression, performance metrics are typically used. In particular, the comparison of the error of the regression model, as well as the similarity measure between the actual and predicted values. The following 15 performance metrics are used to literature to compare regression models [69–71]. The parameters in Equations 2.15-2.29:  $Y_o$  is the observed/true H & Y scores and  $Y_p$  is the predicted H & Y scores.

(1) Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_o - Y_p)^2 \quad (2.15)$$

- Perfect:  $MSE = 0$
- Measures mean square prediction error
- Outliers heavily impact it
- Weights larger errors due to the squaring

(2) Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_o - Y_p| \quad (2.16)$$

- Perfect:  $MAE = 0$
- Measures mean prediction error
- Outliers heavily impact it

(3) Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_o - Y_p|}{Y_o} \quad (2.17)$$

- Perfect:  $MAPE = 0$
- Measures error as a relative percentage
- Outliers heavily impact it

(4) Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\sum \frac{(Y_p - Y_o)^2}{n}} \quad (2.18)$$

- Perfect: RMSE = 0
- Measures error as a relative percentage
- Heavily weights larger errors due to the squaring

## (5) Sum of Squares Error (SSE)

$$SSE = \sum_{i=1}^n (Y_o - Y_p)^2 \quad (2.19)$$

- Perfect: SSE = 0
- Measures variation of predictions
- Larger SSE means greater overall prediction variation
- Tends to weight the larger errors due to the squaring

## (6) Cosine Similarity (CS)

$$CS = \frac{Y_o \cdot Y_p}{\|Y_o\| \|Y_p\|} \quad (2.20)$$

- Perfect: CS = 1
- Measures similarity between 2 vectors
- Only measures similarity of orientation
- Used for cluster cohesion

## (7) Modified Index of Agreement (IA)

$$IA = 1 - \frac{\sum_{i=1}^n |Y_p - Y_o|}{\sum_{i=1}^n (|Y_p - \bar{Y}_o| + |Y_o - \bar{Y}_o|)^2} \quad (2.21)$$

- Perfect: IA = 1
- Measures model prediction error
- Assesses the agreement between the prediction and actual values
- Modified IA formula, overcomes sensitivity to extreme values

## (8) Mutual Information (MI)

$$MI = \sum P(Y_o, Y_p) \cdot \log\left(\frac{P(Y_o, Y_p)}{P(Y_o)P(Y_p)}\right) \quad (2.22)$$

- Greater implies less uncertainty about one random variable given the other
- Measures how much one random variable can tell us about the other
- Effectively a measure of dependency between 2 random variables

## (9) R squared (Rsq)

$$Rsq = \frac{\sum(Y_p - \bar{Y}_p)^2}{\sum(Y_o - \bar{Y}_o)^2} \quad (2.23)$$

- Perfect: Rsq = 1
- Measures the amount of variance explained by the model (Goodness of fit)
- Assumed for linear models
- Not good to compare for linear vs nonlinear models

## (10) Spearman's Correlation

$$Correlation = \frac{COV(Y_p, Y_o)}{\sigma_{Y_p} \sigma_{Y_o}} \quad (2.24)$$

- Perfect: Correlation = 1

- Measures degree of association/relation between 2 vectors
- The statistical significance of the correlation is important

(11) Fractional Bias (FB)

$$FB = \frac{\sum_i (Y_o - Y_p)}{0.5 \sum_i (Y_o + Y_p)} \quad (2.25)$$

- Perfect:  $FB = 0$
- Measures mean bias and indicates systematic error [69]
- Is influenced by high outlier values

(12) Geometric Mean Bias (GMB)

$$GMB = e^{(\overline{\ln(Y_o)} - \overline{\ln(Y_p)})} \quad (2.26)$$

- Perfect:  $GMB = 1$
- Measures mean bias and indicates systematic error
- Is influenced by low outlier values

(13) Normalized Mean Square Error (NMSE)

$$NMSE = \frac{\overline{(Y_o - Y_p)^2}}{\overline{Y_o Y_p}} \quad (2.27)$$

- Perfect:  $NMSE=0$
- Measures scatter and indicates both systematic and random error
- Is influenced by high outlier values

(14) Geometric Variance (GV)

$$GV = e^{(\overline{\ln(Y_o)} - \overline{\ln(Y_p)})^2} \quad (2.28)$$

- Perfect:  $GV=0$
- Measures scatter and indicates both systematic and random error
- Is influenced by low outlier values

(15) Fraction of predictions within a factor of two of observations (FAC2)

$$\text{Fraction satisfies } 0.5 \leq \frac{Y_p}{Y_o} \leq 2.0 \quad (2.29)$$

- Perfect:  $FAC2=1$
- Measures scatter and indicates both systematic and random error
- Is influenced by high outlier values

## 2.6 Statistical Analysis

ANOVA tests are widely used in literature when comparing different means [72]. The ANOVA p-value states the significance of the difference between two or more groups. The one-way ANOVA test has been used in PD literature to evaluate whether there were significant differences between the mean values of PD and control subjects [17]. Despite the widespread usage, a limitation of the ANOVA p-value is that it does not provide an indication of the magnitude of these differences [72]. Moreover, the ANOVA does not indicate overlap or the differences for individual groups.

In solving this conundrum, [72] proposed the usage of confidence intervals (CIs). CIs are an estimated range of values, which have a high probability of covering the true population value [73]. i.e. they provide a probabilistic indication of the lower and upper limits. It was proposed by [72] that 50% or less overlap between CI's, implies a difference between the group means which is statistically significant. Whilst, [73] state that when measuring overlap: less than 50% corresponds to a 95% statistical significance of the difference ( $p \text{ value} \leq 0.05$ ), whilst no overlap



of the CI corresponds to a 99% statistical significance of the difference (p value  $\leq 0.01$ ). It is important to note that unlike typical ANOVA analysis, for within-subject or within-class CI analysis does not make assumptions that the errors in each sample are independent.

# CHAPTER 3

## METHODOLOGY

In order to answer the research question, an adequate dataset was acquired, and signal processing and data mining were applied to derive features from and perform classification on the data. Statistical methods were also applied as further validation. This chapter will describe the data and tools that were implemented.

### 3.1 Dataset

#### 3.1.1 Data

A quantitative answer to the question of the extent to which kinematic signals acquired from an instrumented walker's sensors provide an automated, quantitative and reliable discrimination between PD progression stages, necessitated an appropriate dataset, which includes subjects at different PD stages and an adequate number of patients as a whole.

To satisfy this, the study re-used an existing dataset of 71 patients and 21 controls that was acquired in a movement disorders clinic, in the Sheba Medical Center, Israel. This research is the first analysis of this extensive PD dataset.

The data was acquired during the patient's routine follow-up visit to the clinic. The clinical assessments as expounded in the literature review in Chapter 2 were conducted as per the normal clinic follow-up visit, wherein PD H & Y severity scores were assigned by the clinic's neurologists. Additionally, an unaided TUG

walking test was performed by 48 of the patients and the test times were separately recorded. Thereafter, the patients who agreed to participate in the study and signed an informed consent, performed an instrumented walking protocol (as outlined in Section 3.1.3) using the PAMMII walker, as an addition to the clinic visit. Control subjects were recruited from the medical staff and patients' caregivers who volunteered to participate in the experiment and likewise signed an informed consent.

Ethics approval for re-use of data was obtained from the University of the Witwatersrand, Human Research Ethics Committee. The clearance number is M180202 and clearance certificate is in Appendix A. Additionally, the experiment to collect the data was performed in accordance with the Helsinki Declaration, as well as under the hospital's ethical guidelines (Ethics number: 3036-16-SMC).

### **3.1.2 Instrumentation**

The PAMMII device described in Section 2.3.5 was the instrumentation used to capture the raw data/signals. The device as shown in Figure 3.1 had two force sensors (Tekscan A201) underneath the walker's hand grips to measure a patient's grip force with a range of 0-440 N. Two digital encoders used to measure walker motion were attached to the walker's front wheels using tooth belts. The encoder signals were processed to obtain the position and velocity of the walker. Lastly, a tri-axial accelerometer (ADXL335) was contained in the control box of the walker. The accelerometer scaled the reading to gravitational acceleration of  $9.81m/s^2$  and had a range of  $\pm 16$  g with resolution in increments of 0.1 g. An embedded microcontroller (Arduino Nano V3) executed the commands and control functionality, as well as acquired the data at a sampling rate of 21.5 Hz. The data was written to an SD card in the form of a CSV file. The data consisted of the signals outlined in Section 3.1.4

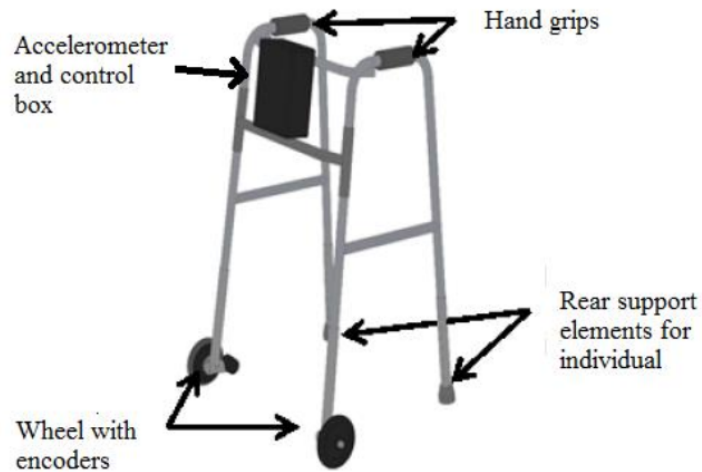


FIGURE 3.1: Labelled diagram of the PAMMII system from [4]

### 3.1.3 Test Protocol

The data was obtained using an instrumented TUG walking protocol, which the patients performed while holding and walking with PAMMII: The patient gets up from a chair then walks along a 3m straight line path, thereafter turning around and walking 3m back to the starting point. This test was then repeated with a 10m walking protocol in order to have a longer distance, which was indicated in previous studies with PAMMII to yield additional information in the signal processing [4]. The raw data described in Section 3.1.4 below was acquired by the device throughout the patient’s walking tests.

### 3.1.4 Dataset Structure

The dataset included a total of 92 subjects (71 patients and 21 controls), with 2 CSV files for each subject. One file for the 3m TUG and the other file for the 10m walking test. Each file contained the raw signals/data obtained from the walker’s sensors: three acceleration signals obtained from the tri-axial accelerometer; two force signals obtained from the right and left handles of the walker and two distance signals obtained from the encoders on the walker’s wheels. These signals were used in the analysis described in Section 3.2.

The patients' H & Y scores, age, gender, clinical comments and the unaided TUG times (for 52% of the patients) were stored in a separate file. The patient H & Y scores included H & Y 0 (Controls), 1, 2, 2.5, 3 and 4. Patient's with H & Y scores of 5 were not included as these patients are bed ridden or confined to a wheelchair without assistance and thus, would be unable to perform the walking tests.

## 3.2 Processing Methodology

### 3.2.1 High-Level Breakdown

Figure 3.2 illustrates the high-level block diagram of the methodology used in the data analysis. The processes involved in each block are described in the following sections.

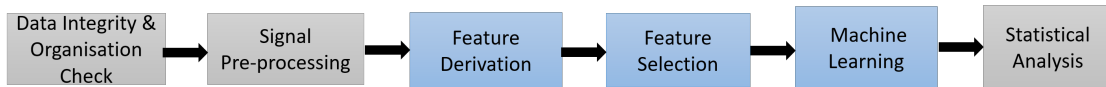


FIGURE 3.2: Research methodology block diagram

## 3.3 Data Integrity and Organisation Check

Prior to the analysis, the integrity and completeness of the dataset needed to be checked and then organized and prepared for the analysis. The following steps were carried out:

1. Check that each patient has conducted an accurate 3m and 10m test. If a patient does not have both tests the analysis is not run for that patient.
2. Check that each patient has recorded clinical scores such as H & Y scores.
3. Check for incomplete or missing data within each file. These data points were then interpolated as explained in Section 3.4.2.

## 3.4 Signal Pre-Processing

A crucial issue in dealing with the low-cost and noisy sensor data is the low quality of the signals. Extensive signal pre-processing was employed to compensate for the errors introduced by the low-cost and noisy sensors thereby ensuring reliable feature derivation and system performance. The signal pre-processing included six stages: velocity signal production, data interpolation, signal denoising, accelerometer rotation, segmentation of movement phases and footfall detection.

### 3.4.1 Velocity Signal Production

Velocity signal was used in previous studies [4], with similar sensors and hence the production of such signals means that it can be used for comparison to these studies. Additionally, although velocity can be inferred from the displacement and acceleration signals, this pre-processing stage simplifies the processing steps to follow.

Prior to producing the velocity signals, the left and right encoder values were averaged to determine the overall distance travelled by the walker. The velocity signals were produced through numeric differentiation using the distance encoder data, as well as the time stamps. The distance encoder ticks were converted to meters to ensure that the units of velocity was m/s. The meters per tick conversion ratio was per the datasheet of the encoders.

### 3.4.2 Interpolation

Based on preliminary analysis of the signals obtained from the PAMMII system, it was noted there was a sample of missing or invalid data, approximately every 40-50 samples. Consequently, cubic splines interpolation was carried out across the signal vector to mitigate this issue.

### 3.4.3 Signals Denoising

The use of low-cost sensors led to the acquisition of signals that contain measurement inaccuracies coupled with environmental and hardware noise. The errors and/or imprecision of the measurements acquired by these sensors must thus be compensated, so that the features extracted from the signals are reliable. Hence, rigorous signal pre-processing was applied to achieve this compensation. The pre-processing consisted of filtering to reduce signal noise.

A moving average filter was employed as a first process to de-noise all the acquired signals in the database. The filter choice was motivated by evidence from wide literature dealing with kinematic and other noisy signals [4, 11, 34–36], which found it to be an ideal filter for reducing random noise (which characterize signals from low-cost sensors), whilst retaining a sharp step response [74].

The output from the moving average filter is given by the difference equation in Equation 3.1

The moving average filter is low-pass Finite Impulse Response (FIR) filter. The design of the filter employed consisted of a window size/sample size of five and was zero-phase lag, in order to mitigate phase distortion that resulted from filtering. These parameters were obtained from experimental analysis of the measured signals. The output from the moving average filter is given by the difference equation in Equation 3.1.

$$y[n] = \frac{1}{5}(x[n - 2] + x[n - 1] + x[n] + x[n + 1] + x[n + 2]) \quad (3.1)$$

Secondly, Empirical Mode Decomposition (EMD) was also applied for the purposes of adaptive filtering. It was applied to further denoise the accelerometer signals for better footfall detection and will be discussed in Section 3.4.6.

### 3.4.4 Accelerometer Rotation

As discussed in Section 2.4.1, when working with 3D accelerometer signals, it is important to align the accelerometer signals with the direction of movement - in this case, antero-posterior (parallel and perpendicular) to patient movement. A rotation matrix method was used to rotate the signal [4, 37, 38]. Specifically, the Rodrigues' Rotation method was used, which rotates a vector in space using an axis and angle. The Rodrigues' Rotation algorithm is given by Equation 3.2.

$$V_{rot} = V_{init} \cos \theta + (k \times V_{init}) \sin \theta + k(k \cdot V_{init})(1 - \cos \theta) \quad (3.2)$$

Where:

$V_{rot}$  = Rotation Matrix

$V_{init}$  = Initial Reference Co-Ordinates

$\theta$  = Angle of rotation

$k$  = Unit vector of the rotation axis given by Equation 3.3

$$k = \frac{a \times b}{|a \times b|} \quad (3.3)$$

The accelerometer signal at the beginning of the test, when the patient was stationary (i.e. as the device is turned on), was used as the reference. The reference (at device turn-on time) had gravity as the only component of acceleration. In this case, the rotation axis and angle were such that the  $z$ -axis, at the turn-on time, was aligned with  $[x, y, z] = [0, 0, -1]$ . Therefore, the rotation matrix was formulated against this reference and all data points of the accelerometer signals were rotated accordingly.

After rotation, gravity was subtracted from the signal, leaving only relevant kinematic measurements of actual patient movement in the  $x$ ,  $y$  and  $z$  axis of acceleration.



### 3.4.5 Segmentation of Movement Phases

The movement phase segments were determined using the velocity signals. When plotted, the velocity signal offers an intuitive view of where a movement phase starts (signal amplitude increases from zero to a non-zero value) and stops (signal amplitude returns to zero).

Based on the testing protocol, all patients walk in a straight line for 3m or 10m, turn around and walk 3m or 10m back to the starting point. As a result, all patients have two phases of straight line walking, separated by a turning phase. That being said, the straight line walking phase can differ from patient to patient. For example, while some patients might have uninterrupted straight line walking, other patients may stop during this phase for short or long periods. These periods could be a manifestation of the disease (e.g. freezing of gait (FOG)) or due to physical limitations of the patient (e.g. fatigue or balance problem).

Hence, this trend was considered when estimating and segmenting movement phases. The periods of negligible movement (stops) or turning were determined using peak detection as shown in Figure 3.3. Peak detection was applied to the velocity signal to determine the extrema (maxima and minima). The maxima was the peak velocity achieved during straight line movement. The minima referred to the slowest velocity which was representative of a turn or stop. The peaks were then used to determine the tolerance band which was the threshold between straight line movement and stop/turn phases. Experimentally, it was found that a tolerance band of 12.5% below the mean of the extrema provided a reliable estimate of the turn and stop periods. The tolerance band is indicated by the blue line in Figure 3.3.

Subsequently, the straight line movement phase was defined as periods of non-negligible velocity (i.e. anything above the tolerance band shown in Figure 3.3). This would include constant velocity, as well as periods of increasing velocity and periods of decreasing velocity signal.

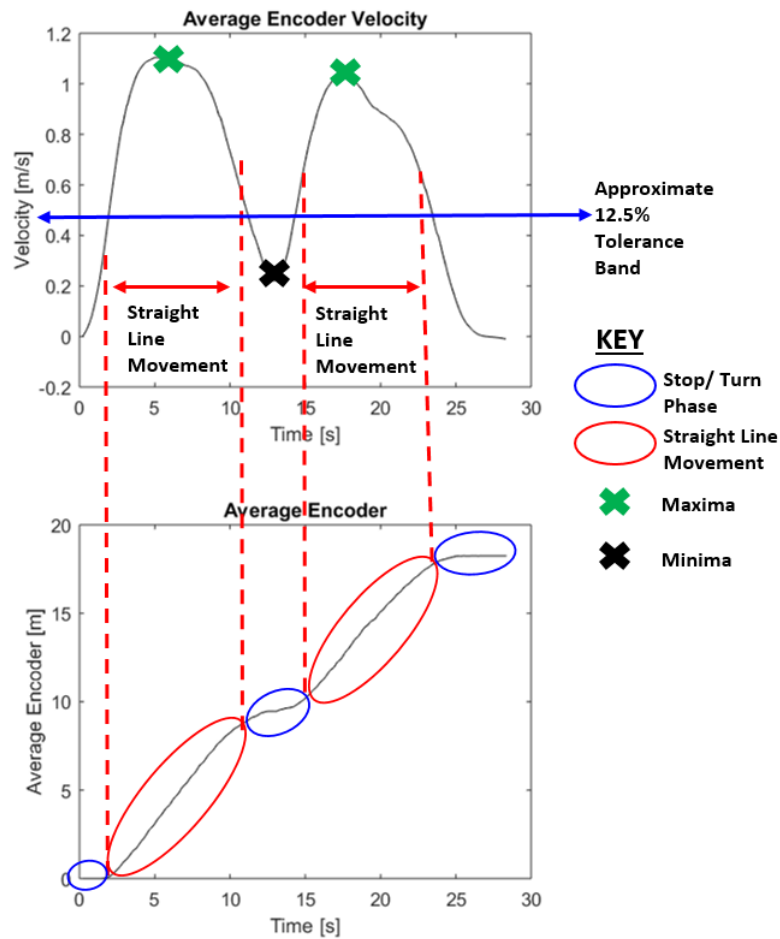


FIGURE 3.3: Graphical explanation of the different aspects of movement phase segmentation

Based on this definition of straight line movement and making use of the tolerance band, the non-movement periods (timestamps) were then indexed in a movement vector as binary 0, whilst the straight line movement periods (timestamps) were then indexed in the vector as binary 1.

Anomaly mitigation was performed to identify single instances of negligible movement introduced by measurement error. These singular instances were not FOG episodes, but rather data points mis-classified as stationary. A sliding window was used to check the binary movement vector ("1" represented straight line movement and "0" negligible movement or turn). Anomalies were characterized based on experimental assessment as a single non-movement index (binary 0), with five straight line movement indices (binary 1) on either side of the binary 0. If an anomaly was detected, the data point was replaced with the mean of the window.

This was to ensure that it was representative of the current state of the patient's movement.

### 3.4.6 Footfall Detection

Gait is composed of strides which are defined as the period from initial contact of one foot to terminal contact of the other foot [16]. This basic element of gait has been derived in many gait assessment systems, as these footfalls were crucial building blocks for the derivation of reliable gait based features.

In the case of exo-body walker-mounted sensors, an accurate detection of footfalls is very challenging compared to systems where the sensors are attached to and move with the subject's body.

Therefore, an algorithm for footfalls detection was developed and implemented for the low-cost sensors mounted on the exo-body walker. Two algorithms were investigated making use of both the force sensors and accelerometer.

- Algorithm 1: Replicated the footfall detection algorithm proposed by [39]. The algorithm made use of force sensors on the handles of a walker to calculate footfalls. The authors proposed that by calculating the difference in force between the right and left handles, thereafter applying peak detection to this difference, the peaks would be representative of footfalls. However, it must be noted that this algorithm was not validated and this assessment was the first quantitative validation of the predictive accuracy.
- Algorithm 2: A novel footfall detection algorithm was developed and made use of the  $z$ -axis accelerometer signal, Empirical Mode Decomposition (EMD), signal reconstruction and peak detection.

