

https://dspace.org/

Gait classification in parkinson's disease using signals from walker mounted sensors

Beder, David

2019

https://hdl.handle.net/10539/28376

Beder, David Alon (2019) Gait classification in Parkinson's disease using signals from walker mounted sensors, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/28376>

Downloaded from a DSpace repository. https://dspace.org/



GAIT CLASSIFICATION IN PARKINSON'S DISEASE USING SIGNALS FROM WALKER MOUNTED SENSORS

Author: David Beder (811071)

Supervisor: Prof. Vered Aharonson

A dissertation submitted to the Faculty of Engineering and the Built Environment, University of Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering.

Johannesburg, 2019

Declaration

I, **David Alon Beder**, declare that this dissertation entitled, "**Gait classification in Parkinson's Disease using signals from walker mounted sensors**" is my own work. 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.



.....

(Signature of Candidate)

8 day of April year 2019

Abstract

Currently, there is no cure for Parkinson's Disease (PD), however, treatments and medications exist that can lessen PD symptoms. The assessments of PD patients' response to these treatments are mostly subjective and existing PD status scales are inadequate to measure small changes, especially where short-term treatments are concerned. This research presents a quantitative measurement system that was able discriminate between three subject groups viz. patients who received a set dosage of medication (treatment patients), patients who received no medication (non-treatment patients), and control subjects. The system also was able to discriminate between the treatment patients according to their pre-assigned clinical scores (such as the Hoehn and Yaer scale), based on their responses to the treatments. The study involves the secondary analysis of encoder and accelerometer signals acquired in a trial conducted at the Rambam Medical Centre, in Israel. Signal pre-processing was used to denoise the signals, identify the movement segments and detect walking footfalls through newly proposed algorithms. Feature extraction and feature selection was implemented to determine feature subsets that could enhance the classifier's performance. A Bagged Decision Trees was applied to differentiate between the subjects according to the subject groups as well as discriminate the treatment patients according to their clinical scores. Results indicated that the three subject groups could be discriminated with an overall accuracy of 92.16% ± 3.92%, a sensitivity of 82.35%, and a specificity of 93.38%. Additionally, key features that were able to classify the three subject groups were identified and have an accuracy of 90.20% ± 4.26%, a sensitivity of 81.37%, and a specificity of 91.64%. However, further examination of these results suggests that this is owing to the systems' ability to differentiate between patients with PD versus controls as opposed to its ability to differentiate between treatment patients and non-treatment patients. Further testing of *only* the treatment patients indicated that the system was able to discriminate treatment patients according to clinical scores with an overall accuracy of 79.41% ± 4.12%, a sensitivity of 85.00%, and a specificity of 73.43%, based on their response to treatment. Across all experiments, the control subjects were correctly classified with an accuracy of at least 90%. Conversely, 50% of non-treatment patients were misclassified as treatment patients. This may indicate that the system is too sensitive trying to discern subtle changes between the treatment patients and non-treatment patients' response to medication that may not be present within the short test frame. Moreover, it may suggest that treatment patients with different clinical scores should receive different dosages of medication to assess the full effects of the treatment.

Dedication

I dedicate this study wholeheartedly to my beloved parents Merle and Barry Beder, who have been a constant source of inspiration and gave me the strength when I thought of giving up, who continually provide their moral, spiritual, emotional, and financial support.

To my brothers, sister, girlfriend, grandparents, relatives, mentor, friends, and colleagues who shared their words of advice and encouragement to finish this study. I would like to express my deepest thanks and gratitude.

Lastly, I would like to acknowledge the works of God, thank you for the guidance, strength, power of mind, protection, and skills and for giving me a healthy life in which to complete this dissertation.

Acknowledgements

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

- To Prof. Vered Aharonson, thank you for all your guidance and mentorship throughout this research task. Your support, advice, and reviews of this research were priceless and helped me learn and grow throughout this research task.

- To all the patients, controls and staff at the Rambam Medical Centre in Israel who were involved in acquiring the data for this research task, thank you.