The literature surveyed indicated that the gait of the patient was substantially manifested in accelerometer signals. In an ideal, clean signal as illustrated in Figure 3.4, the peaks of the accelerometer signal represent footfalls [7]. Two issues

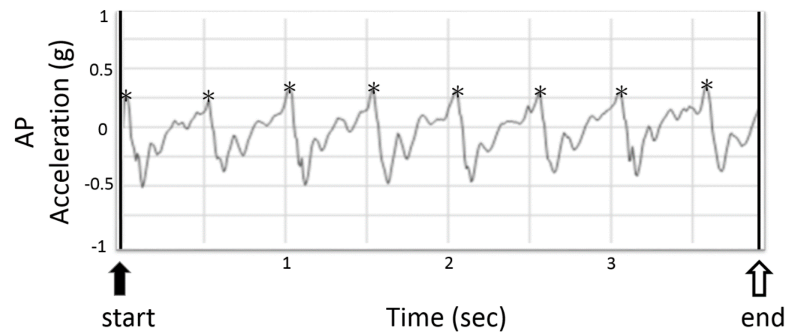


FIGURE 3.4: Ideal, "clean" accelerometer signal where the peaks (stars) represent footfalls from [7]

were addressed in developing this new algorithm. Firstly, the need to accommodate the noisy accelerometer signal. Secondly, since previous works provided no ground truth validation for their algorithm, the accuracy of the predicted footfalls in this study was established using ground truth of footfalls from video data.

Footfall detection is inherently based on peak detection. The signal therefore should be meticulously cleaned in order to avoid false peaks which stem from noise. Therefore after the initial denoising outlined in the pre-processing section, a second rigorous denoising was performed using EMD. EMD as was discussed in Section 2.4.1 is a locally adaptive algorithm which was indicated as suitable for analysis of non-stationary signals [75].

The algorithm iteratively decomposes the signal into a finite number of oscillatory components called Intrinsic Mode Functions (IMFs). These components are sorted such that each IMF consisted of lower frequency oscillations than its preceding IMF. The PAMMII device's signals are such non-stationary time-series signals. Hence, this thesis explored the applicability of the algorithm to the development of a novel footfall detection algorithm. Moreover, explore the benefit of adaptive usage of Intrinsic Mode Functions (IMFs) when compared to other footfall detection algorithms.

The basis of EMD is that a signal is a superposition of fast and slow oscillatory components [75]. The IMFs were extracted using a sifting algorithm. The novel EMD based algorithm applied is detailed in Algorithm 1.

---

**Algorithm 1:** EMD Sifting Algorithm [76]

---

- 1: Obtain maxima and minima of the signal.
  - 2: Construct the upper and lower envelope using cubic splines separately.
  - 3: Calculate the mean value for the upper and lower envelopes.
  - 4: Subtract this mean from the original signal to obtain a new signal residual signal  $r(t)$ .
  - 5: Repeat steps 1) to 4) until the conditions of being an IMF are satisfied i.e. the number of extrema or zero-crossings are equal to each other or differ by one, as well as the mean value of the upper and lower envelope is equal to zero.
  - 6: When the conditions of an IMF from Step 5 are satisfied extract this value as the  $n$ th IMF.
  - 7: Take the remaining signal after extraction, as the new signal. Repeat steps 1) to 6) on the new signal.
  - 8: Stop calculating IMFs once the stopping criterion of standard deviation of 0.3 is obtained.
- 

The signal decomposed by EMD can be represented mathematically by Equation 3.4 [77].

$$x(t) = \sum_{i=1}^n c_i(t) + r(t) \quad (3.4)$$

Where:

- $x(t)$  = Original signal
- $c(t)$  = IMF decomposition
- $r(t)$  = Residual signal

After decomposing the signal using EMD, the  $z$ -axis accelerometer signal was then reconstructed from various IMFs and peak detection applied to the reconstructed

signal. This was done to assess the accuracy of footfall detection for signal reconstructions from different IMFs. Thereafter, peak detection was carried out on the reconstructed signal, whereby the peaks corresponded to predicted footfalls. Different IMF levels were compared to determine the number of IMF's which give the best accuracy in footfall detection.

The performance benchmarking compared the predicted footfalls for both algorithms versus the actual footfalls which were manually timestamped using the video data.

### 3.5 Feature Derivation

Feature derivation is the transformation of the raw signals into a set of features, which could provide distinctive properties and insight into the data. The feature derivation carried out is presented below.

Accurate footfall detection described in Section 3.4.6, provided a basis for many of the derived features. This section lists and describes the features that were derived in this study. The features listed were an amalgamation of features derived in multiple previous gait studies. Some of the features are specific to PD, whilst other features pertain to gait assessment of other diseases (e.g. fibromyalgia, COPD etc) or normal gait [78]. However, the derivation of all the features were adjusted to suit the signals of the walker-mounted sensor system, where sensor type, quality and placement were different from previous studies.

The tables below present different features which represent properties of the data. The features are divided into the following categories: (1) Table 3.1: Spatio-temporal features, (2) Table 3.2: Statistical Features, (3) Table 3.3: Frequency Domain Features and (4) Table 3.4: Information Theory based Features.

The features translated from studies on other diseases are marked with an asterisk (\*). The novel features specific to this study are in bold.

The acceleration features were derived separately for the  $x$ ,  $y$  and  $z$  axis of the accelerometer. These Cartesian axes of the accelerometer were also converted to spherical co-ordinates: azimuth angle, elevation (polar) angle and radius distance, which were useful for some features extraction.

TABLE 3.1: Spatio-Temporal Features

<b>Feature</b>	<b>Description</b>
Number of steps (NS)	Number of steps taken during the walking period
Mean Step Time	Average time taken between each step in seconds
Mean Step Length	Average length of each step in meters
Mean Step Velocity	Average velocity of each step in m/s
Mean Acceleration	Average acceleration in the walking phase
Standard Deviation Step Time	Standard Deviation of the within subject step time divided by mean step time
Standard Deviation Step Length	Standard Deviation of the within subject step length divided by mean step length
Standard Deviation Step Velocity	Standard Deviation of the within subject step velocity divided by mean step velocity
Standard Deviation Acceleration	Standard Deviation of the within subject acceleration divided by mean acceleration
Total TUG time	Time taken for the TUG test (including standing up and turn time)
Total Straight Line Walk time (TWT)	Time taken for the straight line walking only in the test
Total Turn time	Time taken for turn period
Cadence	Ratio of the number of steps taken to the walking time (NS:TWT)



TABLE 3.2: Signal Statistic Features

<b>Feature</b>	<b>Description</b>
Skewness (x,y,z,rms, spherical)	Extent of the accelerometer signals asymmetry
Kurtosis (x,y,z,rms, spherical)	Extent which accelerometer signals distribution lies to the left of the mean
Mean Accelerometer (x,y,z,rms, spherical)	Average measurement of the specific accelerometer in the walking phase
Variance Accelerometer (x,y,z,rms, spherical)	Variance of the specific accelerometer in the walking phase
Accelerometer Q1 (x,y,z,rms, spherical)	First Quartile of the accelerometer signal
Accelerometer Q3 (x,y,z,rms, spherical)	Third Quartile of the accelerometer signal
Accelerometer Median (x,y,z,rms, spherical)	Median of the accelerometer signal
Accelerometer Mode (x,y,z,rms, spherical)	Mode of the accelerometer signal
Accelerometer IQR (x,y,z,rms, spherical)	Inter-Quartile Range of the accelerometer signal
Accelerometer Range (x,y,z,rms, spherical)	Range of the accelerometer signal
Accelerometer MSE (x,y,z,rms, spherical)	Mean Squared Energy of the accelerometer signal
Max and Min Force Difference	Maximum and Minimum difference between the right and left force
Mean Force Difference	Mean difference between the right and left force
Standard Deviation Force Difference	Standard deviation of difference between the right and left force
Mean $F_{Right}$	Mean right force measurement
Mean $F_{Left}$	Mean left force measurement
Mean Turn Velocity	Average velocity of the turn phase in m/s
Standard Deviation Turn Velocity	Standard Deviation of the within subject turn velocity
<b>Mean Encoder Difference</b>	Mean difference between the right and left encoder
<b>Standard Deviation Encoder Difference</b>	Standard deviation of difference between the right and left encoders

TABLE 3.3: Frequency Domain Features

<b>Feature</b>	<b>Description</b>
3dB Bandwidth (x,y,z)	3dB Bandwidth - accelerometer signal (walking phase)
3dB Bandwidth Turn (x,y,z)	3dB Bandwidth - accelerometer signal (turning phase)
<b>99% Occupied Bandwidth (x,y,z)</b>	Bandwidth containing 99% total integrated power - accelerometer signal (walking phase)
<b>99% Occupied Bandwidth Turn (x,y,z)</b>	Bandwidth containing 99% total integrated power - accelerometer signal (turning phase)
Mean Frequency (x,y,z)	Mean Frequency - accelerometer signal (walking phase)
Mean Frequency Turn (x,y,z)	Mean Frequency - accelerometer signal (turning phase)
<b>99% Occupied Power (x,y,z)</b>	99% total integrated Power - accelerometer signal (walking phase)
<b>99% Occupied Power Turn (x,y,z)</b>	99% total integrated Power - accelerometer signal (turning phase)
<b>Half Power (x,y,z)</b>	Frequency when signal drops to Half Power - accelerometer signal (walking phase)
<b>Half Power Turn (x,y,z)</b>	Frequency when signal drops to Half Power - accelerometer signal (turning phase)

TABLE 3.4: Information Theoretic Features

<b>Feature</b>	<b>Description</b>
Cycle Frequency *	Fundamental Frequency of the accelerometer walking pattern from DFT
Step Asymmetry *	Ratio of the 1st ( $A_{d1}$ ) and 2nd ( $A_{d2}$ ) autocorrelation coefficients
Step Regularity *	First, dominant autocorrelation coefficient indicating periodicity
Walk Ratio *	Average step length divided by cadence (indicating rhythmic walking)
Harmonic Ratio	Ratio of sum of even to sum of odd harmonics (accelerometer)
Entropy (x,y,z,rms, spherical )	Uncertainty measure of the accelerometer signal, linked to regularity
ZCR (x,y,z,rms, spherical )	Zero Crossing Rate of accelerometer signal
AR (x,y,z,rms, spherical )	Auto-regression coefficient of accelerometer signal
TKEO (x,y,z,rms, spherical )	Instantaneous Changes in energy (Teager-Kaiser energy) of accelerometer
Total Harmonic Distortion (x,y,z)	Measure of signal distortion compared to the harmonics
Walking Intensity *	Modulus taken of the integral of the accelerometer output
RMS to medio-lateral accelerometer ratio	Proportional measure of rms to medio-lateral
RMS to antero-posterior accelerometer ratio	Proportional measure of rms to antero-posterior
Correlation of R-L force	Correlation between the left and right force sensor measurements
Correlation between each accelerometer axis	Correlation of (x,y,z,rms, spherical) with each other
MI between each accelerometer axis	Mutual Information of (x,y,z,rms, spherical) with each other
Cross Entropy between each accelerometer axis	Cross Entropy of (x,y,z,rms, spherical) with each other

### 3.6 Feature Selection

The feature derivation stage produced a large, high-dimensional feature set of 211 features (listed in the aforementioned Tables 3.1 - 3.4). Not only would such a high dimensional feature set contain redundant features, but machine learning algorithms perform better with reduced dimensionality [41].

Furthermore, a significant aspect of the study was to determine which features are most relevant and provide the most insight into discriminating between different PD severity stages. Feature selection served this purpose through selecting relevant features which best characterize the changes between the different severity groups.

Three techniques of feature selection were conducted in this study for the aforementioned purposes and are described below.

#### 3.6.1 Principal Components Analysis (PCA)

Projection based PCA described in Section 2.4.3 was used to reduce dimensionality. As required by PCA, the features were normalized to the range between 0 and 1 as per Equation 3.5. This ensured features were comparable despite having different units and scales. Moreover, this mitigated the bias introduced by features with high variance.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.5)$$

Where:

$X_{norm}$  = Normalized feature value

$X_{max}$  = Maximum feature value

$X_{min}$  = Minimum feature value

$X$  = Value to be normalized

As is common practice, the number of Principal Components (PCs) chosen explained 95% of the variance of the feature set [79]. This ensured that critical information was not lost due to the reduction in dimensionality.

A major drawback of PCA is its inherent trait, that the selected feature set contains linear combinations of features. This transformation poses a problem from a clinical perspective, as the features could no longer be easily interpreted by clinicians: A particular focus in the selection process was to determine the most important features in terms of discrimination between the different PD stages, as determined by the H & Y scores. Consequently, in an attempt to preserve the units of the feature set so that clinicians can still interpret the results, two further feature selection techniques were implemented:

(1) Selection of features based on relevance to clinical usage and (2) Selection of features using a one-way ANOVA test.

### **3.6.2 Selection of features based on relevance to clinical usage**

Clinicians make use of spatio-temporal observations when assessing PD, in the initial diagnosis and later in determining the disease stage. Consequently, following a consultation with clinicians, the following 13 features were selected from the spatio-temporal parameters of Table 3.1. These features quantify the characteristics used by clinicians in their assessments of gait and have the most clinical relevance according to the clinicians' suggestions.

1. Turn time
2. Mean Step Time
3. Standard Deviation of Step Time
4. Mean Step Length
5. Standard Deviation of Step Length

6. Mean Velocity
7. Standard Deviation of Velocity
8. Mean Acceleration
9. Standard Deviation of Acceleration
10. Number of steps taken (during walking phase)
11. Total TUG time (including both straight line walking and turn time)
12. Total straight line walking time (excluding turn time)
13. Cadence

### **3.6.3 Feature Selection using ANOVA**

The data consisted of subjects from different H & Y groups. These groups had unknown statistical significance of the differences for individual features. The Analysis of Variance (ANOVA) statistical test was used to ascertain if there was a statistical difference between the means of three or more of these H & Y groups. The Null Hypothesis was that the means of each of the groups are equal. However, if a statistical difference was found with a p-value of less than 0.05, then the alternative hypothesis was accepted, which stated that at least two H & Y groups out of the total number groups have means that are statistically different.

The ANOVA test was performed on each individual feature to statistically determine which features were representative of the differences between the different H & Y groups. Those features that rejected the Null Hypothesis were then selected as viable features given the statistical significance of the difference.

## **3.7 Feature Space Analysis**

Prior to applying machine learning for classification it is useful to understand the distribution and other statistical properties of the features across the different

classes. This aligns with the goal of quantitative understanding into the manifestation of different PD stages. Thus, the distributions/probability density functions of each feature were analysed as follows.

### 3.7.1 Normal Distribution Testing

Many processes such as biological, financial, astronomical, physics based etc are often modelled by a normal distribution. The reason is that many measurements fluctuate around a central measure [80]. Furthermore, when the measurement information is limited to precision or error then the assumption of Gaussian (normal) fluctuations are typically seen as a good choice. This inference impacts the statistical tests used, as parametric tests assume normal distributions of the data, whilst non-parametric tests do not assume a normal distribution.

The Kolmogorov-Smirnov and the Anderson Darling tests were run on each individual feature to test whether the data follows a normal distribution. Both tests rejected the null hypothesis of accepting normality, if the p-value was less than 0.05.

### 3.7.2 Distribution Fitting

Since preliminary analysis showed that many of the features were not normally distributed, it was important to ascertain the distributions of each individual feature. While these distributions were not used in machine learning, their pursuit may provide quantitative understanding of the PD features by being a probabilistic model of the features manifestation.

Fitting of different possible distributions to each individual feature was performed. The distributions fitted were the following continuous distributions (supported by MATLAB): Beta, Birnbaum-Sanders, Exponential, Extreme value, Gamma, Generalized extreme value, Generalized Pareto, Inverse Gaussian, Logistic, Log-logistic, Lognormal, Nakagami, Normal, Rayleigh, Rician, t location-scale and

Weibull. The aforementioned extensive list of distributions was fitted to allow for a more specific distribution fit. When fitting the different distributions to each feature, the method employed maximum likelihood estimation (MLE).

The aforementioned extensive list of distributions was fitted to allow for a more specific distribution fit. When fitting the different distributions to each feature, the method employed maximum likelihood estimation (MLE).

Once all the distributions are fitted, the best distribution for each feature was selected using the Bayesian Information Criterion (BIC), which minimizes Equation 3.6 below. The distribution with the lowest score is considered the optimal best fit.

$$BIC = -2 \cdot \ln(L) + w \cdot \ln(v) \quad (3.6)$$

Where:

$v$  = size of the feature vector under examination

$w$  = number of parameters

$L$  = maximum likelihood function -  $P(\text{data}|\text{parameters}, \text{model})$

First, the distributions were fitted for each feature in the full feature set. Secondly, the feature set was segmented based on H & Y group (i.e. severity). Each distribution was then fitted for each feature for the different H & Y groups. e.g. the distributions were fitted for the mean step length feature for each H & Y group. This allowed the distributions of each feature, based on the severity of the disease, to be obtained.

### 3.8 Machine Learning

The primary research question lies in finding the extent in which automated, quantitative discrimination between PD stages can be achieved, based on the acquired signals. Machine learning was used to quantitatively answer this question.



Machine learning can be specified as either a classification or a regression problem. Both options were explored in this study and are described in Sections 3.8.1 and 3.8.2 respectively.

### **3.8.1 Classification**

Classification or supervised learning given the class labels (Hoehn and Yahr - H & Y scores) was first explored. The clinical H & Y scores were used as the class labels. Whilst, the score may be subjective, these scores are given by experts, according to an accepted test, and are therefore considered the current gold standard in Parkinson's disease (PD) assessment. Thus, since clinicians' derived scores are the metric which will serve as the benchmark for performance of the thesis proposed system. Moreover, as described in Section 2.3.1, the H & Y scale is one of the two accepted PD stage metrics (the other being the UPDRS – another clinician derived scale) and both are the only available ground truth tools. Therefore it makes sense to use such scores for comparative purposes.

The choice of supervised learning paradigm was based on the need to provide not only a discrimination accuracy but also an insight into the relevance of the features with respect to PD stage discrimination. Consequently, supervised learning models such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) were unsuitable [65].

Hence, Random Forests were used as the classification algorithm as it solved this problem, since in this classification method, feature importance is quantifiable. A detailed description of random forests was provided in Section 2.4.4.4, thus only implementation and testing is outlined in the forthcoming sections.

#### **3.8.1.1 Random Forest Implementation & Testing**

The implementation of the random forest was done as per the methods outlined in Section 2.4.4.4, as expounded by Breiman [64]. The implementation of the random

forest (RF) then had three different stages: (1) Determine the optimal minimized forest size by analyzing the accuracy of RF's of 1-1000 trees, (2) Determine the important features using random permutation and (3) Test the performance accuracy of the optimal forest with the different feature subsets from Section 3.6. This is described below:

(1) Determine the optimal forest size: In order to determine the optimal forest size, random forests of with 1-1000 trees were evaluated and the accuracy plotted. The knee point of the graph (i.e. the smallest forest size where the error does not decrease anymore such that it tends to a limit) is considered as the optimal forest size. The evaluation is done as per Algorithm 2.

---

**Algorithm 2:** Optimal Forest Size

---

- 1: Initialise state  $i=1$
  - 2: Train a random forest of size  $i$  on the full 211 dimensional feature set. As recommended by Breiman [64] a different random bootstrap training sample which omits  $1/3$  of the feature set was used to train each tree.
  - 3: After training pass the entire training set through the random forest to test performance and record the accuracy.
  - 4: Increment  $i$  by 1
  - 5: Repeat steps (2) to (4) until  $i=1000$  (1000 was chosen to compare how performance changes with increasing forest size)
  - 6: Compare the generalization error of each forest size, by plotting a graph of error vs number of trees
  - 7: The optimal forest size is seen on the graph as the knickpoint where the error tends to a limit. i.e. the smallest forest size without losing accuracy.
- 

When each tree is grown,  $1/3$  of the data is randomly excluded. Hence, a separate test set or cross-validation set was not required to obtain unbiased test results, as randomness is injected internally when each tree is grown using a different random bootstrapped sample, which omits  $1/3$  of the data [64].

(2) Algorithm 3 below, describes how feature importance was determined using random permutation. The random permutation of features was performed on the optimal forest of 100 trees, obtained using Algorithm 2.

---

**Algorithm 3:** Feature Importance [65]

---

- 1: Initialise state  $i=1$
  - 2: For feature  $x_i$ , randomly permute the feature column. (i.e. randomly shuffle the feature column  $x_i$  )
  - 3: Compute the change in the error rate (i.e. accuracy of predictions) compared to the original baseline error of the optimal forest.
  - 4: Record the increase/decrease in error. The larger the error the higher the feature importance. The reason for this is that by permuting important features the accuracy will significantly decrease, whilst permuting unimportant features will have minimal impact on accuracy
  - 5: Increment  $i$  by 1
  - 6: Repeat steps (2) to (4) for all  $i$  features (where  $i=211$ )
- 

The test of feature importance was repeated 5 times. The features which were used (as important features) by the random forest in 80% of the trials, were selected as the "important features".

(3) Finally, the following feature subsets were tested with the optimally sized random forest:

1. Full feature set
2. Principal Components
3. Features selected by the random forest
4. ANOVA selected features
5. Features selected based on clinical relevance

The discrimination accuracy between the aforementioned feature subsets was then compared.

### 3.8.2 Regression

A regression model was used to predict H & Y scores. The rationale is that given the coarse scale of 1 to 4, comparison of patients' symptoms within and across classes may be inaccurate. While clinicians, who are the gold standard assign these scores, it may happen that patients with the same clinical score exhibit different symptoms. Thus, a regression approach, may allow for a finer scale of the H & Y. For example, where two patients have the same H & Y of 2, one may be a more severe case than the other and hence be assigned an H & Y score of 2.3. This finer scale may allow for enhanced comparison between patients and for better monitoring of a patient's decline, which is not manifested using the current H & Y.

A regression approach can provide a mathematical model. Although a determination of importance and relevance for each specific feature cannot be derived from a regression model, this is an acceptable trade-off as feature importance has already been obtained using the classification model. Moreover, the regression model serves to illustrate the model/pattern mapping of input features to clinical scores, whereby the predicted scores can be compared against the ground truth.

The following machine learning regression methods were designed and implemented:

#### 3.8.2.1 Stepwise Linear Regression

Whilst the assumption was that the model was complex and non-linear, it can still be informative to explore linear models to validate the assumption. Thus, a stepwise linear regression model was used and developed as per Algorithm 4.

Testing of the model was done as per Section 3.8.2.7.