- To my research group partner, Mr. Nabeel Seedat with whom I collaborated with on the data pre-processing aspect of this research task, thank you.

- To each and every person who has taken the time to read this work and offer feedback, thank you.

Table of Contents

DECLARATION ..... **II**

ABSTRACT ..... **III**

DEDICATION ..... **IV**

ACKNOWLEDGEMENTS ..... **V**

TABLE OF CONTENTS ..... **VI**

LIST OF FIGURES ..... **X**

LIST OF TABLES ..... **XIII**

ABBREVIATIONS ..... **XV**

PUBLICATIONS ..... **XVI**

CHAPTER 1. INTRODUCTION **1**

1.1 RESEARCH CONTEXT 1

1.2 RESEARCH QUESTION 3

CHAPTER 2. LITERATURE REVIEW **4**

2.1 PARKINSON'S DISEASE 4

2.2 CLINICAL ASSESSMENT AND DIAGNOSIS OF PD 4

2.2.1 Unified Parkinson Disease Rating Scale *5*

2.2.2 Hoehn and Yahr Scale *5*

2.2.3 Timed-Up-and-Go Protocol *6*

2.3 PARKINSON'S DISEASE AND GAIT RELATIONSHIP 7

2.4 ANALYSIS OF GAIT 8

2.4.1 Gait *8*

2.4.2 Gait Abnormalities *8*

2.5 GAIT ANALYSIS SYSTEMS 9

2.5.1 Optical Motion Capture Systems *9*

2.5.2 Instrumented Walkway Systems *9*

2.5.3 Wearable Systems *10*

2.5.4 Instrumented Walker Systems *10*

2.6 DISCRIMINATION SYSTEMS BASED ON KINEMATIC SIGNALS 11

2.6.1 Signal Pre-processing *12*

2.6.1.1 Signal Filtering *13*

2.6.1.2 3-Axis Signal Rotation *14*

2.6.1.3 Footfalls Detection 14
2.6.2 Feature Extraction 15
2.6.3 Feature Selection 16
2.6.3.1 Projection-Based Feature Selection 16
2.6.3.1.1 Principal Component Analysis 17
2.6.3.2 Criteria-Based Feature Selection 17
2.6.3.2.1 Wrapper Feature Selection Method 18
2.6.3.2.2 Filter Feature Selection Method 18
2.6.4 Discrimination Systems 18
2.6.4.1 Unsupervised Learning Algorithm 19
2.6.4.1.1 K-Means Clustering 19
2.6.4.2 Supervised Learning Algorithm 19
2.6.4.2.1 Support Vector Machines 20
2.6.4.2.2 Artificial Neural Networks 20
2.6.4.2.3 Naïve Bayes 20
2.6.4.2.4 Decision Trees 21
2.6.4.2.5 Bagged Decision Trees (Bagging) 22

CHAPTER 3. METHODOLOGY 25
3.1 HIGH LEVEL OVERVIEW 25
3.2 DATABASE 25
3.2.1 Subject Description 25
3.2.2 Experimental Device 26
3.2.3 Experimental Procedure 27
3.2.4 Data and Signals Acquired 27
3.3 DATA AND SIGNAL COMPLETENESS AND INTEGRITY CHECK 28
3.4 SIGNAL PRE-PROCESSING 28
3.4.1 Movement Identification Algorithm 29
3.4.1.1 Encoder Signal Preparation 29
3.4.1.2 Signal Filtering 30
3.4.1.3 Movement Stage Selection 30
3.4.2 Footfall Detection Algorithm 31
3.4.2.1 Accelerometer Signal Preparation 31
3.4.2.2 Filtering and Footfall Detection 32
3.5 FEATURE EXTRACTION 33
3.5.1 Pre-and-Post Feature Extraction 34
3.5.2 Change between Pre-and-Post Extracted Features 34