---

**Algorithm 4:** Stepwise Linear Regression Algorithm

---

- 1: Initialise the empty model
  - 2: Perform regression (mapping of  $x$ - $y$ ) on each individual feature in the dataset. i.e. regression for  $x_1, x_2, \dots, x_N$  separately.
  - 3: Calculate the t-statistic and p-value for each feature. The feature with the lowest p-value is added to the model, provided the p-value is less than 0.15.
  - 4: Feature  $x_i$  will be selected based on the p-value in the previous step. The model is then  $n$  dimensional.
  - 5: Fit each of the remaining features to make a increased  $(n+1)$  dimensional feature model and perform regression on that feature. i.e. regression for  $x_1, x_2, x_3, \dots, x_N$  separately.
  - 6: Calculate the t-statistic and p-value for those remaining feature. The feature with the lowest p-value is added to the model provided the p-value is less than 0.15
  - 7: Recheck the p-values of those predictors originally in the model before the most recent addition. If the significance decreases then remove the added feature
  - 8: Repeat steps 1) to 7). Until the stopping criteria in step 9 is met
  - 9: Stop when adding the additional feature does not yield t-statistic p-value of less than 0.15 (Based on literature [81]).
- 

### 3.8.2.2 Robust Linear Regression

Robust linear regression uses a weighting function to address the issue of outliers that plagues Ordinary Least Squares (OLS) regression. The designed Robust linear regression model used a bi-square weighting function given by Equation 3.7.

$$w = (|r| < 1) \times (1 - r^2)^2 \quad (3.7)$$

Where:

$r$  = specific data point

Initially, each data point/variable got the same weight and the model coefficients were estimated using OLS. The weights were then iteratively recomputed such that points closer to the line got higher weights and those further from the line got reduced weights. Thereafter, coefficients were recomputed using the new weights.

The model was tested as outlined in Section 3.8.2.7

### **3.8.2.3 Support Vector Regression**

Support Vector Regression (SVR) is a non-parametric method of regression which relies on kernels to produce a function which maps input to output relationships as was outlined in Section 2.4.4.2.

The problem studied here after preliminary analysis of the linear models was found to be complex and multi-dimensional. Hence, a non-linear SVR with a Gaussian kernel function was used as the regression model. The model testing was done as described in Section 3.8.2.7.

### **3.8.2.4 Random Forest Regression/ Regression Forests**

Random Forests that are used for regression are called Regression Forests. The difference between the classification and regression applications is that classification uses majority voting while regression uses averaging, as discussed in Section 2.4.4.4.

A regression forest was grown with the optimal number of trees obtained in the classification. The model was tested as outlined in Section 3.8.2.7.

### **3.8.2.5 Boosted Trees Regression**

Boosted Trees Regression are an ensemble learner which makes use of decision trees, but in a different manner to random forests as was discussed in the Section 2.4.4.5.

An ensemble of boosted regression trees was grown with the optimal number of trees as obtained by the random forest. The model was tested as outlined in Section 3.8.2.7.

### **3.8.2.6 Artificial Neural Network (ANN)**

A Bayesian Regularized Artificial Neural Network (BRANN) was designed for the regression problem. As discussed in Section 2.4.4.1, BRANNs address the issues of overfitting and increased bias and variance associated with typical ANNs [59]. Moreover, when compared to early stopping which is typically used in deep neural networks and heavily relies on intuition for good performance, Bayesian regularization is a more formalized method of preventing overfitting.

A BRANN was designed with 20 hidden layers (using a tan-sigmoidal activation function [54, 55]) and a single output layer to predict H & Y scores. Training and testing of the model was conducted as outlined in Section 3.8.2.7.

### **3.8.2.7 Model Testing**

The different regression models were tested and compared. The different regression models used the full derived feature set as inputs. The inputs are then mapped by the regression model to outputs which are finer scaled H & Y scores. The generalization performance of each regression model was tested using 10-fold cross validation. The patient cohort was divided randomly into 10 equal sized subsets. The model was trained on 90% of the dataset (i.e. 9 subsets) and then 10% of the data (i.e. 1 subset) was retained as the test set. This process was then repeated 10 times, so that each patient subset served as the test set.

The predicted H & Y in the 10 cross-validated tests were then compared with the dataset's H & Y scores, which were determined by the neurologists. The comparison employed 15 performance metrics discussed in Section 2.5. These metrics are summarized in Table 3.5. Each round of cross-validation calculated a separate value for the performance metrics. Finally, the mean of the 10 tests was computed to obtain average values for each performance measure.

The different regression models were then compared. The most appropriate model minimized the error between predicted and actual H & Y (e.g. MSE, RMSE, MAPE etc) and maximized the similarity between predicted and actual H & Y (e.g. cosine similarity, FAC2, correlation etc). This evaluation was done using the regression performance metrics shown in Table 3.5.

TABLE 3.5: Regression Performance Metrics

Performance Measure	Equation
Mean Squared Error (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n (Y_o - Y_p)^2$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n  Y_o - Y_p $
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{ Y_o - Y_p }{Y_o}$
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\sum \frac{(Y_p - Y_o)^2}{N}}$
Sum of Squares Error (SSE)	$SSE = \sum_{i=1}^n (Y_o - Y_p)^2$
Cosine Similarity (CS)	$CS = \frac{Y_o \cdot Y_p}{\ Y_o\  \ Y_p\ }$
Modified Index of Agreement (IA)	$IA = 1 - \frac{\sum_1^N  Y_p - Y_o }{\sum_{i=1}^N ( Y_p - \bar{Y}_o  +  Y_o - \bar{Y}_o )^2}$
Mutual Information (MI)	$MI = \sum P(Y_o, Y_p) \cdot \log\left(\frac{P(Y_o, Y_p)}{P(Y_o)P(Y_p)}\right)$
R squared (Rsq)	$Rsq = \frac{\sum (Y_p - \bar{Y}_p)^2}{\sum (Y_o - \bar{Y}_o)^2}$
Spearman's Correlation	$Correlation = \frac{COV(Y_p, Y_o)}{\sigma_{Y_p} \sigma_{Y_o}}$
Fractional Bias (FB)	$FB = \frac{\sum_i (Y_o - Y_p)}{0.5 \sum_i (Y_o + Y_p)}$
Geometric Mean Bias (GMB)	$GMB = e^{(\ln(Y_o) - \ln(Y_p))}$
Normalized Mean Square Error (NMSE)	$NMSE = \frac{(\bar{Y}_o - \bar{Y}_p)^2}{Y_o Y_p}$
Geometric Variance (GV)	$GV = e^{(\ln(Y_o) - \ln(Y_p))^2}$
Fraction of predictions- factor of two (FAC2)	<i>Fraction satisfies</i> $0.5 \leq \frac{Y_p}{Y_o} \leq 2.0$



### 3.9 Statistical Analysis

The ANOVA test was used to compare the means of the different H & Y groups. However, the ANOVA p-value does not indicate overlap or the differences between individual groups as was discussed in Section 2.6. Confidence intervals (CI's) were used to address this issue and it was proposed by [72], that a  $\leq 50\%$  overlap of CI's indicates a statistically significant difference in group means. Whilst, [73] stated that  $\leq 50\%$  overlap corresponds to a statistical significance of 95%, whilst no overlap corresponds to a 99% statistical significance.

#### 3.9.1 Confidence Interval Calculation

The means of each H & Y group for each individual feature was compared using CI's. The CI's were calculated using Equation 3.8 [72]. The CI equation produces an estimated range of values (lower and upper bounds), which have a high probability of covering the true population value [73]. The pairwise overlap of that group with every other H & Y group was compared to ascertain statistical significance of the difference. For example: the overlap of H & Y 0 is compared to H & Y 1, H & Y 2, H & Y 2.5, H & Y 3 and H & Y 4. This comparison is done pairwise for every group and had 15 possible comparison permutations.

$$CI = \mu_i \pm \frac{\sqrt{2}}{2} (t_{n-1, 1 - \frac{\alpha}{2}} \sqrt{\frac{J}{J-1}}) (\sigma_i / \sqrt{n}) \quad (3.8)$$

Where:

- $\mu_i$  = mean of the *i*th group
- $t_{n-1, 1 - \frac{\alpha}{2}}$  = value from the t-distribution ( $\alpha$  is 0.05)
- $J$  = number of levels (number of classes in the experiment)
- $(\sigma_i / \sqrt{n})$  = standard error of the *i*th group
- $\sigma_i$  = standard deviation of the *i*th group
- $n$  = number of subjects in that group

### 3.9.2 Feature Classification Performance Comparison

The assessment of CI overlap between groups allowed the comparison to determine which features had the least and most overlap between different H & Y groups. This was an indication of class separability. The features with the least overlap (best discrimination performance) were then compared to the ANOVA selected features, as well as the features selected by the random forest.

It is reasonable to assume that different features behave and manifest differently at different stages of the disease. This assumption was corroborated by the clinicians as well.

Thus, the CI significant features chosen were those that for any of the 15 possible permutations which compared all H & Y pairs; at most 2 permutations were not statistically significant (i.e. > 50% overlap).

The common features between the CI statistical method and the aforementioned computational based methods served as a further validation step of the features which are crucial to the discrimination between PD stages.

## CHAPTER 4

### RESULTS

The results of the research are presented in this chapter, in the same order as the methodology steps described in Chapter 3. A detailed analysis and discussion of the results and the insights derived from the results is presented in Chapter 5. In this chapter, the data integrity and organisation check is followed by the results of the signal pre-processing steps. Thereafter, the outcomes of different methods of feature selection are presented. Findings from feature space analysis are followed by the results of the different machine learning algorithms. Finally, the results of statistical analysis using CI's is presented.

#### 4.1 Data Integrity and Organisation Check

The breakdown of the dataset according to H & Y severity group, after the data integrity and completeness checks is listed in Table 4.1 below. Patients were excluded as per the criteria outlined in Section 3.3.

TABLE 4.1: Dataset Integrity Analysis comparing the number of subjects excluded and remaining for each H &amp; Y group

<b>H &amp; Y Group</b>	<b>Total Number</b>	<b>Number Excluded</b>	<b>Number Remaining</b>
0 (Control)	21	0	21
1	7	0	7
2	26	2	24
2.5	18	1	17
3	14	2	12
4	6	1	5
<b>Total</b>	<b>92</b>	<b>6</b>	<b>86</b>

There is a distinct class imbalance among the different H & Y groups especially in the smaller number of H & Y 1 and 4 subjects. This will be taken into consideration and the analysis discussed accordingly in Chapter 5.

## 4.2 Signal Pre-processing

The results of signal pre-processing is outlined below. It covers the results of signal denoising, segmentation of movement phases, as well as footfall detection.

### 4.2.1 Signal Denoising

The moving average filter coupled with zero-phase lag compensation was successfully applied to denoise the noisy signal. An example result for the velocity signal is shown in Figure 4.1. The small window size can be seen to have allowed for localized attenuation of the high frequency noise.

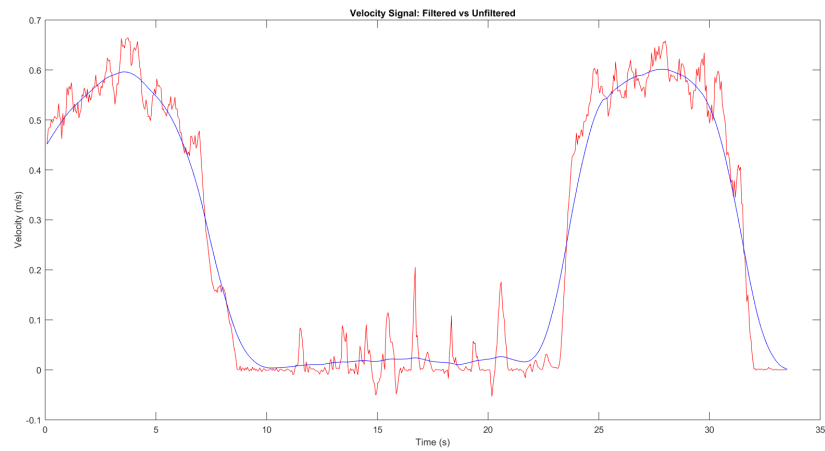


FIGURE 4.1: Example of a velocity signal before (red line) and after the application of a moving average filter (blue line).

## 4.2.2 Segmentation of movement phases

Movement phase segmentation involved identifying regions of straight line movement, as well as regions of stationary or turning movement. The algorithm was applied to the velocity signal and can be seen in Figure 4.2 below.

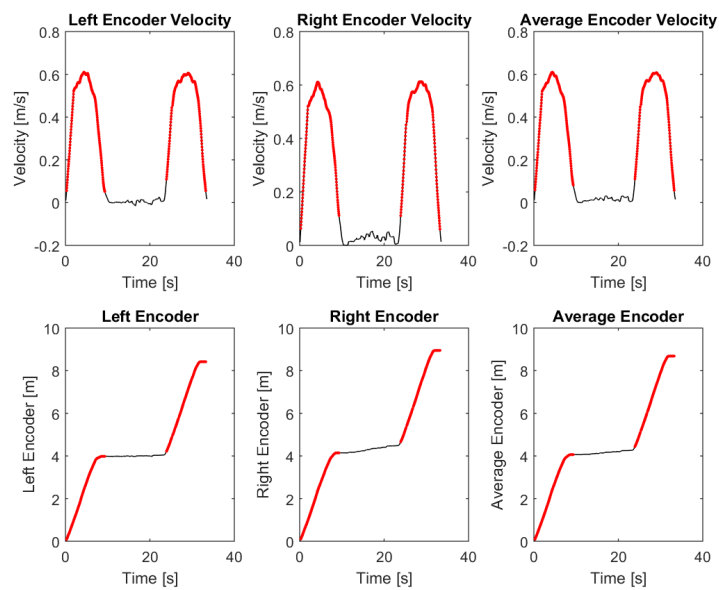


FIGURE 4.2: Movement Phases Algorithm Results where red lines are predicted movement periods and black lines are predicted stationary or turning periods

### 4.2.3 Footfall Detection Algorithm

The two footfall detection algorithms made use of the accelerometer and force sensor signals respectively. Firstly, the EMD algorithm was used to extract IMFs. An example of the initial four IMFs can be seen in Figure 4.3 below. The different extracted oscillatory components are clearly visible, with decreasing frequency in the subsequent IMFs.

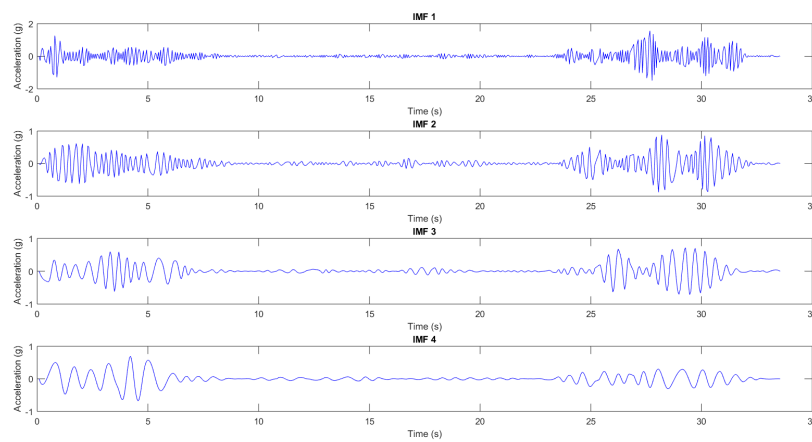


FIGURE 4.3: Example of the First, Four IMFs which show a decreasing frequency

The signal was then reconstructed, starting from different IMF levels (from IMF 1-4). After signal reconstruction, peak detection was performed on the signals, where the peaks were representative of footfalls.

Footfall detection accuracy was calculated as follows. The video validated footfall range was between heel strike and toe-off. The algorithm predicted footfall was deemed accurate if it was detected within the timestamped range of the validated footfall.

Figure 4.4 shows an example of the predicted footfalls (circles) compared to the validated video footfalls ranges between the red and oranges X's.

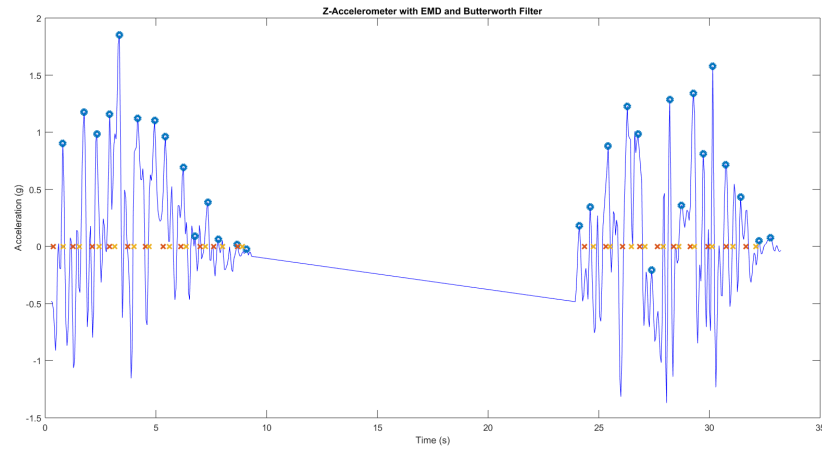


FIGURE 4.4: Example of the Footfall detection algorithm: The predicted footfalls are represented by circles and the validated video-captured footfalls ranges are plotted as “x”s when a red “x” is a start timestamp of the range and oranges “x” the end of the range .

The algorithm accuracy was validated against four tests captured by video, which had different walking patterns (simulating PD shaking, foot dragging, normal walking, slow and irregular walking). The accuracy performance of the novel accelerometer algorithm is shown in Figures 4.5 and 4.6.

The second footfall detection algorithm was the replicated force sensor algorithm proposed by [39]. The accuracy performance of this algorithm is depicted in Figures 4.7 and 4.8.

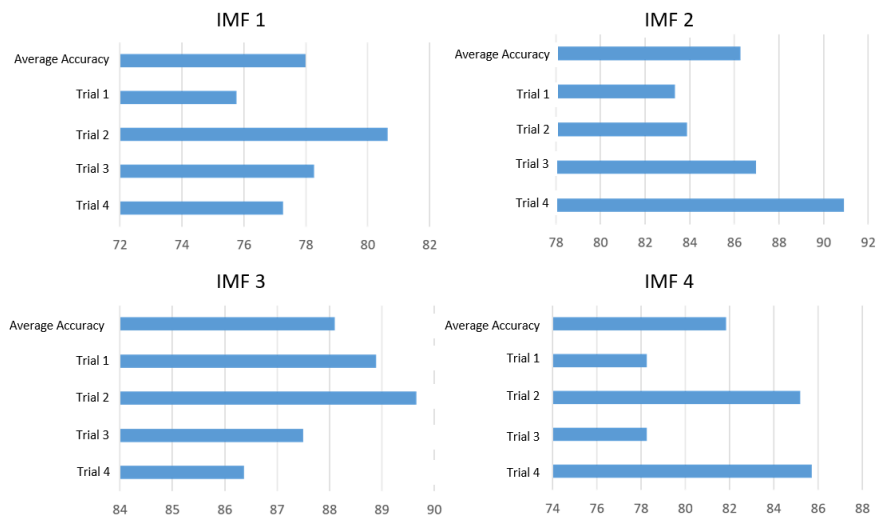


FIGURE 4.5: Novel Accelerometer Algorithm: Percentage Accuracy of footfall detection with reconstruction from different IMF's. Average accuracy indicates the mean accuracy of the four trials for the specific IMF

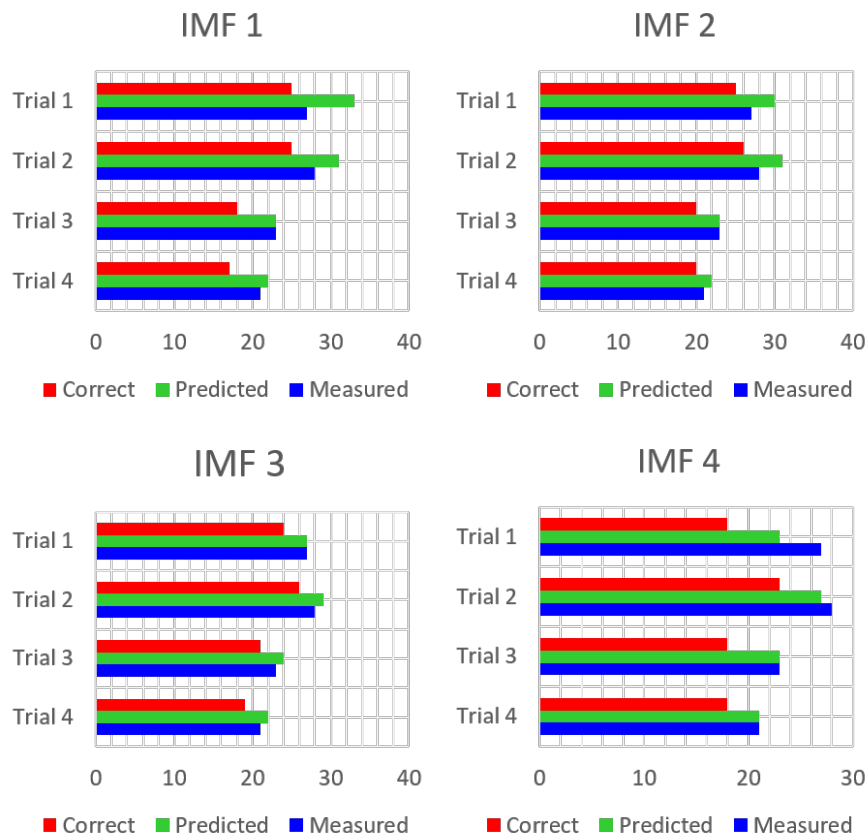


FIGURE 4.6: Novel Accelerometer Algorithm: Comparison of number of actual footfalls (segmented from the video) to the predicted ones: the blue bar represents the number of true measured footfalls, the green bar represents the number of predicted footfalls, the red bar represents the number of correctly predicted footfalls



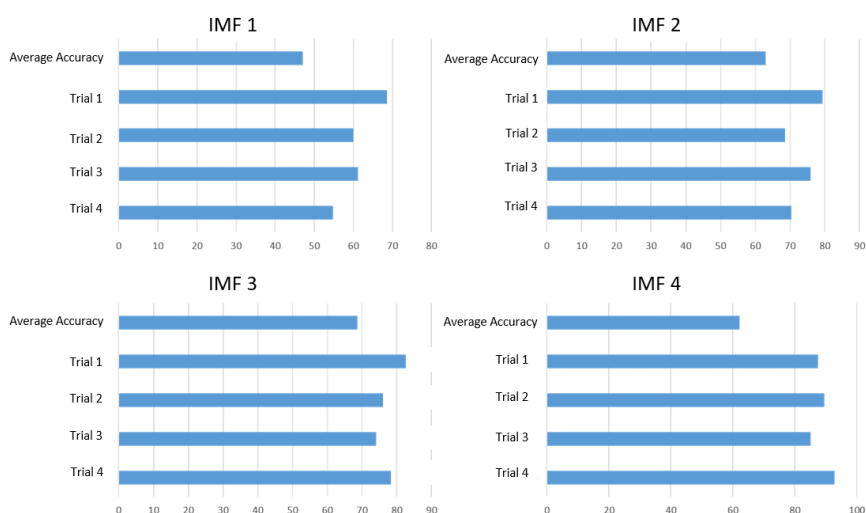


FIGURE 4.7: Force Sensor Algorithm: Percentage Accuracy of footfall detection when reconstructing from different IMFs. Average accuracy indicates the mean accuracy of the four trials for the specific IMF

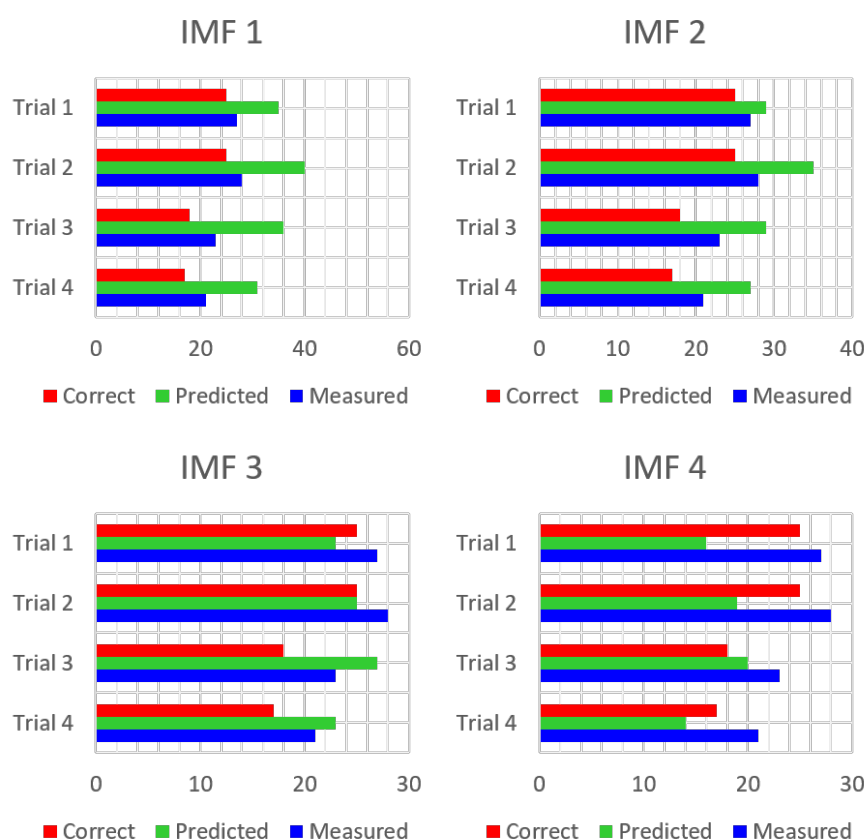


FIGURE 4.8: Force Sensor Algorithm: Comparison of number of actual footfalls (segmented from the video) to the predicted ones: the blue bar represents the number of true measured footfalls, the green bar represents the number of predicted footfalls, the red bar represents the number of correctly predicted footfalls

Table 4.2 conveys that reconstructing from IMF 3 yields the highest accuracy for both algorithms. Additionally, the newly proposed  $z$ -axis accelerometer footfalls detection algorithm was compared to the replicated force sensor detection results. The performance of the two algorithms is illustrated in Table 4.2 where accuracies of all combinations of IMF reconstruction levels are computed.

TABLE 4.2: Predicted Footfalls Accuracy for the two Algorithms. Each column contains the accuracy (in percentage) obtained for the different IMF reconstructions. The four rows are the 4 walking tests. The fifth row is the average across tests for each IMF.

		<b>Algorithm Detection Rate (%)</b>							
		<b>New Algorithm</b>				<b>Literature Algorithm</b>			
<b>IMF</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
		64	75	89	67	55	74	70	51
		74	76	86	83	46	57	67	60
		78	87	84	78	44	62	64	73
		74	87	83	85	41	57	72	62
	<b>Average</b>	<b>72</b>	<b>82</b>	<b>86</b>	<b>78</b>	<b>47</b>	<b>63</b>	<b>69</b>	<b>62</b>

## 4.3 Feature Selection

### 4.3.1 Feature Selection based on clinical usage

As described in Chapter 3, the spatio-temporal feature subset listed below were the most intuitive features that can be easily communicated to and understood by clinicians. Moreover, they quantify the characteristics used by clinicians in their assessments of gait. The following features met this criteria of clinical relevancy.

**Features based on clinical relevance**

1. Turn time
2. Mean Step Time
3. Standard Deviation of Step Time
4. Mean Step Length
5. Standard Deviation of Step Length
6. Mean Velocity
7. Standard Deviation of Velocity
8. Mean Acceleration
9. Standard Deviation of Acceleration
10. Number of steps taken (during walking phase)
11. Total TUG time (including both straight line walking and turn time)
12. Total straight line walking time (excluding turn time)
13. Cadence

**4.3.2 PCA**

In this feature selection method, PCA was applied and the PC's that explained 95% of the variance of the feature space were chosen. The number of the chosen PC's were 47 and 42 for the 3m tests and the 10m tests, respectively.

### 4.3.3 ANOVA feature selection

One-way ANOVA was performed on each feature in the full feature set and the selected features were those that rejected the null hypothesis of equal means with a p-value less than 0.05. This implied a 95% certainty of a statistical difference between the means of the six H & Y groups. The ANOVA, Type I error probability in rejecting the Null Hypothesis was maintained at 5%.

This selection method yielded 67 features (3m test) and 66 features (10m test) out of the 211 features. A summary of the chosen features grouped by feature category is shown in Figure 4.9. The full list of chosen features which met this criteria is detailed in Appendix B, Tables B.1 and B.2 for the 3m and 10m tests respectively.

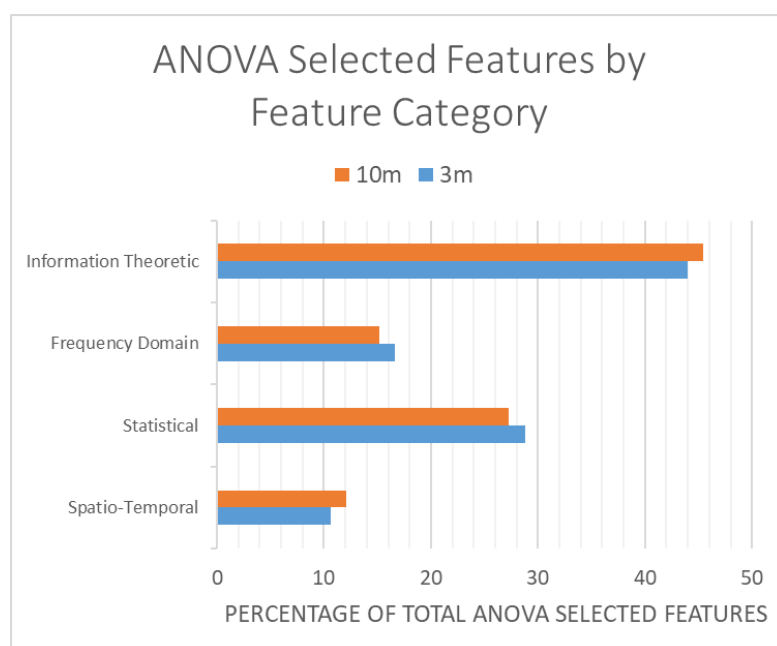


FIGURE 4.9: The ANOVA selected features grouped by feature category for the 3m and 10m tests

## 4.4 Feature Space Analysis

### 4.4.1 Normal Distribution Tests

The Kolmogorov Smirnov (KS) test rejected the Null Hypothesis for 100% of the features for both the 3m and 10m tests. The Anderson Darling (AD) test rejected the Null Hypothesis for 86% and 82% of the features for the 3m and 10m tests respectively.

The difference between the KS test and AD tests is that the KS test places greater weight on the center of the distribution, whilst the AD test allows for fatter tails to the distribution. The implication of both tests is that normality could not be assumed when conducting statistical tests. Therefore, the distribution fit for the features was assessed.

### 4.4.2 Distribution Fitting

#### 4.4.2.1 Entire Feature Space

After fitting different probability distributions to each of the individual features, the Bayesian Information Criteria (BIC) was used to obtain the "best fit" probability distribution. The lowest BIC score as described in Section 3.7.2 corresponded to the "best fit" distribution to the specific feature. The result yielded 17 different probability distributions for the features of the 3m test and 16 different probability distributions for the features of the 10m test. The probability distributions were tallied and the results are illustrated in Figures 4.10 and 4.11 for the 3m and 10m tests respectively.

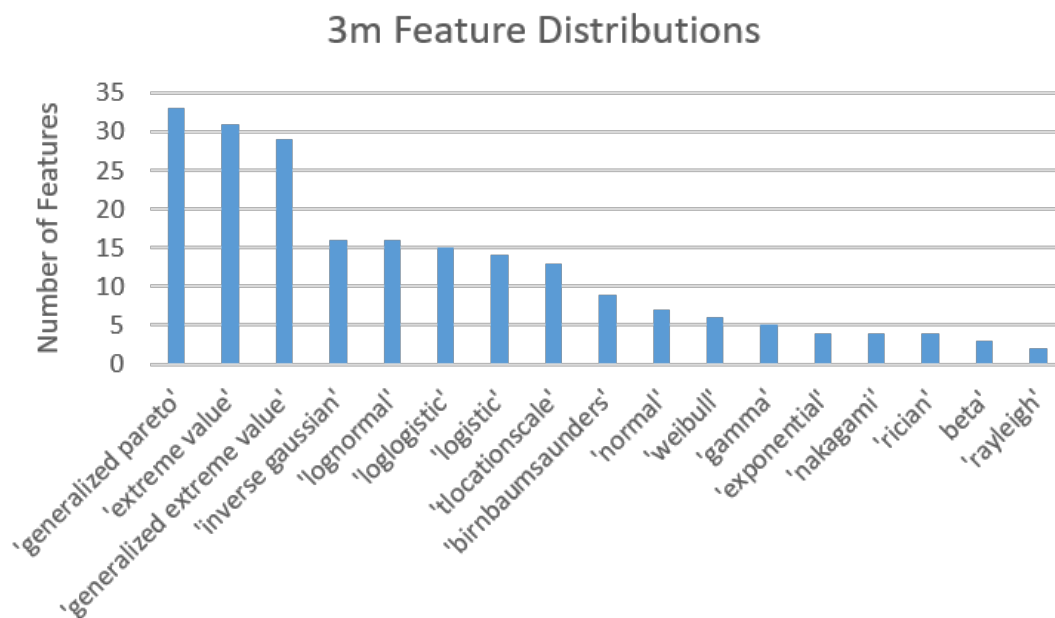


FIGURE 4.10: Number of features per each best fit (based on BIC) distribution for the 3m test

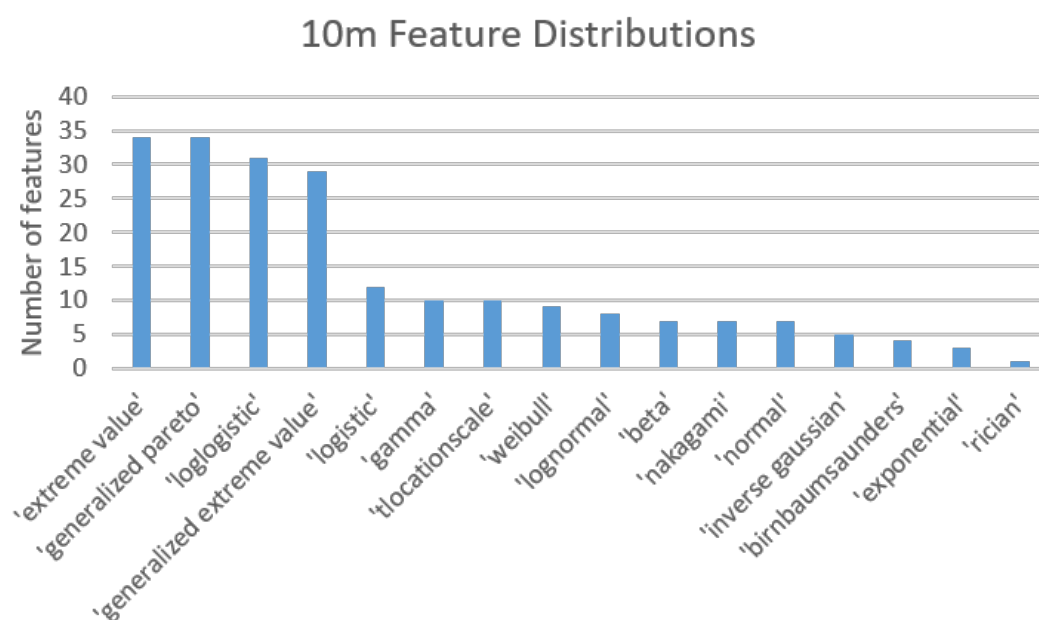


FIGURE 4.11: Number of features per each best fit (based on BIC) distribution for the 10m test

#### 4.4.2.2 Distributions of H & Y Severity groups

In order, to assess how the different probability distributions change as a function of disease severity, the features were then segmented into 6 groups, based on the different H & Y group (0, 1, 2, 2.5, 3 and 4). The probability distributions were then fitted to each feature for each H & Y group. The small number of H & Y 4 subjects and the implication the representation provided by the probability distribution is discussed in Chapter 5.

The 3 best probability distributions for the different H & Y stages are listed below. The results of the 3m test is shown in Table 4.3 and the 10m test in Table 4.4.

TABLE 4.3: The 3 Best Distributions for the different H & Y Groups in the 3m test. The percentage of features out of the total number of features is depicted below the distribution name

<b>Top 3 Distributions</b>	<b>H &amp; Y - 0</b>	<b>H &amp; Y - 1</b>	<b>H &amp; Y - 2</b>	<b>H &amp; Y - 2.5</b>	<b>H &amp; Y - 3</b>	<b>H &amp; Y - 4</b>
1	Gen Pareto - 48%	Gen. Pareto - 28%	Gen. Pareto - 85%	Gen. Pareto - 55%	Gen. Pareto - 85%	Gen. Pareto - 79%
2	loglogistic - 21%	Rayleigh - 20%	t- location scale - 8%	t- location scale - 18%	t- location scale - 8%	t- location scale - 10%
3	t- location scale - 10%	loglogistic - 18%	Gen. extreme value - 4%	Rayleigh - 14%	Gen. extreme value - 4%	Gen. extreme value - 4%

TABLE 4.4: The 3 Best Distributions for the different H & Y Groups in the 10m test. The percentage of features out of the total number of features is depicted below the distribution name

<b>Top 3 Distributions</b>	<b>H &amp; Y - 0</b>	<b>H &amp; Y - 1</b>	<b>H &amp; Y - 2</b>	<b>H &amp; Y - 2.5</b>	<b>H &amp; Y - 3</b>	<b>H &amp; Y - 4</b>
1	Gen. Pareto - 58%	Gen. Pareto - 44%	Gen. Pareto - 75%	Gen. Pareto - 48%	Gen. Pareto - 62%	Gen. Pareto - 79%
2	t-location scale - 13%	Rayleigh - 14%	t-location scale - 8%	Exponential- 10%	Rayleigh - 10%	Exponential- 14%
3	Gen. extreme value - 7%	loglogistic - 7%	Gen. extreme value - 7%	t-location scale - 7%	Gen. extreme value - 10%	t-location scale - 10%

The generalized pareto distribution is shown to be the most prevalent probability distribution among the features, in all groups. Figures 4.12-4.15 illustrate examples of the change in the generalized pareto distribution for the different PD stages.



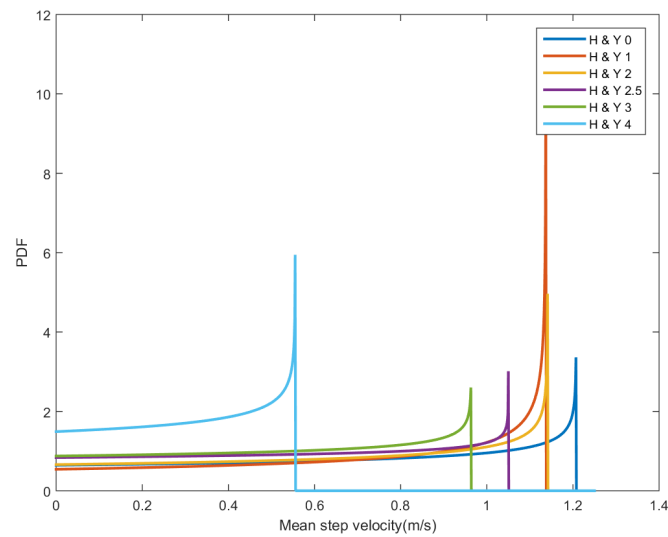


FIGURE 4.12: The probability distribution of the Mean Step Velocity feature. The distribution for each different PD stage is shown in different colours as per the figure legend.

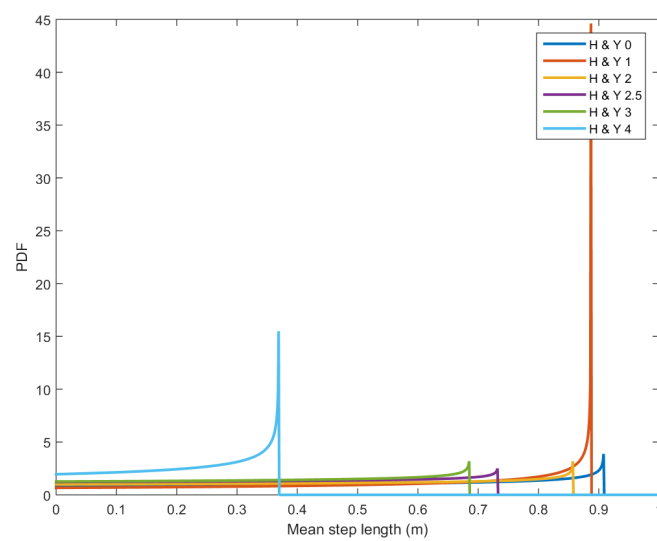


FIGURE 4.13: The probability distribution of the Mean Step Length feature. The distribution for each different PD stage is shown in different colours as per the figure legend.

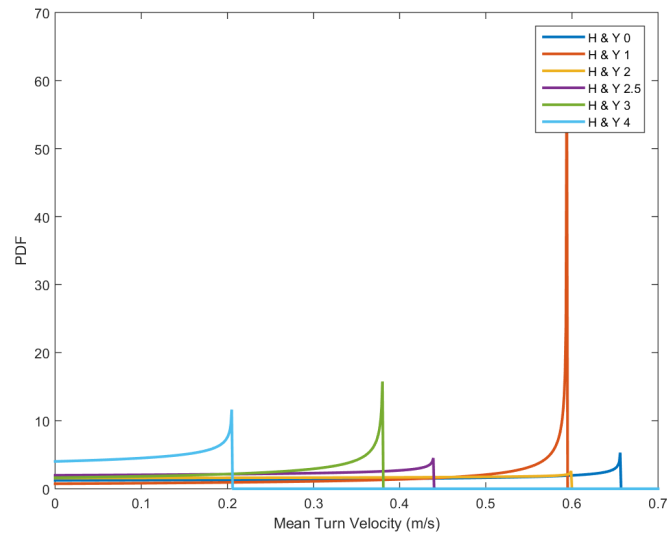


FIGURE 4.14: The probability distribution of the Mean Turn Velocity feature. The distribution for each different PD stage is shown in different colours as per the figure legend.

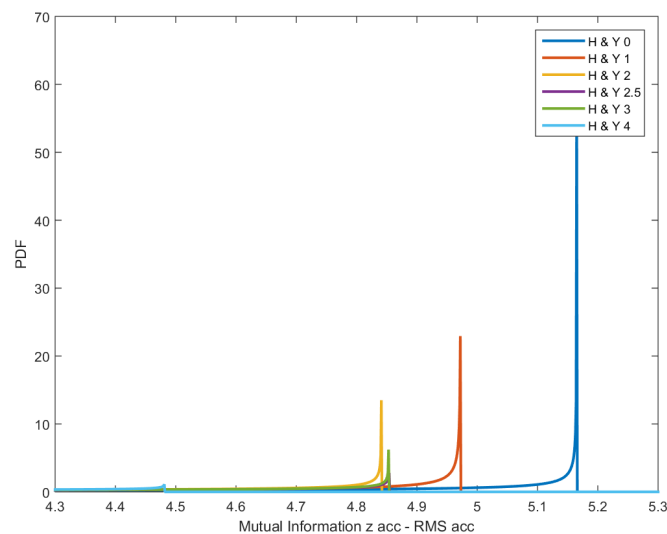


FIGURE 4.15: The probability distribution of the MI between z acceleration - RMS acceleration feature. The distribution for each different PD stage is shown in different colours as per the figure legend.

## 4.5 Machine Learning

The experiments and partitions utilized for both classification and regression experiments was described in detail in Sections 3.8.1.1 and 3.8.2.7 respectively.

## 4.5.1 Classification: Random Forest

### 4.5.1.1 Full Feature Vector

The random forest was trained as was described in Chapter 3. Different forest sizes ranging from 1-1000 trees were grown and tested to determine an optimized number of trees.