3.6 FEATURE SELECTION 35
3.7 DISCRIMINATION SYSTEM 36
3.7.1 Dataset Size Considerations 36
3.7.2 Classification Algorithm 37
3.7.3 Software Implementation 37
3.7.4 Performance Metrics 37
3.7.5 Classification Experiments 38
3.8 KEY FEATURE IDENTIFICATION AND VALIDATION 39
3.8.1 Key Feature Identification 39
3.8.2 Key Feature Validation 39
CHAPTER 4. RESULTS 40
4.1 DATABASE 40
4.2 DATA/SIGNAL COMPLETENESS AND INTEGRITY CHECK 40
4.3 SIGNAL PRE-PROCESSING 41
4.3.1 Movement Identification Algorithm 41
4.3.1.1 Signal Filtering 41
4.3.1.2 Movement Stage Selection Output 42
4.3.2 Footfall Detection Algorithm 43
4.3.2.1 Signal Filtering 43
4.3.2.2 Footfall Detection Output and Accuracy Evaluation 44
4.4 FEATURE EXTRACTION 47
4.5 FEATURE SELECTION 50
4.5.1 PCA Selected Features 50
4.5.2 1-way ANOVA Selected Features 50
4.5.3 Medically Observable Features 51
4.5.4 FSS Features 52
4.6 DISCRIMINATION SYSTEM 52
4.6.1 Experiment 1 – Finding the Optimal Number of Trees 54
4.6.2 Experiment 2 – Classification of the 3 Subject Groups 54
4.6.3 Experiment 3 – Classification between Treatment Patients and Non-treatment Patients 57
4.6.4 Experiment 4 – Classification between Treatment Patients H&Y 2 and H&Y 3 Clinical Scores 59
4.7 KEY FEATURE IDENTIFICATION 61
4.8 KEY FEATURE VALIDATION 61
CHAPTER 5. DISCUSSION AND CONCLUSION 65

5.1 KEY FINDINGS 65
5.1.1 Research Goals 65
5.1.1.1 Experiment 2 – Classification of the 3 Subject Groups 65
5.1.1.2 Experiment 3 – Classification between Treatment Patients and Non-treatment Patients 68
5.1.1.3 Experiment 4 – Classification between Treatment Patients H&Y 2 and H&Y 3 69
5.1.1.4 Feature Selection Methods 70
5.1.1.5 Key Features Identification and Validation 71
5.1.1.5.1 Interpretation of the Key Features 72
5.1.2 Differences between the 3m and 10m TUG Data 72
5.1.3 Conclusions regarding the use of the PAMMII Instrumented Walker 73
5.1.3.1 Accurate Movement Stage Identification 73
5.1.3.2 Accurate Footfalls Detection 74
5.2 KEY FINDINGS IMPACT 75
5.2.1.1 New Methods to Overcome Noise and Limitations of Low-cost Sensors 75
5.2.2 New Footfalls Detection Algorithm 75
5.3 RESEARCH LIMITATIONS AND FUTURE WORK RECOMMENDATIONS 76
5.4 CONCLUSION 76
REFERENCES 79
APPENDIX A. ETHICS CLEARANCE CERTIFICATE 85
APPENDIX B. 10M TUG DATA RESULTS 86