Figures 4.16 and 4.17 illustrate the percentage error in discrimination between different PD stages for the different forest sizes for the 3m and 10m tests respectively. A random forest of approximately 100 trees is the optimal forest size. The reason for the choice of optimal forest size of 100 is that this number of 100 trees in the random forest met the optimization criteria (i.e. knee point of Figures 4.16 and 4.17 ) as outlined in Section 3.8.1.1.

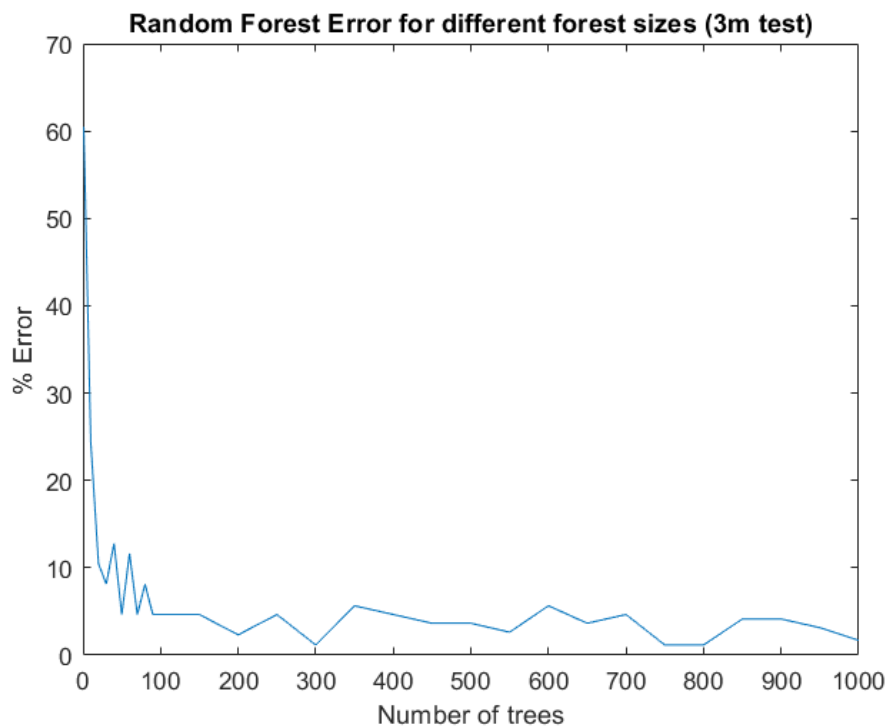


FIGURE 4.16: Discrimination error (in percentage) between PD stages, for 1-1000 trees, for the 3m test

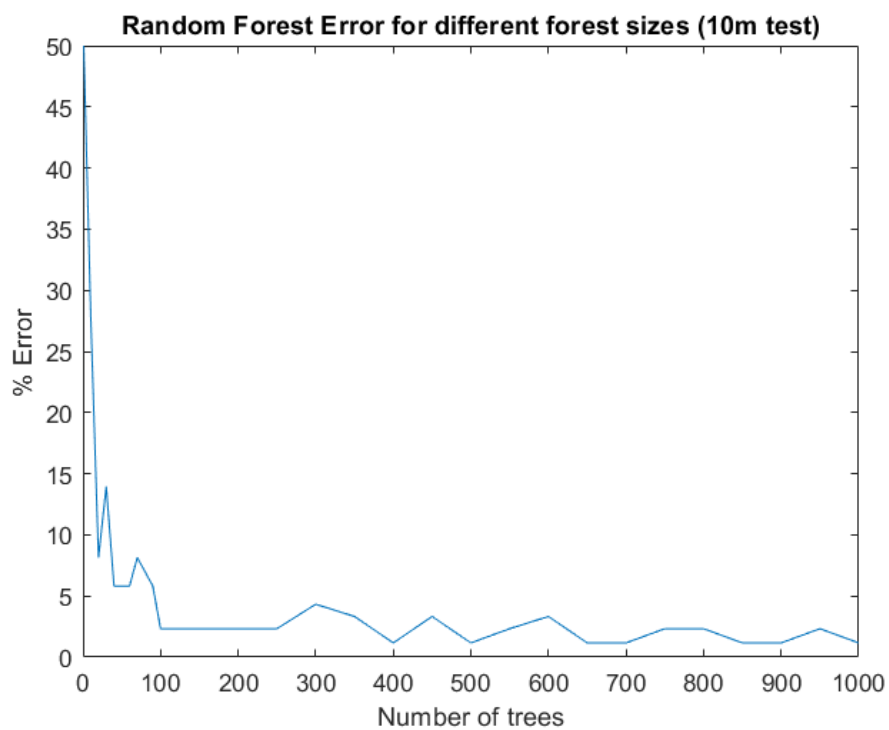


FIGURE 4.17: Discrimination error (in percentage) between PD stages, for 1-1000 trees, for the 10m test

#### 4.5.1.2 Random Forest Selected Features

The random forest classifier provides feature importance by indicating the features that were used in the classification. The 30 features in Table 4.5 represent the important features selected by the random forest classifier as was described in Section 3.8.1.1.

TABLE 4.5: Random Forest Selected Features which are the features used by the random forest in discriminating between different PD stages

Mean Encoder Difference	Radius Q1
Number of Steps	y Accel Q3
Total Tug time	RMS Accel Median
Standard Deviation Turn Velocity	x Accel IQR
Mean Turn Velocity	x Accel MSE
Walk Ratio	y Accel MSE
y Accel half power	azimuth MSE
y Accel 99% power	azimuth Entropy
y Accel Q1	Correlation z Accel - Azimuth
RMS Accel Entropy	Mutual Information x Accel - y Accel
Elevation Mean	Mutual Information y Accel- Radius
Mean Step Length	Mutual Information z Accel- Radius
Standard Deviation Accel	Mutual Information z Accel- RMS Accel
y Accel Variance	Skewness Azimuth
Mean Step Velocity	Total Harmonic Distortion y

#### 4.5.1.3 Performance comparisons with the optimal number of trees

Table 4.6 outlines the classification performance accuracy and execution time in discriminating between different PD stages for the different feature selection subsets. All feature subsets were evaluated and classified using the optimal random forest of 100 trees.

TABLE 4.6: Accuracy (Acc) and Execution Time Comparison using the optimal 100 tree Random Forest

Feature Subset	3m- Acc(%)	3m- Time (s)	10m-Acc(%)	10m- Time (s)
PCA	97	10.56	97	9.22
All	94	41.4	95	42
ANOVA Selected	93	15.2	94	16.2
RF Selected	90	6.36	90	6.94
Spatio-Temporal	87	4.05	83	4.22

Figure 4.18 and Figure 4.19 presents the accuracy results for a random forest with 100 trees (optimal number) using the different feature selection subsets for the 3m and 10m tests respectively.

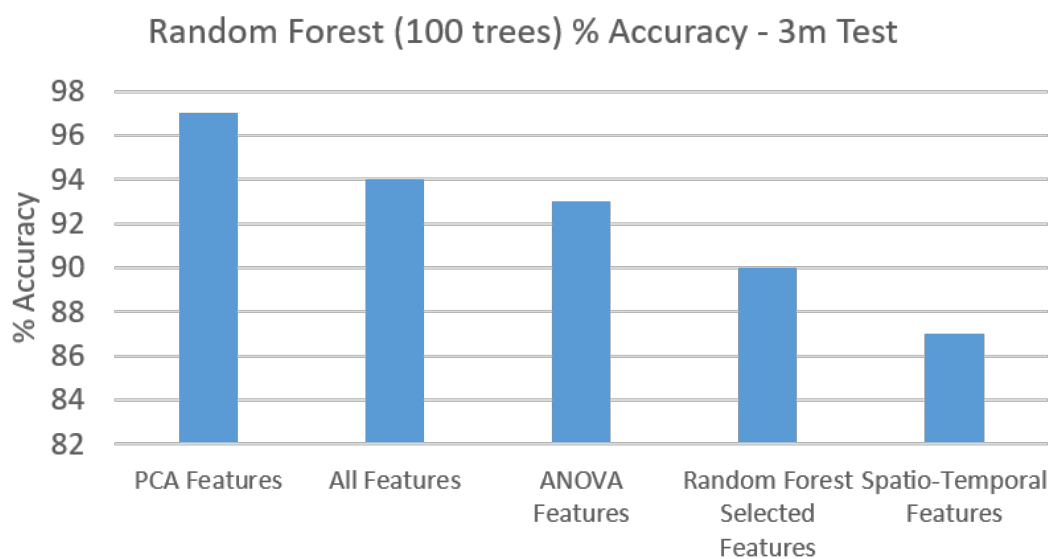


FIGURE 4.18: Random Forest Percentage Accuracy of discrimination between different PD stages for different feature selection subsets, for the 3m test

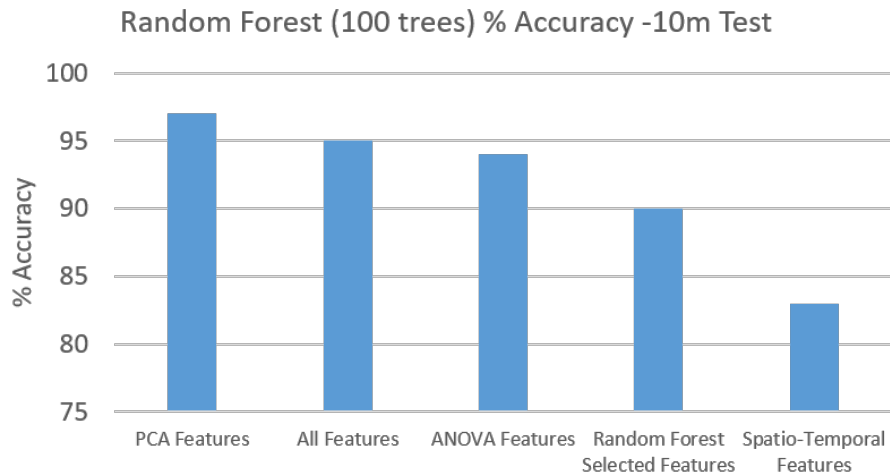


FIGURE 4.19: Random Forest Percentage Accuracy of discrimination between different PD stages for different feature selection subsets, for the 10m test

Figure 4.20 presents the comparative accuracy results for a random forest with 100 trees using different feature subsets for both the 3m and 10m tests

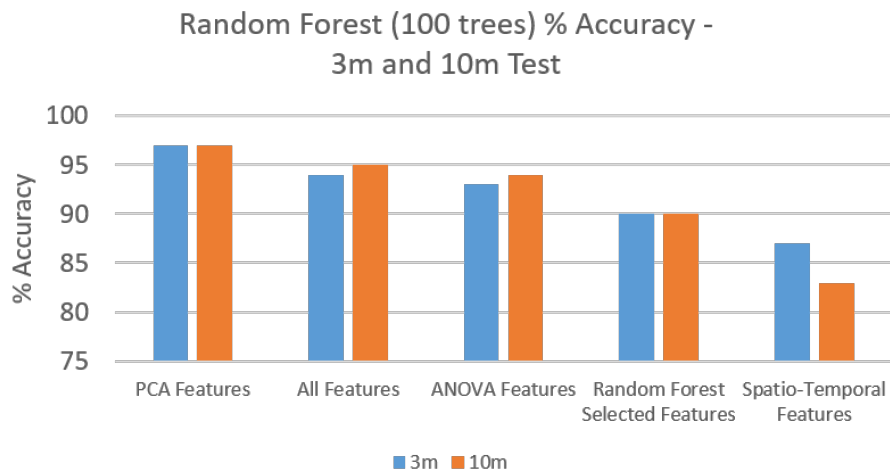


FIGURE 4.20: Comparison of the Random Forest Percentage Accuracy of discrimination between different PD stages for 3m and 10m Tests

The PCA selection method demonstrates the highest accuracy. The ANOVA feature selection is indicated as the best trade-off of high discrimination accuracy (only slightly lower than PCA), fast execution time and unlike the PCA, provides untransformed features. A thorough analysis of the trade-offs described is presented in Chapter 5.

The change in percentage accuracy as a function of execution times (s) of the different feature subsets is shown for 100 trees in Figure 4.21 and for 1-100 trees in Figure 4.22.

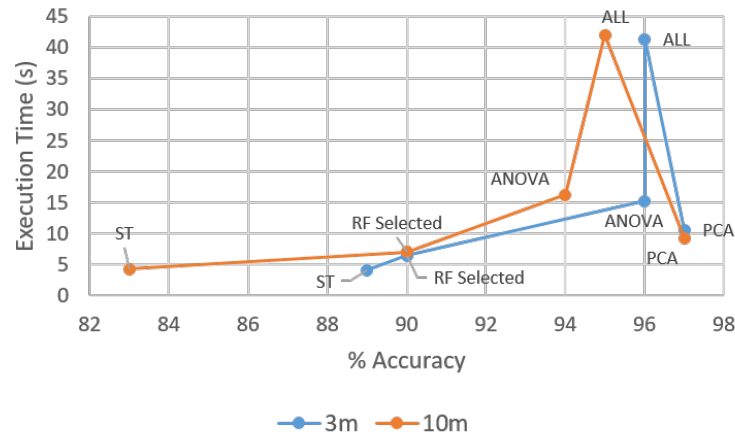


FIGURE 4.21: Execution time vs. 100 trees Classifier accuracy for the 5 feature subsets

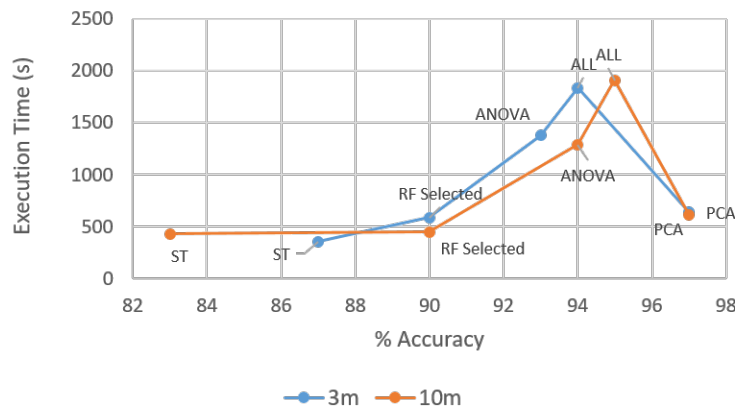


FIGURE 4.22: Execution time vs. 1-100 trees Classifier accuracy for the 5 feature subsets

The large spike in execution times for both 3m and 10m tests was for the classifier using the full 211-dimensional features space.

The confusion matrices using the ANOVA features for the 3m test which had a 93% accuracy and 10m test which had a 94% accuracy are shown in Figures 4.23 and 4.24 respectively. These confusion matrices provide the classification errors and highlights the nature of the errors, in terms of how many classes apart misclassifications were.



**Accuracy: 93.02%**

H & Y 0	100.0% 19	0.0% 0	3.8% 1	5.6% 1	0.0% 0	0.0% 0
H & Y 1	0.0% 0	100.0% 5	7.7% 2	0.0% 0	0.0% 0	0.0% 0
H & Y 2	0.0% 0	0.0% 0	84.6% 22	5.6% 1	0.0% 0	0.0% 0
H & Y 2.5	0.0% 0	0.0% 0	3.8% 1	88.9% 16	0.0% 0	0.0% 0
H & Y 3	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 13	0.0% 0
H & Y 4	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 5
	H & Y 0	H & Y 1	H & Y 2	H & Y 2.5	H & Y 3	H & Y 4

Target Class

FIGURE 4.23: Confusion Matrix for the 3m test using ANOVA selected features

**Accuracy: 94.19%**

H & Y 0	95.5% 21	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
H & Y 1	0.0% 0	100.0% 6	3.8% 1	0.0% 0	0.0% 0	0.0% 0
H & Y 2	0.0% 0	0.0% 0	88.5% 23	0.0% 0	0.0% 0	0.0% 0
H & Y 2.5	0.0% 0	0.0% 0	0.0% 0	94.4% 17	0.0% 0	0.0% 0
H & Y 3	4.5% 1	0.0% 0	7.7% 2	0.0% 0	100.0% 10	0.0% 0
H & Y 4	0.0% 0	0.0% 0	0.0% 0	5.6% 1	0.0% 0	100.0% 4
	H & Y 0	H & Y 1	H & Y 2	H & Y 2.5	H & Y 3	H & Y 4

Target Class

FIGURE 4.24: Confusion Matrix for the 10m test using ANOVA selected features

## 4.5.2 Regression for H & Y Prediction

Different regression models for the H & Y predictions were compared and the results are reported below.

### 4.5.2.1 Performance Metrics

The results of the 10-fold cross validation for the different regression models is shown in Table 4.7 and 4.8 for the 3m and 10m tests respectively. The values per performance metric (i.e. per row) were normalized to the range of 0-1, so that different models could be compared to each other. Consequently, in the new range a 0 is the minimum score for the performance metric, whilst a 1 is the maximum score for the performance metric. The assessment of these models is discussed in Chapter 5.

TABLE 4.7: Normalized Regression Performance Metrics for the 3m test, where the columns are the different regression models and the rows are the different performance metrics

	Stepwise	SVM	Robust	Boosting	ANN	RF
MAE	1	0.014	0.038	0.032	0	0.003
SSE	1	0.000087	0.00061	0.00031	0	0.000021
MSE	1	0.000089	0.00063	0.00033	0	0.000018
MAPE	1	0.0063	0.0289	0.0174	0.0036	0
RMSE	1	0.0048	0.02329	0.01570	0	0.0011
RSQ	1	0	0.00085	0.0003	0.00011	0.00011
CORR	0.205	0.823	0	0.647	1	0.971
MI	1	0.708	0.625	0.416	0.583	0
Cosine Sim	0	0.931	0.681	0.841	1	0.977
IA	1	0	0.227	0.5	0.409	0.227
Fractional bias	0	0.884	1	0.971	0.927	0.884
Geo mean bias	0	0	0.392	0.392	0.428	1
NMSE	1	0.00187	0.0359	0.0154	0	0.0009
Geo var	0	0	0.393	0.393	0.428	1
Fac2	0	0.94	0.79	0.76	1	1

TABLE 4.8: Normalized Regression Performance Metrics for the 10m Test, where the columns are the different regression models and the rows are the different performance metrics

	Stepwise	SVM	Robust	Boosting	ANN	RF
MAE	1	0.028	0.161	0.163	0.033	0
SSE	1	0.008	0.0493	0.0461	0.0074	0
MSE	1	0.0083	0.0482	0.0438	0.007	0
MAPE	1	0	0.1200	0.1903	0.0437	0.0025
RMSE	1	0.0313	0.1390	0.1241	0.0196	0
RSQ	1	0	0.038	0.087	0.0259	0.0089
CORR	0	0.314	0.155	0.550	1	0.939
MI	1	0	0.792	0.0604	0.548	0.319
Cosine Sim	0	0.956	0.833	0.858	0.975	1
IA	1	0	0.409	0.675	0.263	0.434
Fractional bias	1	0.0059	0.0029	0.00236	0.0022	0
Geo mean bias	0	0	0.925	0	1	0
NMSE	1	0.0026	0.0393	0.0444	0.00194	0
Geo var	0	0	0.925	0	1	0
Fac2	0	0.97	0.90	0.54	0.96	1

#### 4.5.2.2 Relationship between feature values and H & Y scores

Scatter plots and the best fit line of the H & Y scores vs. feature values was performed for all features. Figures 4.25-4.29 illustrate examples of 5 features, for the 3m test. The results for the 10m test were similar to the 3m test. The plots show the values for the H & Y stages 1-3. H & Y 0 and 4 were omitted since the values are significantly larger or smaller and mis-scale the figure. The H & Y 1-3 are also the PD severity stages that are hardest to quantify and would greatly benefit from a finer H & Y scale.

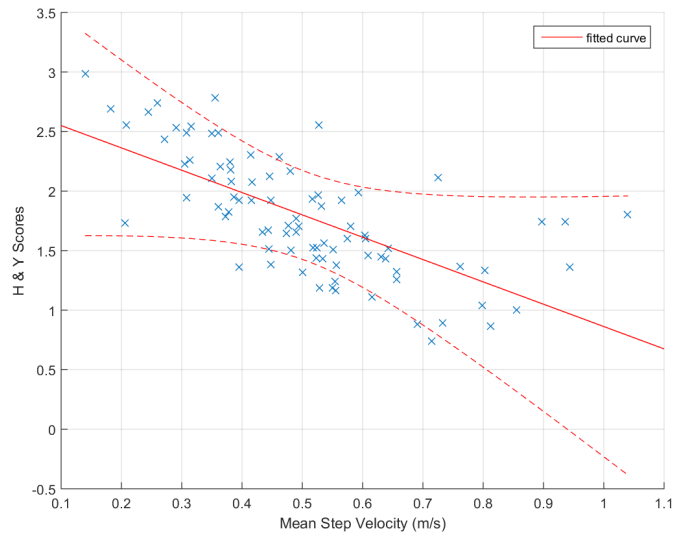


FIGURE 4.25: H & Y vs Mean Step Velocity. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound

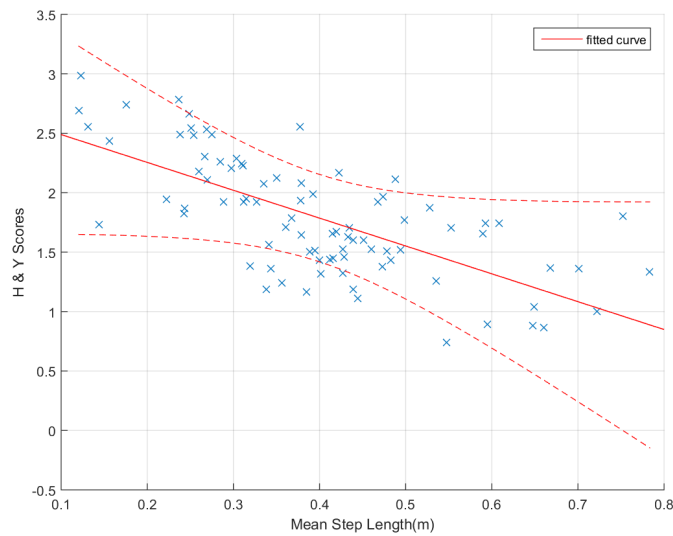


FIGURE 4.26: H & Y vs Mean Step Length. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound

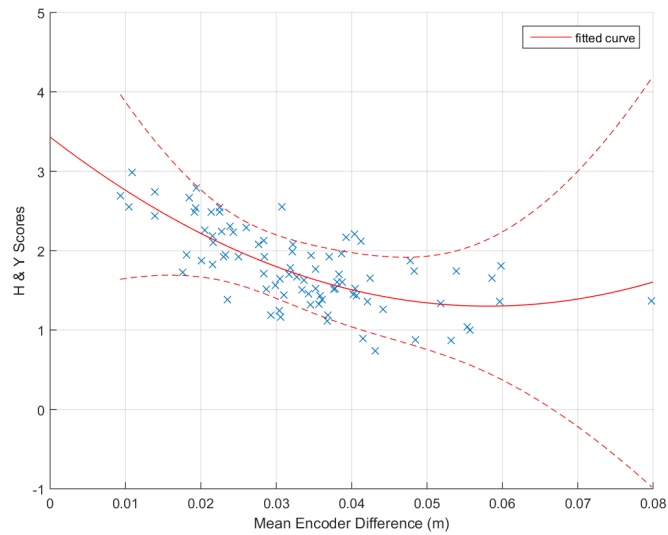


FIGURE 4.27: H & Y vs Mean Encoder Difference. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound

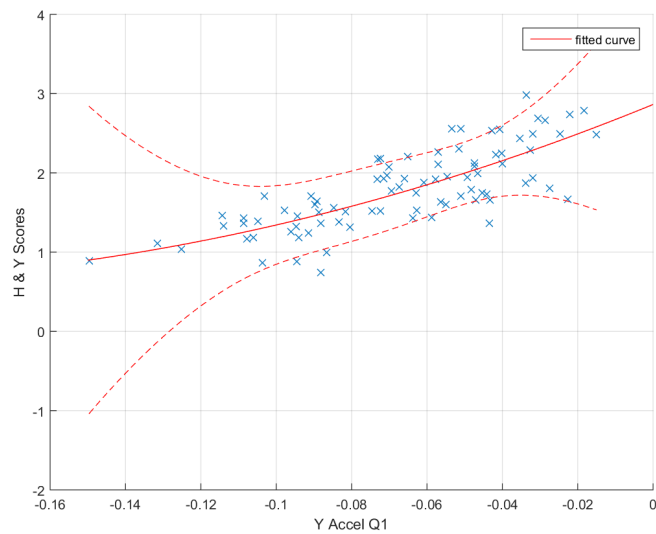


FIGURE 4.28: H & Y vs Y accel Q1. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound

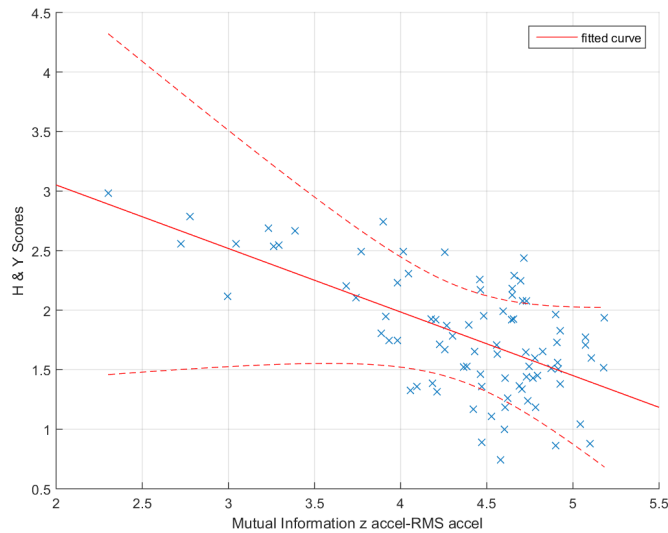


FIGURE 4.29: H & Y vs Mutual Information z accel-RMS accel. The solid red line is the best fit line and the dotted red lines is the 95% confidence bound

The line fit analysis indicates that different features have different trends in terms of their relation to H & Y scores. The examples of the 5 figures demonstrate there are both negative and positive linear relationships, as well as negative and positive exponential relationships. However, different as the mathematical relationships are, most of the features exhibit change in finer increments than the H & Y score. These trends are discussed fully in Chapter 5.

#### 4.6 Statistical Analysis using CI's and relation between CI's and selected features

The statistical analysis made use of CI's and computed the overlap between the groups. The CI features (which met the significance criteria outlined in Section 3.9), were compared to the ANOVA selected features and the features selected by the random forest. These results are presented in the forthcoming sub-sections.

### 4.6.1 Confidence Intervals

The overlap between the CI's of the 6 different H & Y groups was computed using Equation 3.8 for all the features. This method estimates the significance of the difference between group means [72, 73]. The significant difference criteria, as defined by [73] are: No overlap indicates a 99% significance, less than 50% overlap indicates 95% significance and more than 50% overlap indicates no significance.

Figures 4.30 and 4.31 are examples of features that demonstrated no overlap greater than 50%. The groups compared in the figures had either a 95% or 99% significance for the difference between the H & Y groups.

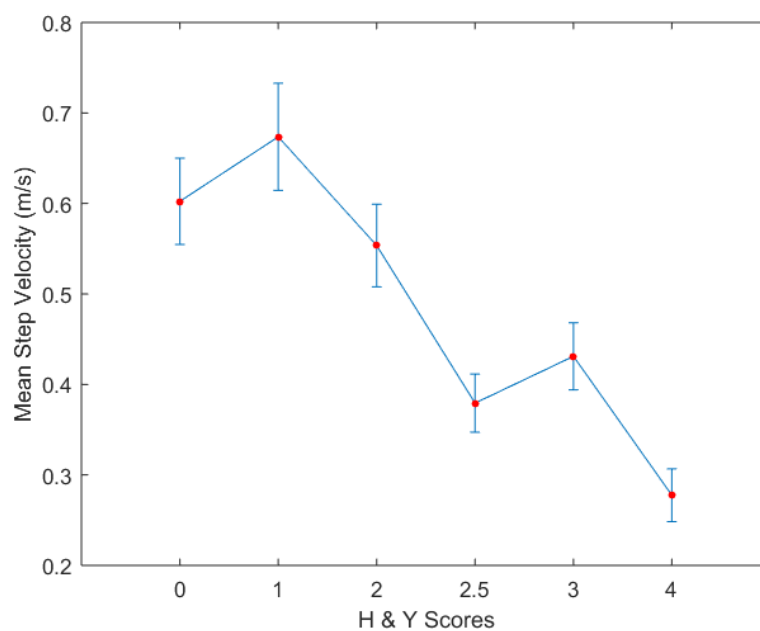


FIGURE 4.30: CI Plot for Mean Step Velocity, which shows no group overlap of greater than 50% and either a 95% or 99% significance of difference between H & Y group pairs

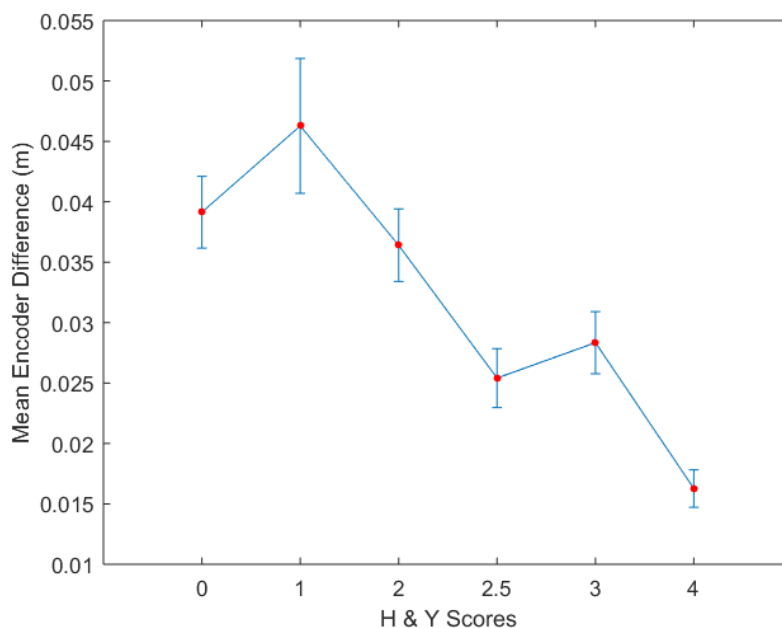


FIGURE 4.31: CI Plot of Mean Encoder Difference, which shows no group overlap of greater than 50% and either a 95% or 99% significance of difference between H & Y group pairs

Figures 4.32 and 4.33 are examples where the feature had one instance of overlap greater than 50%, indicating no significant difference. This was between H & Y 0 and 1 in Figure 4.32 and H & Y 2.5 and 3 in Figure 4.33. The rest of the comparisons between H & Y groups had 95% and 99% significance of difference. It was found that none of the features had greater than 50% overlap (no significant difference) for every comparison between H & Y group pairs. The implication is that even the less discriminative features demonstrated some discrimination between different H & Y stages, even if it was a coarse difference i.e. between the lower severity groups (H & Y 0, 1) and the higher severity groups (H & Y 3, 4)



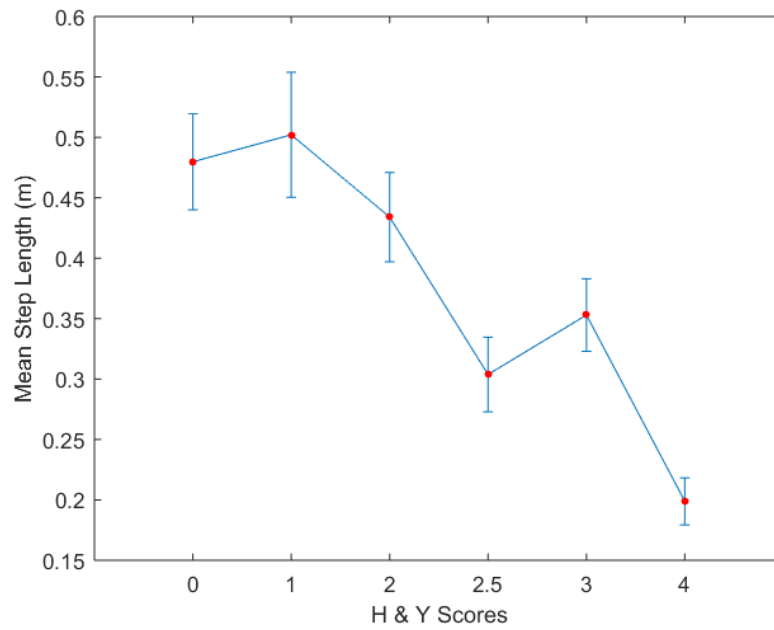


FIGURE 4.32: CI Plot of Mean Step Length, which shows greater than 50% overlap for H & Y 0 and 1

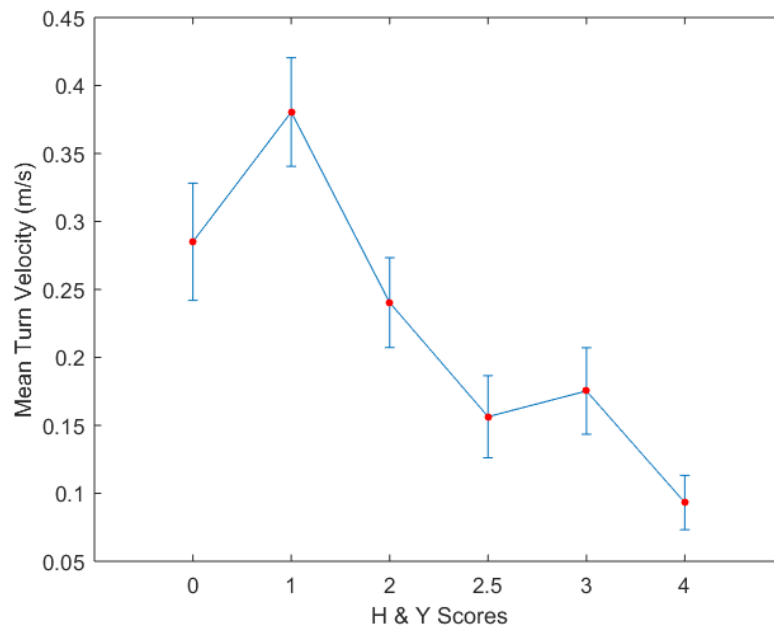


FIGURE 4.33: CI Plot of Turn Velocity, which shows greater than 50% overlap for H & Y 2.5 and 3

## 4.6.2 Feature Comparison

The features which met the CI criteria (as described in Section 3.9.2 based on the pairwise comparison of overlap between H & Y groups) were then compared to the ANOVA selected features and to the random forest selected features.

Table 4.9 lists the percentage of features that were similar/overlapping, in being discriminative between the H & Y groups, in the CI significance analysis and the ANOVA selection and the random forest selection methods. Figures 4.34 and 4.35 demonstrate these similar features by category, for the ANOVA and the random forest selection respectively. Table 4.10 lists these common features between the features subsets which allowed for the highest accuracy of discrimination between PD stages. The features that appeared in each pair of selection methods are highlighted in cyan in Appendix B, Table B.3 for the ANOVA features and Table B.4 for the random forest features.

TABLE 4.9: Percentage of features overlapping between the significant CI features and the ANOVA and random forest selected features

Comparison	% Features Overlapping
ANOVA Features - CI Features	80%
Random Forest Features- CI Features	63%
Random Forest Features - ANOVA Features	87%

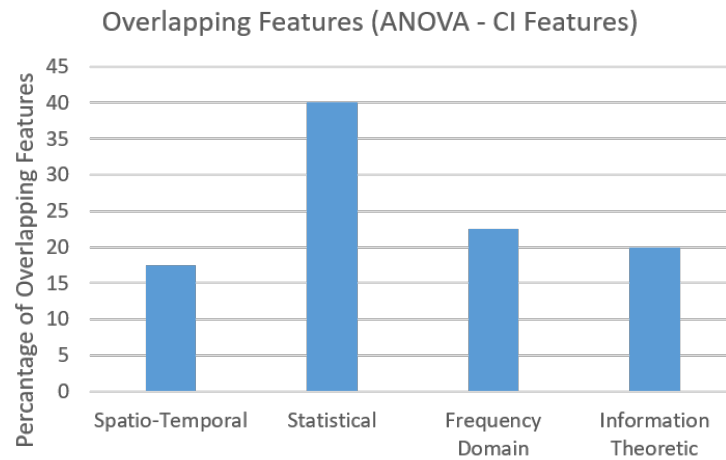


FIGURE 4.34: Percentage of overlapping features by category between the ANOVA features and CI Features

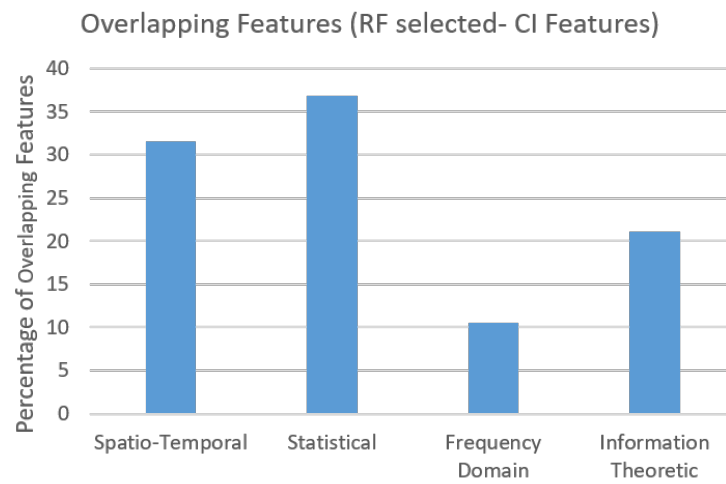


FIGURE 4.35: Percentage of overlapping features by category between the ANOVA features and Random Forest Selected Features

TABLE 4.10: Common features between the features subsets which allowed for the highest accuracy of discrimination between PD stages

Mean Encoder Difference	Radius Q1
Number of Steps	y Accel Q3
Total TUG time	RMS Accel Median
Standard Deviation Turn Velocity	x Accel IQR
Mean Turn Velocity	azimuth Entropy
Walk Ratio	y Accel MSE
y Accel half power	y Accel Variance
y Accel 99% power	Correlation z Accel - Azimuth
y Accel Q1	Mutual Information z Accel- RMS Accel
RMS Accel Entropy	Mutual Information x Accel - y Accel
Total Harmonic Distortion y	Mutual Information y Accel- Radius
Mean Step Length	Mutual Information z Accel- Radius
Mean Step Velocity	Standard Deviation Accel

# CHAPTER 5

## DISCUSSION AND CONCLUSIONS

This chapter presents a detailed discussion and analysis of the results presented in Chapter 4 and conveys concluding remarks of the dissertation.

### 5.1 Key Findings

The following key findings are discussed: (1) The answer to the research question, (2) Patterns in different PD severity groups and (3) Implications of the signal pre-processing methods. Finally, possible applications of these findings, their limitations and suggestions for future work are presented.

#### 5.1.1 Research Goal

##### 5.1.1.1 Extent of Discrimination between PD stages

The primary focus of this research was to assess the extent to which kinematic signals obtained from walker-mounted sensors could provide an automated and reliable discrimination between PD stages. The results that lead to the answer of this question are outlined in the next sections.

The research question and hypothesis has been answered by the result that a random forest classifier of a 100 trees, which used ANOVA-selected features can reliably discriminate between the different stages of PD with an accuracy of 93% and 94% for the 3m and 10m walking tests respectively. The confusion matrices in

Figure 4.23 and 4.24 highlight the type of errors obtained and indicate that errors typically result in mis-classification to the adjacent PD stage (i.e. either to the H & Y stage immediately above or below the actual class). The highest number of mis-classifications occur between the H & Y 2 and 2.5 stages.

The two walking tests employed in the data acquisition experiment – 3m and 10m tests yielded similar discrimination performance of 93% and 94% respectively. This result implies that only one of them is needed, which can be useful from the clinical point of view: where the assessment can be shortened to one of the tests. This result is similar to an earlier study conducted with the PAM-MII instrumented walker which showed that a 10m walking test provided better discrimination between PD patients and controls [4]. The earlier study showed a sensitivity and specificity between PD and controls were 91% and 95% for the 3m test and 96% and 100% for the 10m test. It must be noted, however, that the difference in accuracy between the 3m and 10m tests is different in this study when compared to the aforementioned earlier study conducted with the PAMM instrumented walker [4]. Possible reasons for the accuracy difference may be that the earlier study performed a binary classification between PD patients and controls, where the differences between the classes are larger. Whereas, this study involved more intricate discrimination hence the reduced accuracy.

#### **5.1.1.2 Best Feature Subset for Discrimination**

Kinematic signals and the usage of spatio-temporal parameters from these signals have been used to assess gait abnormalities in both qualitative and quantitative studies [4, 38, 39, 42, 43, 82, 83]. The present study aimed to improve the analysis performed in these studies, by assessing a wider variety of features, as well as incorporating a larger number of features (211 overall). Moreover, a comparative analysis of the features in terms of their category (spatio-temporal, frequency domain, statistical and information theoretic) was performed.

The large number of features necessitated feature selection to reduce dimensionality for the classification. Four different feature selection methods were applied

to the entire feature set and the four feature subsets were compared. The selection methods were: spatio-temporal, PCA selected, ANOVA selected and random forest selected features. These 4 features subsets were compared based on the discrimination obtained using a random forest classifier. The random forest was used, since unlike other machine learning methods explored, it allowed features important to the discrimination to be determined [64, 65].

The comparison of the 4 selection subsets yielded that the spatio-temporal features subset provided the worst discrimination accuracy, as well as the largest difference between the discrimination accuracies for the 3m (87% accuracy) and 10m (83% accuracy) test. The spatio-temporal subset was important to study since these features could be better communicated to and accepted by clinicians and are a part of the observation process used in clinical assessments [82–84]. These spatio-temporal features that may correspond to the qualitative, “observable by clinicians” features of gait, yielded poor performance in the quantitative discrimination, implying that the other categories of features are necessary to improve accuracy. However, the poor performance of the classifier using the 13 spatio-temporal features can be attributed to the fact that this subset of features included a smaller number of features when compared to the other subsets of features which were obtained by mathematical selection methods and included a mixture of spatio-temporal, statistical, information theoretic and frequency domain features. Also, the human visual perception, and particularly specialist’s, most probably pick up many other cues when observing these spatio-temporal characteristics of gait.

The discrimination accuracy results in Table 4.6 revealed that the full feature set, PCA selected feature set and ANOVA selected feature set produced similar, high discrimination accuracy (greater than 90%) for both the 3m and 10m tests (Figure 4.20). The objective of a further comparative analysis was to ensure fast classifier execution time and features which are untransformed as to not obscure the individual features and their units. Using the full feature set maintained the units of the features, however the execution time was 3 times slower than PCA features and 4 times slower than the ANOVA-selected features as shown in Table 4.6. The PCA feature subset had a short execution time, however this selection

method results in features which can no longer be individually identified, nor easily interpretable. The ANOVA selected features addressed both short-comings by providing a short execution time, whilst not transforming the features.

The random forest selection method yielded a feature subset which produced a 5% lower discrimination accuracy compared to the PCA features and ANOVA selected features. This may imply that for this data, the internal feature selection performed in the random forest during the classification is less accurate compared to PCA and ANOVA.

### **5.1.1.3 Finer H & Y scores using regression models**

The PD stages of the patients, used as the labels of the classifier, are given by the H & Y scale [15], which is determined by neurologists through clinical assessment. The scale has the following discrete scores: 0- Healthy, non-PD patient ; 1- Unilateral disease, the patient exhibits minor or no disability ; 2- Bilateral disease without impairment of balance ; 3- Bilateral disease. The patient exhibits mild to moderate disability and instability; 4- Severe disease, however the patient can still walk unassisted; 5- The patient is bed ridden or confined to a wheelchair without assistance (Not included in the experiment as discussed in Section 3.1.4 ). Although the original scale consists of 5 discrete numbers (1 to 5), clinicians already try to make it finer by using the Modified Hoehn & Yahr Scale [85] which adds a "1.5" and "2.5" score. Even with this modified scale, the Movement Disorder Society found that 69% of clinicians (of 1593 who participated in the survey) still found that the PD categories were too broad to characterise severity accurately [85]. Although the scale deals mainly with differences in stability, a finer gait scale could be useful for clinicians.