List of Figures

FIGURE 2.1: THE HUMAN GAIT CYCLE, IN TERMS OF STANCE AND SWING PHASE, WHERE ONE COMPLETE GAIT CYCLE IS MEASURED AS THE TIME RECORDED BETWEEN TWO CONSECUTIVE FOOTFALLS OF THE SAME FOOT [24] ..8
FIGURE 2.2: DIAGRAM OF THE PAMMII INSTRUMENTED WALKER, HIGHLIGHTING THE PLACEMENT OF ITS SENSORS AND CONTROL BOX 11
FIGURE 2.3: FLOW DIAGRAM OF A TYPICAL DISCRIMINATION SYSTEM RESEARCH TASK 12
FIGURE 2.4: ACCELEROMETER BASED FOOTFALLS DETECTION FROM [50] WHERE RED STARS ON EACH PEAK OF THE SIGNAL INDICATE FOOTSTEPS.....15
FIGURE 2.5: EXAMPLE OF A DECISION TREE SPLIT 21
FIGURE 2.6: BAGGED DECISION TREE ALGORITHM OVERVIEW, HIGHLIGHTING HOW BOOTSTRAP AGGREGATING IS USED TO RESAMPLE DATA WITH REPLACEMENT 23
FIGURE 3.1: BLOCK DIAGRAM DESCRIBING THE METHODOLOGY IMPLEMENTED TO ANSWER THE RESEARCH QUESTION.....25
FIGURE 3.2: PAMMII SCHEMATIC DIAGRAM, HIGHLIGHTING THE PLACEMENT OF ITS SENSORS AND CONTROL BOX 27
FIGURE 3.3: FLOW DIAGRAM DESCRIBING HOW THE MOVEMENT STAGE IDENTIFICATION ALGORITHM DETERMINES MOVEMENT AND NON-MOVEMENT PERIODS.....30
FIGURE 3.4: EXAMPLE OF HOW EACH IMF LEVEL IS CALCULATED 33
FIGURE 4.1: EXAMPLE OF A FILTERED VS. UNFILTERED DISTANCE ENCODER SIGNAL.....42
FIGURE 4.2: EXAMPLE OF A FILTERED VS. UNFILTERED DERIVED VELOCITY SIGNAL 42
FIGURE 4.3: MOVEMENT STAGE IDENTIFICATION ALGORITHM OUTPUT RESULTS, HIGHLIGHTING THE SEGMENTATION BETWEEN STRAIGHT LINE WALKING AND TURNING PERIODS.....43
FIGURE 4.4: EXAMPLE OF AN EMD FILTERED VS. UNFILTERED Z-AXIS ACCELEROMETER SIGNAL.....44
FIGURE 4.5: EXAMPLE OF STRAIGHT-LINE WALKING FOOTFALLS DETECTION VALIDATION, FOR THE 4 TESTS PERFORMED, FOR IMF LEVELS 1-4.....45
FIGURE 4.6: NUMBER OF CORRECTLY PREDICTED FOOTFALLS, FOR THE 4 TESTS PERFORMED, FOR IMF LEVELS 1-4 46
FIGURE 4.7: FOOTFALL DETECTION ACCURACY, FOR THE 4 TESTS PERFORMED AND AN AVERAGE ACCURACY, FOR IMF LEVELS 1-4.....46
FIGURE 4.8: ACCURACY OF THE BDT CLASSIFIER VS. NUMBER OF TREES, HIGHLIGHTING THE OPTIMAL NUMBER OF TREES NEEDED ACROSS THE IMPLEMENTED FEATURE SELECTION METHODS, FOR THE 3M TUG DATA ...54
FIGURE 4.9: ACCURACY OF THE BDT CLASSIFIER USING THE FIVE FEATURE SELECTION METHODS AS AN INPUT TO CLASSIFY THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA.....55
FIGURE 4.10: CONFUSION MATRIX OF THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY PCA AS AN INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO THEIR REAL SUBJECT GROUPS (TREATMENT PATIENTS, NON-TREATMENT PATIENTS, AND CONTROLS) AND THE PREDICTED CLASSES (ROWS) ARE THE BDT CLASSIFIER OUTPUT FOR EACH SUBJECT INTO THE 3 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 3M TUG DATA 56