Studies have used the concept of regression models to predict PD scores. However, typically speech derived features (e.g. vocal frequency, jitter, amplitude etc) are used to predict UPDRS scores [86–88]. Gait features were also used along with regression by [45, 46], to predict a PD score which was then correlated to UPDRS. The regression models in this study filled this gap of H & Y regression, by using



walker derived gait features to predict H & Y scores. The results indicate a viability to fill values between the discrete, integer H & Y scores and thus to produce a finer "gait-related H & Y scale". This finer scale was quantitatively assessed by comparing the model predictions against the clinician's predictions, these included "half scores" of 2.5. The best fit model aimed to minimize the error metrics, whilst maximizing the similarity metrics. This mirrors the criteria used by [88], where the optimized UPDRS regression model minimized error and maximized similarity.

The regression models which met these best-fit criteria were the SVM, random forest and ANN. All 3 methods were able to provide a good fit to the clinical score as conveyed by the low normalized error metrics. The success of these non-linear models indicates that non-linearity is necessary to accurately capture the dynamics of the features. This corroborates both [86] and [87] in demonstrating the best performance using non-linear regression models to predict PD UPDRS scores.

Various performance fit metrics were employed to ensure both low error and high similarity measures. When comparing the SVM, random forest and ANN it was found that for similarity metrics, the SVM had lower correlation metrics when compared to the random forest and ANN. Both the random forest and ANN had low error metrics, as well as a high similarity metrics. Neither the ANN nor the random forest regression models had distinct superiority when compared to each other, thus the best fit model was based on architectures. The random forest model has an advantage, as unlike the ANN, it is not a black box model.

The results of the random forest regression model are conveyed in Figures 4.25-4.29. The graphs suggest that a robust model can be developed to predict H & Y scores to a finer degree. This is additionally validated when comparing the predicted H & Y scores which highlight the inter-subject difference (for subjects from the same original H & Y class). That being said, the predictions closely match the labelled H & Y scores (i.e. no discrepancy in the general PD stage), as the error scores between the two are low ( $\text{MAPE} \leq 26\%$ ) and the similarity scores

are high (e.g. cosine similarity  $\geq 0.88$ ). However, this "gait quality scale" would require further testing of a larger cohort of patients from all H & Y stages and an evaluation of its benefits in aiding clinical assessment.

The stepwise linear model, which is widely used in medical research [89], performed inadequately, based on the large values for all the error metrics (e.g. MAE, SSE, MAPE) and the low values for all the similarity metrics (e.g. cosine similarity, FAC2, correlation). This matches the findings of [89] who in epidemiological research found that stepwise models, produce high errors when the data has a large number variables and multicollinearity - which is the case for the features in this study.

#### **5.1.1.4 Most significant features for PD discrimination**

The agreement between the selected features of the random forest and the ANOVA selection amounted to 26 features (87% agreement between the 2 sets). These features which, the techniques implied as significant to the PD stage discrimination are listed in Table 4.10. The ANOVA and random forest selected features were compared as they produced the highest discrimination accuracy and untransformed features. The PCA features were not considered in this comparison, as the principal components (PCs) are transformed as a linear combination of features.

The features in Table 4.10 yield an insight on the instrumentation. The finding that force features do not play a significant role implies that collecting force sensor data may be redundant for future experimentation. Rather the accelerometer and encoder data are important.

The analysis of the feature categories may also provide an insight into PD gait characteristics which can be used in future research. Previous studies focused on spatio-temporal features, particularly on step time, step length and step velocity, derived from either wearable sensors ([38], [48]), as well as instrumented walkers [4, 39].

The significant features in this study— those that offered the best H & Y discrimination - however contain features from all the feature categories: spatio-temporal (4/26), frequency domain (2/26), information theoretic (9/26) and statistical features (11/26). Although, step length and step velocity are among the significant features in this analysis, corroborating earlier studies, the percentage of information theory and statistical features in the selected feature set was much larger, as conveyed in both Table 4.10 (important features according to all 3 selection methods in the study) and Figure 4.9 (ANOVA features—the best selection method according to this study analysis). This finding implies that PD gait analysis of PD should not be limited to the observable, time-based features, but need also include the more abstract and mathematical features. Including all four feature categories in future research on PD gait analysis, whether using similar or different sensors data, may yield additional insight into PD gait deterioration with disease severity.

The larger feature sets, resulting from an initially larger feature space, compared to those used in previous studies, and the expanded categories are due to the fact that many of these features are either unique to this study, due to the data acquisition procedure (e.g. turning features, walk ratio etc) or were proposed in more recent research (post-2016)[45, 46]. Among these features are the statistical features such as Q1, Q3, IQR, median and information theory features such as entropy and mutual information. The results support a recommendation to use the procedure as well as the additional features in future studies.

### **5.1.2 Patterns in different PD severity groups**

The patterns found for the different PD groups will be discussed in the following sections. Firstly, inter-group variability and similarity will be discussed. Secondly, it will address the statistical model of a generalized pareto probability distribution. Thirdly, the within-group CI analysis and overlap validation of the features will be discussed.

### 5.1.2.1 Inter-Group variability and similarity patterns

The analysis of the CI's for different features, as well as, the probability distributions (e.g. Figures 4.12-4.15) illustrated that although accurate discriminations were achieved between all PD stages, certain PD stage pairs are more closely related to each other than others. The findings are discussed in the following points.

(i) *H & Y 0 and 1 are closely related*: Control subjects and subjects with H & Y 1 have differences but are still closely related, as per the CI plots in Figures 4.30 - 4.33. This finding was corroborated by the neurologist in the clinic as reasonable, since H & Y 1 is unilateral disease - with minimal or no functional disability, which may be similar to controls. Conversely there are heterogeneous symptoms in this stage, where certain patients may display different combinations of symptoms and which may not pertain to gait [15].

(ii) *H & Y 2 is significantly different from H & Y 1 and 2.5*: The finding that H & Y 2 is different from H & Y 1 is understandable as the disease becomes bilateral at H & Y 2 and produces more severe symptoms, among which are the gait related symptoms. An interesting finding was the difference in CI's between H & Y 2 and 2.5, as it highlights the benefit of a finer H & Y scale as the disease progresses. The difference between H & Y 2 and 2.5 is further corroborated by [90] who found differences between H & Y 2 and 2.5 when assessing features not used in this study such as the ground reaction force (which is the force exerted on the body per footfall typically measured using wearables).

(iii) *H & Y 2.5 and 3 are closely related*: H & Y of 2.5 and 3 were more closely related than H & Y 2 and 2.5. This observation, however, does not pertain to all the features used for the discrimination: Some features had CI overlap (e.g. turn velocity), whilst others had no overlap (e.g. mean step length). Interestingly, a large percentage of features with no CI overlap appeared in both feature subsets, selected by the ANOVA and random forest as shown in Table 4.9.

(iv) *H & Y 4 is significantly different from H & Y 3*: The difference between H & Y 3 and 4 is reasonable as H & Y 4 patient's manifest with more significant instability and gait impairment. Additionally, the probability distributions shown

in Figures 4.12-4.15 further demonstrate the significant difference that the H & Y 4 stage has compared to other PD stages.

The class imbalance of the number of subjects in each of the different H & Y groups was noted in this study, especially in the smaller number of H & Y 4 subjects. However, the analysis yielded that the H & Y 4 patients, exhibit marked differences in their gait features compared to the other PD stages, that are even easily discernible in visual inspection of their gait recorded signals. This discernible difference is also conveyed in the CI plots in Figures 4.30-4.33, as well as, in the probability distributions in Figures 4.12-4.15. This discernible difference mitigates some of the effects of the imbalance. Rather, it is H & Y 2 and 3 that are typically more challenging to distinguish from each other and thus required more test subjects (as was the case for this study).

#### **5.1.2.2 Statistical model: Most prevalent probability distribution**

Table 4.3 and Table 4.4 for the 3m and 10m tests show that the most prevalent distribution fit to the features is the generalized pareto distribution. Moreover, a greater number of features are modelled by that distribution as the H & Y score increases. The percentage of features out of the total feature set, that followed the generalized pareto was under 50% for the earlier PD stages (H & Y 1,2) and greater than 80% for the more severe PD stages (H & Y 3,4). It was noted that there is an exception in this trend for H & Y of 2.5, where the percentage of features with that distribution drops. This finding may imply that the H & Y stages of 0 and 1 are not well characterized and that the H & Y 2.5 may be heterogeneous in the characteristics of the features when compared to the other H & Y groups.

The second and third best fit distributions for the features varied and include: t-location scale, Rayleigh (special version of the Weibull distribution), Generalized Extreme Value, Loglogistic and Exponential. All these distributions except the Exponential distribution are heavy-tailed distributions like the generalized pareto distribution.

It was found that the features which changed their best fit probability distribution, as PD severity increased were mainly information theoretic (mutual information, correlation, entropy) and statistical (MSE, IQR, standard deviation of acceleration). This demonstrates the value of the information theoretic and statistical features to the discrimination. That being said, the spatio-temporal features of the number of steps and TUG time also exhibited this trend.

### **5.1.2.3 Implication of the generalized pareto distribution**

Analysing the generalized pareto probability distributions for each different PD stage as per Figures 4.12-4.15 shows the differences between the PD stages and how the parameters of the distribution changes based on the PD stage. Whilst some class differences are observable in the distributions, it did also further highlight the similarities between stages. The following have differences between the distributions but are still closely related to each other: H & Y 0 and 1 and then H & Y 2, 2.5 and 3. H & Y 4 on the other hand is completely different from the other stages. This corroborates the discussion in Section 5.1.2.1. That being said, there is a limitation in the small sample size of H & Y 4 subjects and this finding needs to be verified with a greater number of subjects in this group.

The prevalence of heavy-tailed distributions and the pronounced, narrow peaks and long tails for the generalized pareto distribution, as can be seen in Figures 4.12-4.15 implies that there is a large probability of a disproportionate value in the sample [91]. The term disproportionate refers to a feature value which is significantly different from the other values within the same class (PD stage). This is an understandable finding as at the same stage of PD, certain symptoms manifest strongly for certain patients and not for others [15]. This would lead to a disproportionate value for the specific feature, which these heavy-tailed distributions model. These preliminary observations need to be validated using a much larger number of patients, in all groups.

The distribution findings was then compared to a study by [92, 93]. They proposed that a log-normal model of the velocity signal (obtained from inertial sensors) could

be used to estimate the cranio-caudal signature during a turn (distinct motion of the head followed by the trunk). This was then applied to PD to assess the viability of a lognormal model in cases of mobility disorders [93]. The lognormal distribution was shown to adequately model the cranio-caudal feature [93]. Thus, the distribution could be used to estimate the cranio-caudal feature value and may even represent the generating process associated with the neuromuscular system [92, 93].

The lognormal distribution was then contrasted with the generalized pareto distribution model found through the current study. Firstly, both the lognormal and generalized pareto distributions are heavy tail distributions. Secondly, [94] states that log-normal and generalized pareto distributions are alternative distributions to describe the same quantities. This validates the generalized pareto distribution model fitted to PD features and the distribution could also be used to estimate feature values by sampling values from the distributions for each PD stage.

#### **5.1.2.4 CI Analysis and overlap validation**

The analysis of the pairwise CI overlap of H & Y groups for each individual feature, served as a statistical validation of the class separability. The results in Table 4.9 indicate that the random forest selected features are worse than the ANOVA selected features: Only 63% of the random forest selected features met the CI significance criteria compared to 80% of the ANOVA selected features. These findings may serve as another indication that the ANOVA selected feature set enables a better discrimination, as indicated by the machine learning classification performance, and a validation, by using a different statistical analysis

Finally, the higher discrimination accuracy using the ANOVA selected features, when compared to the random forest selected features may be explained using the within-group CI results. The greater percentage overlap between the CI significant features and the ANOVA selected features implies that ANOVA feature selection yielded features which had greater discrimination ability and may account for the superior discrimination accuracy when using these features.

### 5.1.3 Implications from signal pre-processing methods

A challenge in this research was the instrumentation used. Although chosen for its low-cost and enhanced usability, the inexpensive sensors had large levels of signal noise, as well as measurement imprecision. Moreover, the mounting of the sensors on a walker instead of directly on the body, made basic gait parameters, that are easily obtained from wearables much harder to compute. Specifically, these properties, necessitated an accurate detection of footfalls, the basic parameter of the gait cycle, to ensure reliable feature derivation.

#### 5.1.3.1 Accurate footfall detection using EMD and peak detection applied to the z-axis accelerometer

Accurate footfall detection is necessary to quantify human walking patterns especially in assessment of gait disorders [76]. The novel  $z$ -axis accelerometer algorithm proposed in this study which made use of EMD and peak detection showed superior accuracy (86% footfall detection accuracy) when compared to a previously reported algorithm by [39]. This work was used for performance comparison, as it was one of few similar studies which used exo-body sensors rather than a wearable.

It was also found as per the results in Table 4.2, that IMF 3 was the most suitable IMF to reconstruct the signal for both implemented algorithms, obtaining the best results by removing high frequency noise and baseline drift, whilst maintaining the peaks in the signal, which were not as easily discernible in the raw signal. The discernible peaks facilitated footfall detection and the subsequent extraction of relevant gait features. Reconstruction from IMF 1 performed poorly as it is simply the original noisy signal. Reconstruction from IMF 2 did remove signal noise and had an accuracy score of 82%, however reconstruction from IMF 3 still yielded superior results. Reconstructing from IMF 4 also performs poorly, by discarding the first 3 IMFs, thus losing important features and signal information.



### 5.1.4 Main Findings Summary

The following are the main findings of the research:

1. PD stages can be discriminated using kinematic signals obtained from a low-cost and noisy walker mounted sensor data.
2. A subset of the most relevant features, derived from this data, that provided the best discrimination between PD stages was assessed.
3. A generalized pareto distribution was indicated as the best fit probability distribution for the majority of PD features.
4. Confidence intervals and probability distributions validate the machine learning methods' class separability between different PD stages.
5. A novel quantitative footfall detection algorithm achieves an accuracy of 86% in detecting footfalls and improved on previous studies.

### 5.1.5 Application of the findings

This study demonstrated the feasibility that walker mounted sensors can reliably discriminate between PD stages. This capability opens up the potential for health-care professionals to use such device for PD progression evaluation, providing a patient-centred care with minimal cost and labour. Moreover, the low-cost and simplicity may enable patients to use this device at home and integrate in an eHealth monitoring scheme.

In addition, the novel footfall detection algorithm and the statistical model may be used as tools for other studies. Footfall detection is not only relevant to PD and could be used in gait assessment (using similar instrumentation) for other gait and movement disorders (such as elderly after falls or people after surgery). The probability distribution is shown to model feature values for the different PD stages. These distributions could be used to estimate values for features by sampling a value from the probability distribution for the PD stage.

### 5.1.6 Limitations and Future Work

1. While the footfall detection algorithm was quantitatively validated using video data, its validation was done using control subjects who mimicked PD gait and was therefore limited. Future studies could validate the algorithm using a ground truth of video with actual PD patients.
2. The regression model proved the possibility of a finer H & Y scale, that pertains to gait. However, a significant limitation was that there was no ability to validate the finer, intermediate scores against ground truth scores. A further study could be conducted, where the finer scores are assigned by healthcare professionals, to analyse the significance of this scale from a clinical perspective.
3. The H & Y scores, which served as labels to the data, were assigned in a single clinic, by the same neurologists, which may produce a bias. A multi-center study that would employ different clinicians can assess that there is no bias in the H & Y scores.
4. In order to validate the generalized pareto distribution as a valid statistical model, a larger study with more subjects in each H & Y group is needed.

## 5.2 Conclusion

The primary goal of the research was to determine the extent to which kinematic signals obtained from walker mounted sensors could be used to provide an automated, quantitative and reliable discrimination between PD progression stages. The signal processing and machine learning methods applied to the noisy walker mounted sensor data achieved a 93% (3m) and 94% (10m) accuracy in discriminating between different H & Y stages of PD.

The most relevant features that provided this discrimination were obtained and may provide a quantitative insight into gait changes as the disease progresses. The majority of these features were information theoretic and statistical.

The research also indicated a generalized pareto distribution as a best fit distribution of PD features, as manifested in gait, as well as the possibility of estimating feature values from the distribution. A novel footfall detection algorithm improved upon other algorithms.

There is scope for future improvements to this study, particularly in terms of validating the algorithms and the findings with larger cohorts of patients. However, this research has shown that optimized signal processing and machine learning can overcome the limitations of low-cost and noisy sensors and offer good discrimination based on their acquired signals. The study also implied that quantitative methods have the potential to assist medical practitioners in the evaluation of their PD patient's disease stage.

In conclusion, this research confirmed the hypothesis that signal processing and machine learning techniques applied to kinematic signals acquired from low-cost and noisy sensors mounted on an exo-body walker can provide a quantitative measure of PD progression and discriminate between clinical disease stages.

# **Appendix A**

## **Ethics Approval**

This appendix provides all documentation related to ethics clearance for re-use of data. The following pages contain copies of the documentation.


UNIVERSITY OF THE  
WITWATERSRAND,  
JOHANNESBURG



R14/49 Mr Nabeel Seedat et al

**HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)**

**CLEARANCE CERTIFICATE NO. M180202**

**NAME:** Mr Nabeel Seedat et al  
**(Principal Investigator)**  
**DEPARTMENT:** School of Electrical and Information Engineering  
**PROJECT TITLE:** Quantitative Analysis of Parkinson's Disease Progression  
Using Signal Processing and Machine Learning  
**DATE CONSIDERED:** 23/02/2018  
**DECISION:** Approved unconditionally  
**CONDITIONS:**  
**SUPERVISOR:** Prof Vered Aharonson  
**APPROVED BY:**   
Prof C Penny, Chairperson, HREC (Medical)  
**DATE OF APPROVAL:** 26/02/2018

**This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.**

**DECLARATION OF INVESTIGATORS**

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report.** The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in February and will therefore be due in the month of February each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

  
Principal Investigator Signature

\_\_\_\_\_  
Date

**PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES**

**DECLARATION:****Adherence to HREC (Medical) Ethics Application Terms and Conditions**

I, the undersigned, hereby declare that I have not collected data/ done secondary data analysis or any other form of research, prior to obtaining clearance certificate from the HREC (Medical) for study no: M180202.

I have read and understood the terms and conditions on page 8-9 of the [HREC \(Medical\) application form](#). I confirm that it is my responsibility to ensure that I have received final HREC (Medical) Clearance before commencing any research.

Nabeel Seedat

\_\_\_\_\_  
Name, Surname and Signature

Student/Staff no if applicable: 719484

Date: 28/02/2018

Prof. Vered Aharonson

\_\_\_\_\_  
Name, Surname and Signature

Supervisor (if applicable)

Date: 28/02/2018

## Appendix B

### Additional Results

This appendix presents additional results including the full list of ANOVA significant features and features similar between the significant CI features and the ANOVA features.