FIGURE 4.11: ACCURACY OF THE BDT CLASSIFIER USING THE FIVE FEATURE SELECTION METHODS AS AN INPUT TO CLASSIFY TREATMENT PATIENTS VS. NON-TREATMENT PATIENTS, FOR THE 3M TUG DATA 57
FIGURE 4.12: CONFUSION MATRIX OF THE BAGGED DECISION TREE ALGORITHM USING THE FEATURE-SET SELECTED BY PCA AS AN INPUT FOR CLASSIFICATION OF TREATMENT PATIENTS VS. NON-TREATMENT PATIENTS. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO TREATMENT PATIENTS AND NON-TREATMENT PATIENTS AND THE PREDICTED CLASSES (ROWS) ARE THE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THESE 2 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 3M TUG DATA 58
FIGURE 4.13: ACCURACY OF THE BDT CLASSIFIER USING THE FIVE FEATURE SELECTION METHODS AS AN INPUT TO CLASSIFY TREATMENT PATIENTS IN TERMS OF THEIR H&Y 2 AND H&Y 3 CLINICAL SCORES, FOR THE 3M TUG DATA 59
FIGURE 4.14: CONFUSION MATRIX OF THE BDT ALGORITHM USING THE FEATURE-SET SELECTED BY ANOVA AS AN INPUT FOR CLASSIFICATION OF THE TREATMENT PATIENTS ACCORDING TO THEIR H&Y CLINICAL SCORES. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO THEIR H&Y 2 AND H&Y 3 CLINICAL SCORES AND THE PREDICTED CLASSES (ROWS) ARE THE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THESE 2 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 3M TUG DATA 60
FIGURE 4.15: PERFORMANCE COMPARISON OF THE BDT CLASSIFIER COMPARING THE KEY FEATURES PERFORMANCE AGAINST THE PCA SELECTION METHOD IN TERMS OF CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA.....62
FIGURE 4.16: CONFUSION MATRIX OF THE BDT CLASSIFIER USING THE KEY FEATURES AS AN INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO THEIR REAL SUBJECT GROUPS (TREATMENT PATIENT, NON-TREATMENT PATIENT AND CONTROL) AND THE PREDICTED CLASSES (ROWS) ARE THE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THE 3 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 3M TUG DATA 63
FIGURE B.1: ACCURACY OF THE BAGGED DECISION TREE CLASSIFIER VS. NUMBER OF TREES, HIGHLIGHTING THE OPTIMAL NUMBER OF TREES NEEDED ACROSS THE IMPLEMENTED FEATURE SELECTION METHODS, FOR THE 10M TUG DATA 86
FIGURE B.2: AVERAGE ACCURACY OF THE BDT CLASSIFIER USING THE FIVE FEATURE SELECTION METHODS AS AN INPUT TO CLASSIFY THE 3 SUBJECT GROUPS, FOR THE 10M TUG DATA 87
FIGURE B.3: CONFUSION MATRIX OF THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY PCA AS AN INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO THEIR REAL SUBJECT GROUPS (TREATMENT PATIENTS, NON-TREATMENT PATIENTS AND CONTROL) AND THE PREDICTED CLASSES (ROWS) ARE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THE 3 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 10M TUG DATA 88
FIGURE B.4: ACCURACY OF THE BDT CLASSIFIER USING THE FIVE FEATURE SELECTION METHODS AS AN INPUT TO CLASSIFY TREATMENT PATIENTS VS. NON-TREATMENT PATIENTS, FOR THE 10M TUG DATA 89

FIGURE B.5: CONFUSION MATRIX OF THE BAGGED DECISION TREE ALGORITHM USING THE FEATURE-SET SELECTED BY PCA AS AN INPUT FOR CLASSIFICATION OF TREATMENT PATIENTS VS. NON-TREATMENT PATIENTS. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO TREATMENT PATIENTS AND NON-TREATMENT PATIENTS AND THE PREDICTED CLASSES (ROWS) ARE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THESE 2 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 10M TUG DATA 90
FIGURE B.6: ACCURACY OF THE BDT CLASSIFIER USING THE FIVE FEATURE SELECTION METHODS AS AN INPUT TO CLASSIFY PD PATIENTS H&Y 2 AND H&Y 3 CLINICAL SCORES, FOR THE 10M TUG DATA 91
FIGURE B.7: CONFUSION MATRIX OF THE BDT ALGORITHM USING THE FEATURE-SET SELECTED BY ANOVA AS AN INPUT FOR CLASSIFICATION OF THE TREATMENT PATIENTS ACCORDING TO THEIR H&Y CLINICAL SCORES. THE ACTUAL CLASSES (COLUMNS) ARE THE TREATMENT PATIENTS DIVIDED INTO H&Y 2 AND H&Y 3 CLINICAL SCORES AND THE PREDICTED CLASSES (ROWS) ARE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THESE 2 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 10M TUG DATA 92
FIGURE B.8: PERFORMANCE COMPARISON OF THE BDT CLASSIFIER COMPARING THE KEY FEATURES PERFORMANCE AGAINST THE PCA FEATURE SELECTION METHOD IN TERMS OF CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 10M TUG DATA 94
FIGURE B.9: CONFUSION MATRIX OF THE BDT CLASSIFIER USING THE KEY FEATURES AS AN INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS. THE ACTUAL CLASSES (COLUMNS) ARE THE SUBJECTS DIVIDED INTO THEIR REAL SUBJECT GROUPS (TREATMENT PATIENT, NON-TREATMENT PATIENT AND CONTROL) AND THE PREDICTED CLASSES (ROWS) ARE THE BDT CLASSIFIER OUTPUT OF EACH SUBJECT INTO THE 3 GROUPS. EACH CELL CONTAINS BOTH THE NUMBER AND PERCENTAGE OF SUBJECTS CLASSIFIED BY THE BDT ALGORITHM, FOR THE 10M TUG DATA 95

List of Tables

TABLE 2.1: EXAMPLE TABLE OF DERIVED FEATURES, IN TERMS OF GAIT AND SENSOR FEATURES, FOUND IN LITERATURE 15
TABLE 3.1: LIST OF THE PERFORMED EXPERIMENTS DESCRIBING THEIR PURPOSE, THE DATA USED FOR EACH EXPERIMENT AND THE RESPECTIVE PERFORMANCE EVALUATION METRICS 38
TABLE 4.1: PERCENTAGE OF CORRECTLY LOGGED CLINICAL INFORMATION AND SENSOR DATA 41
TABLE 4.2: SUBJECT DEMOGRAPHICS 41
TABLE 4.3: DISTRIBUTION OF THE TREATMENT PATIENTS WITHIN THE DATASET IN TERMS OF THEIR H&Y CLINICAL SCORES 41
TABLE 4.4: LIST OF THE FEATURES EXTRACTED FROM STRAIGHT LINE WALKING PERIODS OF SUBJECTS' TUG DATA. 47
TABLE 4.5: LIST OF THE FEATURES EXTRACTED FROM TURNING PERIODS OF SUBJECTS' TUG DATA 49
TABLE 4.6: SUMMARY OF THE STATISTICALLY SIGNIFICANT FEATURES SELECTED BY THE 1-WAY ANOVA STATISTICAL TEST FOR THE 3M AND 10M TUG DATA 51
TABLE 4.7: LIST OF THE MEDICALLY OBSERVABLE FEATURES SELECTED BY A HEALTHCARE PROFESSIONAL 52
TABLE 4.8: SUMMARY OF THE PERFORMED EXPERIMENTS DESCRIBING THEIR PURPOSE, THE DATA USED FOR EACH EXPERIMENT AND THE RESULTS OF EACH EXPERIMENT 53
TABLE 4.9: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY PCA AS ITS INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA 56
TABLE 4.10: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY PCA AS ITS INPUT FOR CLASSIFICATION OF TREATMENT PATIENTS VS. NON-TREATMENT PATIENTS, FOR THE 3M TUG DATA 58
TABLE 4.11: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY ANOVA AS ITS INPUT TO CLASSIFY TREATMENT PATIENTS IN TERMS OF THEIR H&Y 2 AND H&Y 3 CLINICAL SCORES, FOR THE 3M TUG DATA 60
TABLE 4.12: LIST OF THE 43 KEY FEATURES THAT HAD A MODERATE-VERY HIGH ABSOLUTE CORRELATION COEFFICIENT BETWEEN THE ORIGINAL FEATURES AND THE FIRST PRINCIPAL COMPONENT 61
TABLE 4.13: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE KEY FEATURES FEATURE-SET AS ITS INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA 63
TABLE 4.14: BDT CLASSIFIER RESULTS, COMPARING THE KEY FEATURES AGAINST THE PCA FEATURE SELECTION METHOD INPUTS FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA 64
TABLE B.1: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY PCA AS ITS INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 10M TUG DATA 88
TABLE B.2: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY PCA AS ITS INPUT FOR CLASSIFICATION OF TREATMENT PATIENTS VS. NON-TREATMENT PATIENTS, FOR THE 10M TUG DATA 90
TABLE B.3: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE FEATURE-SET SELECTED BY ANOVA AS ITS INPUT TO CLASSIFY TREATMENT PATIENTS IN TERMS OF THEIR H&Y 2 AND H&Y 3 CLINICAL SCORES, FOR THE 10M TUG DATA 92