TABLE B.1: ANOVA Significant Features - 3m Test

Mutual Information x Accel - radius	Mutual Information y Accel - z acc	x Accel feature Mean tkeo
Mean Step Velocity	y Accel IQR	z Accel Q3
Mean Encoder Difference	Kurtosis y	y 99 power turn
Mean Step length	z 3db turn	Kurtosis x
Mutual Information y Accel - Radius	y half power	Max Force Difference
Cross Entropy z Accel - Radius	max x Accel - y Accel Correlation	z half power
Mutual Information z Accel - Radius	max zy Correlation	Mutual Information y Accel - Azimuth
Mutual Information z Accel - RMS	y Accel Q1	z 99 Power Turn
Mutual Information x Accel - y Accel	Cross Entropy y Accel - Radius	x 3dB Bandwidth Turn
RMS Accel Entropy	y feature Mean tkeo	z Accel Q1
x Accel Entropy	y half power turn	Kurtosis azimuth
y Accel Entropy	Cross Entropy y Accel - Radius	Radius Mode
z Accel Entropy	Total Walk Time	RMS Accel Mode
Azimuth Entropy	y Accel Q3	Mutual Information z Accel - Azimuth
Elevation Entropy	Mutual Information x Accel - Elevation	Mutual Information x Accel - Azimuth
Radius Entropy	Standard Deviation turn Velocity	x Accel Median
Mean Turn Velocity	z Accel IQR	Mutual Information y Accel - Elevation
Mutual Information x Accel - z Accel	max x Accel - z Accel Correlation	Standard Deviation Accel
Number of Steps	RMS Accel Q1	Correlation x Accel - Radius
Walk Ratio	radius Q1	Azimuth Range
y 99 power	x Mean Frequency	Mean Force Difference
y Accel MSE	x Mean Frequency	
y Accel Variance	z Accel 3dB Bandwidth	



TABLE B.2: ANOVA Significant Features - 10m Test

Mean Turn Velocity	Mean Encoder Difference	Cross Entropy y Accel - Radius
Mean Step Length	Cross Entropy z Accel - Radius	x Accel IQR
Walk Ratio	Mutual Information z Accel- Elevation	y 99 Bandwidth
Number of Steps	Turn Time	z Accel 3db Bandwidth Turn
Total Tug Time	Mutual Information x Accel- z Accel	RMS Accel Mean
Total Walk Time	Mutual Information x Accel - Elevation	Radius Mean
Mutual Information y Accel - Elevation	Mutual Information z Accel - Azimuth	y half power
Mean Step Velocity	Mutual Information y Accel - Azimuth	Correlation z Accel - Azimuth
RMS Accel Entropy	Standard Deviation turn Velocity	Total Harmonic Distortion y
x Accel Entropy	z Mean turn Frequency	Mean step time
y Accel Entropy	Mutual Information x Accel - Azimuth	Correlation y Accel - Radius
z Accel Entropy	z Accel IQR	cadence
Azimuth Entropy	y Accel Q3	RMS Accel qe
Elevation Entropy	x Mean Turn Frequency	Radius Q3
Radius Entropy	Correlation y Accel - Elevation	y Accel Variance
Cross Entropy y Accel - Radius	Correlation y Accel - z Accel	x Accel Q3
Mutual Information y Accel - Radius	radius Q1	feature Mean tkeo
Mutual Information y Accel - z Accel	RMS Accel Q1	y 99 power
Mutual Information z Accel - Radius	RMS Accel Median	y Accel MSE
Mutual Information z Accel - RMS	radius Median	Standard Deviation Step Velocity
Mutual Information x Accel - Radius	y Accel Q1	z Mean Frequency
Mutual Information x Accel - y Accel	Mean Accel	z Mean Frequency turn

TABLE B.3: Features Similar between the significant CI features and the ANOVA features

Mean Step Velocity	rms accel Q1
Mean Encoder Difference	y accel Q1
Standard Deviation Turn Velocity	radius Q1
Mean Step Length	y accel IQR
Correlation x-z	z accel IQR
Mean Turn Velocity	azimuth range
y half power	y accel MSE
z accel 3db bandwidth turn	y accel TKEO
y accel Q3	Correlation x accel-radius
z accel Q3	Mutual Information z accel-radius
RMS Accel Mode	Mutual Information z accel-rms
Radius Mode	Cross Entropy y accel-radius
x accel median	Standard Deviation Step Length
kurtosis azimuth	Standard deviation acceleration
x accel mean freq	Number of steps
z accel mean freq	z Accel Median
mean f diff	z Accel 3dB Bandwidth
y accel variance	elevation median
walk ratio	skewness elevation
total harmonic distortion y	z accel mean
x accel mean freq	step regularity
z accel mean freq	harmonic ratio
y 99 power	y accel ZCR
y half power turn	Cross Entropy z accel-elevation
y 99 power turn	elevation mean

TABLE B.4: Features Similar between the significant CI features and the random forest selected features

Mean encoder diff	y accel MSE
Number of steps	Mutual Information z accel-radius
Std turn Velocity	Mutual Information z accel- RMS
Mean turn Velocity	Mean Step Velocity
Walk ratio	Total Tug time
y half power	RMS accel entropy
y 99 power	RMS accel median
y accel Q1	x accel IQR
elevation mean	x accel MSE
mean step length	azimuth MSE
std accel	azimuth entropy
y accel var	Correlation z accel- azimuth
Total Harmonic Distortion y	Mutual Information x accel - y accel
radius Q1	Mutual Information y acce- radius
y accel Q3	skewness azimuth

## References

- [1] D. Sutherland, “What is parkinson’s disease?” Online. [Online]. Available: <http://www.parkinsonsneurochallenge.org/sarasota-parkinsons-disease-resources/sarasota-what-is-parkinsons-disease.html>
- [2] “Gene therapy for parkinson disease boasts remarkable results,” Mar 2011. [Online]. Available: <https://www.zmescience.com/medicine/genetic/gene-therapy-for-parkinson-disease-boasts-remarkable-results-313412/>
- [3] A. Dzhagaryan, A. Milenkovic, E. Jovanov, and M. Milosevic, “Smart button: A wearable system for assessing mobility in elderly,” in *2015 17th International Conference on E-health Networking, Application Services (HealthCom)*, Oct 2015, pp. 416–421.
- [4] V. Aharonson, I. Schlesinger, A. M. McDonald, S. Dubowsky, and A. D. Korczyn, “A practical measurement of parkinson’s patients gait using simple walker-based motion sensing and data analysis,” *Journal of Medical Devices*, vol. 12, p. 011012, Feb. 2018.
- [5] F. Azevedo, “Complete system for quadcopter control,” Ph.D. dissertation, Federal University of Rio Grande, 2014.
- [6] J. O. Pedro and A. J. Crouse, “Direct adaptive neural control of a quadrotor unmanned aerial vehicle,” in *2015 10th Asian Control Conference (ASCC)*, May 2015, pp. 1–6.
- [7] C.-Y. Hsu, Y.-S. Tsai, C.-S. Yau, H.-H. Shie, and C.-M. Wu, “Test-retest reliability of an automated infrared-assisted trunk accelerometer-based gait analysis system,” *Sensors*, vol. 16, no. 8, p. 1156, 2016.

- 
- [8] M. J. Katz, R. B. Lipton, C. B. Hall, M. E. Zimmerman, A. E. Sanders, J. Verghese, D. W. Dickson, and C. A. Derby, “Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and alzheimer dementia in blacks and whites,” *Alzheimer Disease & Associated Disorders*, vol. 26, pp. 335–343, 2012.
- [9] A. W. Willis, B. A. Evanoff, M. Lian, S. R. Criswell, and B. A. Racette, “Geographic and ethnic variation in parkinson disease: A population-based study of us medicare beneficiaries,” *Neuroepidemiology*, vol. 34, pp. 143–151, 2010.
- [10] L. M. L. de Lau and M. M. B. Breteler, “Epidemiology of parkinson’s disease,” *The Lancet Neurology*, vol. 5, pp. 525–535, Jun. 2006.
- [11] A. Salarian, H. Russmann, F. J. G. Vingerhoets, C. Dehollain, Y. Blanc, P. R. Burkhard, and K. Aminian, “Gait assessment in parkinson’s disease: toward an ambulatory system for long-term monitoring,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 8, pp. 1434–1443, Aug 2004.
- [12] E. Naqvi, “Parkinson’s disease statistics.” [Online]. Available: <https://parkinsonsnewstoday.com/parkinsons-disease-statistics/>
- [13] J. Cancela, M. Pastorino, M. Arredondo, M. Pansera, L. Pastor-Sanz, F. Villagra, M. Pastor, and A. Gonzalez-Marcos, “Gait assessment in parkinson’s disease patients through a network of wearable accelerometers in unsupervised environments,” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2011, pp. 2233–6, 08 2011.
- [14] M. M. Hoehn and M. D. Yahr, “Parkinsonism: onset, progression, and mortality. 1967.” *Neurology*, vol. 50, no. 2, p. 318 and 16 pages following, feb 1998. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9484345>
- [15] M. Wang, “Predictive ability of cerebrospinal fluid biomarkers in diagnosing and evaluating parkinson’s disease,” Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

- [16] T. Sayeed, “Methods and models in signal processing for gait analysis using waist-worn accelerometer: A contribution to parkinson’s disease,” Ph.D. dissertation, Universitat Politècnica de Catalunya, 2015.
- [17] S. V. Perumal and R. Sankar, “Gait and tremor assessment for patients with parkinson’s disease using wearable sensors,” *ICT Express*, vol. 2, no. 4, pp. 168 – 174, 2016, special Issue on Emerging Technologies for Medical Diagnostics. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405959516301382>
- [18] P.-H. Chen, R.-L. Wang, D.-J. Liou, and J.-S. Shaw, “Gait Disorders in Parkinson’s Disease: Assessment and Management,” *International Journal of Gerontology*, vol. 7, no. 4, pp. 189–193, dec 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1873959813000410>
- [19] E. Zahnd, “Prediction methods for 1=f processes with application to the analysis of stride interval time series,” Ph.D. dissertation, University of Pittsburgh, 2015.
- [20] B. Mariani, M. Castro Jimenez, F. J G Vingerhoets, and K. Aminian, “On-shoe wearable sensors for gait and turning assessment of patients with parkinson’s disease,” *IEEE Transactions on Bio-Medical Engineering*, vol. 60, pp. 155–8, 01 2013.
- [21] R. Felder and M. Alwan, *Eldercare technology for clinical practitioners*, 1st ed. Humana Press, 2008, vol. 1.
- [22] M. S. Bryant, D. H. Rintala, J. E. Graham, J. G. Hou, and E. J. Protas, “Determinants of use of a walking device in persons with Parkinson’s disease.” *Archives of physical medicine and rehabilitation*, vol. 95, no. 10, pp. 1940–5, oct 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24953250><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4182107>

- 
- [23] B. Mariani, “Assessment of foot signature using wearable sensors for clinical gait analysis and real-time activity recognition,” Ph.D. dissertation, Ecole polytechnique federale de Lausanne, 2012.
- [24] J. Machado, “Smartphone based closed-loop auditory cueing system,” Ph.D. dissertation, University of Porto, 2014.
- [25] K. Watanabe and M. Hokari, “Kinematical analysis and measurement of sports form,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 36, no. 3, pp. 549–557, May 2006.
- [26] R. Barcik, “Assessment of parkinson gait through digital signal processing and machine learning,” Ph.D. dissertation, Dalarna University, 2017.
- [27] M. P Murray, S. B Sepic, G. M Gardner, and W. J Downs, “Walking patterns of men with parkinsonism,” *American journal of physical medicine*, vol. 57, pp. 278–94, 01 1979.
- [28] S. Morris, “A shoe integrated sensor system for wireless gait analysis and real-time therapeutic feedback,” Ph.D. dissertation, Massachusetts Institute of Technology, 2004.
- [29] J. D. Schaafsma, N. Giladi, Y. Balash, A. L. Bartels, T. Gurevich, and J. M. Hausdorff, “Gait dynamics in parkinson’s disease: relationship to parkinsonian features, falls and response to levodopa,” *Journal of the Neurological Sciences*, vol. 212, no. 1-2, pp. 47–53, 2003.
- [30] J. Perlmutter, “Assessment of parkinson disease manifestations,” *Current protocols in neuroscience*, vol. Chapter 10, p. Unit10.1, 10 2009.
- [31] M. Milosevic, E. Jovanov, and A. Milenković, “Quantifying timed-up-and-go test: A smartphone implementation,” in *2013 IEEE International Conference on Body Sensor Networks*, May 2013, pp. 1–6.
- [32] D. Podsiadlo and S. Richardson, “The timed “up & go”: A test of basic functional mobility for frail elderly persons,” *Journal of the American Geriatrics Society*, vol. 39, pp. 142–8, 02 1991.

- [33] C. Godinho, J. Domingos, G. Cunha, A. T. Santos, R. M. Fernandes, D. Abreu, N. Gonçalves, H. Matthews, T. Isaacs, J. Duffen, A. Al-Jawad, F. Larsen, A. Serrano, P. Weber, A. Thoms, S. Sollinger, H. Graessner, W. Maetzler, and J. J. Ferreira, “A systematic review of the characteristics and validity of monitoring technologies to assess parkinson’s disease,” *Journal of NeuroEngineering and Rehabilitation*, vol. 13, no. 1, p. 24, Mar 2016. [Online]. Available: <https://doi.org/10.1186/s12984-016-0136-7>
- [34] S. Patel, R. Hughes, N. Huggins, D. Standaert, J. Growdon, J. Dy, and P. Bonato, “Using wearable sensors to predict the severity of symptoms and motor complications in late stage parkinson’s disease,” in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2008, pp. 3686–3689.
- [35] A. Handojoseno, “Detection of freezing of gait in patients with parkinson’s disease using electroencephalography and computational intelligence,” Ph.D. dissertation, University of Technology Sydney, 2015.
- [36] C. Stamate, G. D. Magoulas, S. Kueppers, E. Nomikou, I. Daskalopoulos, M. U. Luchini, T. Moussouri, and G. Roussos, “Deep learning parkinson’s from smartphone data,” in *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2017, pp. 31–40.
- [37] M. A. Brodie, N. H. Lovell, C. G. Canning, H. B. Menz, K. Delbaere, S. J. Redmond, M. Latt, D. L. Sturnieks, J. Menant, S. T. Smith, and S. R. Lord, “Gait as a biomarker? accelerometers reveal that reduced movement quality while walking is associated with parkinson’s disease, ageing and fall risk,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2014, pp. 5968–5971.
- [38] A. Akbari, R. B. Dewey, and R. Jafari, “Validation of a new model-free signal processing method for gait feature extraction using inertial measurement units to diagnose and quantify the severity of parkinson’s disease,” in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, July 2017, pp. 1–5.



- [39] J. Ballesteros, C. Urdiales, A. B. Martinez, and M. Tirado, "Automatic assessment of a rollator-users condition during rehabilitation using the i-walker platform," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2009–2017, Nov 2017.
- [40] L. Li, L. Atallah, B. Lo, and G. Z. Yang, "Feature extraction from ear-worn sensor data for gait analysis," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, June 2014, pp. 560–563.
- [41] S. Iram, "Early detection of neurodegenerative diseases from bio-signals: A machine learning approach," Ph.D. dissertation, Liverpool John Moores University, 2014.
- [42] A. Ferrari, P. Ginis, M. Hardegger, F. Casamassima, L. Rocchi, and L. Chiari, "A mobile kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 7, pp. 764–773, July 2016.
- [43] M. Qi, "A comprehensive performance comparison of signal processing features in detecting alcohol consumption from gait data," Ph.D. dissertation, Worcester Polytechnic Institute, 2016.
- [44] L. Palmerini, "Data mining in clinical practice for the quantification of motor impairment in parkinson's disease," Ph.D. dissertation, University of Bologna, 2012.
- [45] A. Zhan, M. A. Little, D. A. Harris, S. O. Abiola, E. Ray Dorsey, S. Saria, and A. Terzis, "High frequency remote monitoring of parkinson's disease via smartphone: Platform overview and medication response detection," *ArXiv*, 01 2016.
- [46] A. Zhan, S. Mohan, C. Tarolli, R. B. Schneider, J. L. Adams, S. Sharma, M. J. Elson, K. L. Spear, A. Glidden, M. A. Little, A. Terzis, E. Ray Dorsey, and S. Saria, "Using smartphones and machine learning to quantify parkinson disease severity: The mobile parkinson disease score," *Jama Neurology*, 03 2018.

- [47] S. Patel, D. Sherrill, R. Hughes, T. Hester, N. Huggins, T. Lie-Nemeth, D. Standaert, and P. Bonato, “Analysis of the severity of dyskinesia in patients with parkinson’s disease via wearable sensors,” in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN’06)*, April 2006, p. 126.
- [48] N. Tahir and H. Manap, “Parkinson disease gait classification based on machine learning approach,” *Journal of Applied Sciences*, vol. 12, pp. 180–185, 02 2012.
- [49] Q. W. Oung, M. Hariharan, S. N. Basah, S. Yaacob, M. Sarillee, and H. L. Lee, “Use of technological tools for parkinson’s disease early detection: A review,” in *2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014)*, Nov 2014, pp. 343–348.
- [50] S. Bind, A. Tiwari, and A. Sahani, “A survey of machine learning based approaches for parkinson disease prediction,” *International Journal of Computer Science and Information Technologies*, vol. 6, no. 2, pp. 1648–1655, 2015.
- [51] H. H. Manap, N. M. Tahir, and A. I. M. Yassin, “Statistical analysis of parkinson disease gait classification using artificial neural network,” in *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec 2011, pp. 060–065.
- [52] A. Khemphila and V. Boonjing, “Parkinsons disease classification using neural network and feature selection,” in *International Journal of Mathematical and Computational Sciences*, 2012.
- [53] L. F. Araghi, M. H. Korayem, A. Nikoobin, and F. Setoudeh, “Neural network controller based on pid controller for two links- robotic manipulator control.”
- [54] I. Thangaraju, M. Muruganandam, and C. Nagarajan, “Implementation of pid trained artificial neural network controller for different dc motor drive,” in *Middle-East Journal of Scientific Research*, 2015.

- [55] T. Callinan, “Artificial neural network identification and control of the inverted pendulum,” *MEng Project Reports, School of Electronic Engineering Dublin City University*, 2003.
- [56] S. S. Haykin, *Neural networks and learning machines*, 3rd ed. Pearson Education, 2009.
- [57] M. Chamanirad, “Design and implementation of controller for robotic manipulators using artificial neural networks,” 04 2009.
- [58] Y. SONG, S. WU, and Y. YAN, “Development of self-tuning intelligent pid controller based on bpnn for indoor air quality control,” in *International Journal of Emerging Technology and Advanced Engineering*, 2013.
- [59] C. Dung Doan and S.-Y. Liong, “Generalization for multilayer neural network bayesian regularization or early stopping,” *2nd Conference Asia Pacific Association of Hydrology and Water Resources*, 01 2004.
- [60] M. Kayri, “Predictive abilities of bayesian regularization and levenberg–marquardt algorithms in artificial neural networks: A comparative empirical study on social data,” *Mathematical and Computational Applications*, vol. 21, pp. 1–11, 05 2016.
- [61] A. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, 01 1998.
- [62] G. Yadav, Y. Kumar, and G. Sahoo, “Predication of parkinson’s disease using data mining methods: A comparative analysis of tree, statistical and support vector machine classifiers,” in *2012 National Conference on Computing and Communication Systems*, Nov 2012, pp. 1–8.
- [63] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.
- [64] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

- [65] A. Cutler, D. R. Cutler, and J. R. Stevens, “Random forests,” in *Ensemble machine learning*. Springer, 2012, pp. 157–175.
- [66] B. Pogorelc, Z. Bosnić, and M. Gams, “Automatic recognition of gait-related health problems in the elderly using machine learning,” *Multimedia Tools and Applications*, vol. 58, no. 2, pp. 333–354, May 2012. [Online]. Available: <https://doi.org/10.1007/s11042-011-0786-1>
- [67] L. Huan, “Boosting and additive models,” Lecture, 2011.
- [68] C. Li, “A gentle introduction to gradient boosting,” Lecture, 2016.
- [69] L. Patryla and D. Galeriua, “Statistical performances measures—models comparison,” *CEA: Paris, France*, 2011.
- [70] J. Chang and S. Hanna, “Air quality model performance evaluation,” *Meteorology and Atmospheric Physics*, vol. 87, no. 1-3, pp. 167–196, 2004.
- [71] A. Zheng, *Evaluating Machine Learning Models*. Sebastopol, CA: O’Reilly, 2015.
- [72] T. Baguley, “Calculating and graphing within-subject confidence intervals for anova,” *Behavior Research Methods*, vol. 44, no. 1, pp. 158–175, Mar 2012. [Online]. Available: <https://doi.org/10.3758/s13428-011-0123-7>
- [73] G. Cumming and S. Finch, “Inference by eye: Confidence intervals and how to read pictures of data,” *American Psychologist*, pp. 170–180, 2005.
- [74] S. Smith, *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Pub., 1997. [Online]. Available: <https://books.google.com/books?id=rp2VQgAACAAJ>
- [75] J.-C. Nunes and E. Delchelle, “Empirical mode decomposition: Applications on signal and image processing,” *Advances in Adaptive Data Analysis*, vol. 1, no. 01, pp. 125–175, 2009.
- [76] N. Seedat, D. Beder, V. Aharonson, and S. Dubowsky, “A comparison of foot-fall detection algorithms from walker mounted sensors data,” in *2018 Electric*

- Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, April 2018, pp. 1–4.
- [77] P. Trnka and M. Hofreiter, “The empirical mode decomposition in real-time,” *Proceedings of the 18th International Conference on Process Control*, 01 2011.
- [78] D. Jarchi, J. Pope, T. K. M. Lee, L. Tamjidi, A. Mirzaei, and S. Sanei, “A review on accelerometry-based gait analysis and emerging clinical applications,” *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 177–194, 2018.
- [79] K. D. Toennies, *Guide to Medical Image Analysis: Methods and Algorithms*, 2nd ed. Springer Publishing Company, Incorporated, 2017.
- [80] E. T. Jaynes, *Probability Theory: The Logic Of Science*. Mathematical Intelligencer, 01 2003, vol. 27.
- [81] E. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, 01 2009, vol. 19.
- [82] J. Perry and J. Burnfield, “Gait analysis, normal and pathological function,” *Journal of Pediatric Orthopaedics*, vol. 2, p. 576, 1992.
- [83] M. Galli, V. Cimolin, M. F. De Pandis, D. Le Pera, I. Sova, G. Albertini, F. Stocchi, and M. Franceschini, “Robot-assisted gait training versus treadmill training in patients with parkinson’s disease: a kinematic evaluation with gait profile score,” *Functional Neurology*, vol. 31, pp. 163–70, 07 2016.
- [84] J. C. Schlachetzki, J. Barth, F. Marxreiter, J. Gossler, Z. Kohl, S. Reinfelder, H. Gassner, K. Aminian, B. M. Eskofier, and J. Winkler, “Wearable sensors objectively measure gait parameters in parkinson’s disease,” *PloS one*, vol. 12, no. 10, p. e0183989, 2017.
- [85] C. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. E. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, M. D. Yahr, and L. Seidl, “Movement disorder society task force report on the hoehn and yahr staging scale: status and recommendations.” *Movement disorders : official journal of the Movement Disorder Society*, vol. 19 9, pp. 1020–8, 2004.

- [86] M. Nilashi, O. Ibrahim, and A. Ahani, “Accuracy improvement for predicting parkinson’s disease progression,” in *Scientific reports*, 2016.
- [87] O. Eskidere, F. Ertas, and C. Hanilci, “A comparison of regression methods for remote tracking of parkinson’s disease progression,” *Expert Syst. Appl.*, vol. 39, pp. 5523–5528, 2012.
- [88] M. Kim, J. Kim, S.-H. Lee, and H. Park, “Imaging genetics approach to parkinson’s disease and its correlation with clinical score,” in *Scientific reports*, 2017.
- [89] O. Morozova, O. Levina, A. Uusküla, and R. Heimer, “Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in russia,” *BMC Medical Research Methodology*, vol. 15, no. 1, p. 71, Aug 2015. [Online]. Available: <https://doi.org/10.1186/s12874-015-0066-2>
- [90] C. Lescano, S. Rodrigo, and D. Christian, “A possible parameter for gait clinimetric evaluation in parkinson’s disease patients,” in *Journal of Physics: Conference Series*, vol. 705, no. 1. IOP Publishing, 2016, p. 012019.
- [91] “Heavy tailed distribution & light tailed distribution: Definition & examples.” [Online]. Available: <http://statisticshowto.com/heavy-tailed-distribution/>
- [92] K. Lebel, H. Nguyen, C. Duval, R. Plamondon, and P. Boissy, “Capturing the cranio-caudal signature of a turn with inertial measurement systems: Methods, parameters robustness and reliability,” *Frontiers in Bioengineering and Biotechnology*, vol. 5, p. 51, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fbioe.2017.00051>
- [93] K. Lebel, C. Duval, H. P. Nguyen, R. Plamondon, and P. Boissy, “Cranio-caudal kinematic turn signature assessed with inertial systems as a marker of mobility deficits in parkinson’s disease,” *Frontiers in Neurology*, vol. 9, p. 22, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fneur.2018.00022>

- [94] Q. Keenan, *Handbook of Probability Distributions*, 2nd ed. 48 West 48 Street, Suite 1116, New York, NY 10036, United States: Orange Apple, 2016.