TABLE B.4: CLASSIFICATION RESULTS FOR THE BDT CLASSIFIER USING THE KEY FEATURES FEATURE-SET AS ITS INPUT FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA.....95
TABLE B.5: BDT CLASSIFIER RESULTS, COMPARING THE KEY FEATURES AGAINST THE PCA SELECTION METHOD INPUTS FOR CLASSIFICATION OF THE 3 SUBJECT GROUPS, FOR THE 3M TUG DATA.....95

Abbreviations

ANN	**A**rtificial **N**eural **N**etworks
ANOVA	**AN**alysis **O**f **VA**riance
BDT	**B**agged **D**ecision **T**ree
DT	**D**ecision **T**ree
EMD	**E**mpirical **M**ode **D**ecomposition
FSS Features	Frequency, Sensor and Statistical Features
H&Y	**H**oehn and **Y**ahr
IMF	**I**ntrinsic **M**ode **F**unction
PAMMII	**P**ersonal **A**id for **M**obility and **H**ealth **M**onitoring II
PCA	**P**rincipal **C**omponent **A**nalysis
PD	**P**arkinson's **D**isease
SVM	**S**upport **V**ector **M**achines
TUG	**T**imed **U**p and **G**o
UPDRS	**U**nified **P**arkinson's **D**isease **R**ating **S**core

Publications

Publications directly related to this dissertation:

1. 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 Engineering Meeting (EBBT)*, pp. 1-4, 2018.

Chapter 1. Introduction

1.1 Research Context

As of 2017, the Parkinson's Disease Foundation estimates that 7-10 million people worldwide suffer from Parkinson's Disease (PD) [1]. This disease is manifested in the degeneration of brain neurons, starting with the motor areas of the brain, reducing the patient's motor functionality and affecting their gait and other movements. Consequently, the disease reduces the patient's quality of life and their independence [2, 3]. After wide-ranging research over several decades, PD is still not fully understood, and no cure exists to date (at the time of writing this dissertation) [4]. Both the diagnoses of PD and its treatments rely on the observed and/or measured symptoms of the disease.

Current Parkinson's Disease diagnosis and monitoring necessitate the presence of clinical signs which are observed by a healthcare professional [5]. The Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr (H&Y) scale are commonly used standardised clinical assessment scales that incorporate subjective data acquired from the patient and/or caregiver during infrequent patient visits in a hospital setting to describe the severity of PD [6, 7]. Consequently, patients may be misdiagnosed or undiagnosed resulting in incorrect treatment [8].

Patients with PD exhibit motor impairment symptoms which compromise their independence and flexibility when moving and walking [4]. This affects a patient's gait or cyclical manner of walking [9]. Gait analysis is used to measure and monitor the biomechanical irregularities in PD and other movement disorder patients i.e. to assess the manner in which the subject walks (or runs if possible) [10]. Therefore, a patients' gait assessment is seen as an integral part of the long-term management of the patient's condition.

Technological advances in signal processing, electronics, and neuroscience has facilitated the design of devices that could improve PD management by providing physicians with quantitative tools to assess the symptoms [11, 12]. This has allowed for the design of devices that could help physicians to better and more quantitatively assess the symptoms of neurological diseases such as Parkinson's Disease [11, 12]. However, most of these devices are expensive, require controlled test environments, complex to set up, time-consuming to strap on patients, and uncomfortable to wear [13, 14, 15]. Additionally, patients with more severe PD may require a walker, which hampers the measurements of these devices [16]. These deficiencies limit the applicability of these devices to the majority of PD patients. Treatment efficacy measurements for these patients, if performed at all, revert back to the subjective methods, necessitating a healthcare professional's observations. The treatment cannot then be properly and timeously adjusted to the individual patient's needs.

Amongst the various tools proposed for affordable, simple diagnostics of PD symptoms, a novel method, the Personal Aid for Mobility and Health Monitoring II (PAMMII) instrumented walker, was developed. This device is designed to quantify gait disorders, with a focus on PD gait. The device comprises a set of low-cost sensors mounted on a common aluminium walker frame that collect kinematic signals as the user walks [16, 17]. Previous studies have indicated PAMMII as a time saving, affordable and easy-to-use device, even to severely motor-impaired patients [17]. These studies also showed that with an application of signal processing to PAMMII data, the system provided good sensitivity in discriminating PD patients from controls [16, 17]. Additionally, another on-going study strives to use this device data for discrimination between the four different stages of PD, from mild to severe.

The current research strives to further explore the sensitivity of data acquired by PAMMII sensors and to discriminate subtle changes of PD patients following a short-term treatment. This research makes use of a database acquired in a specific clinical trial that obtained data from PD patients who received an IV amantadine-sulphate treatment (treatment patients), PD patients who received no treatment (non-treatment patients) and control subjects [16, 17]. The current study aims to discriminate between the three subject groups as well as classify the treatment patients according to their pre-assigned H&Y clinical scores. In doing so, this research aims to identify the key features of gait change based on this dataset. Gait feature extraction heavily depends on the reliability of the kinematic signals used, as well as the detection of basic gait elements like walking footfalls. This poses a particular challenge when

the signals are acquired from low-cost, noisy, walker-mounted off-body sensors. The applied methodology will address these issues using customised signal pre-processing. If successful, this research could assist healthcare professional with the diagnosis and constant monitoring of PD patients.

This study forms part of a larger research collaboration between the University of the Witwatersrand, the Massachusetts Institute of Technology (MIT-USA) and Halmstad University (Sweden) regarding patient-centred diagnosis and monitoring technologies for Parkinson's Disease.

1.2 Research Question

The primary goal of this research focuses on answering the following question "To what extent can features extracted from signals acquired by walker-mounted sensors identify changes in PD patient gait due to treatment."

This question is translated into three objectives as follows:

1. To determine to what extent combinations of these features provide a classification of the three subject groups.
2. To determine to what extent these features or their combinations identify the severity of the disease according to H&Y clinical scales.
3. To determine which of these features, derived from kinematic signals from walker mounted sensors best discriminate between PD patients gait changes due to treatment compared to non-treatment patients and controls.

Chapter 2. Literature Review

2.1 Parkinson's Disease

Like other neurodegenerative diseases, PD is a result of deterioration in a person's nervous system. More specifically, it is a consequence of a patient's inability to produce sufficient levels of dopamine in the substantia nigra section of the brain [18, 2]. The substantia nigra is a basal ganglia structure which is responsible for movement control [2, 3]. Therefore, motor control and performance issues occur in Parkinson's patients [2, 3].

Commonly, PD patients exhibit motor impairment symptoms which compromise their independence and flexibility when moving [4]. These symptoms include but are not limited to: tremors, muscle rigidity, impaired balance, postural instability, bradykinesia (slowed voluntary movements), ataxia (reduced co-ordination), and dyskinesia (erratic involuntary movement) [5, 18, 4]. These symptoms may translate into impaired gait. Consequently, a patient's quality of life is reduced and in severe stages, may require a caregiver [4].

PD occurs predominantly in older patients; however, it exhibits a relatively wide age distribution compared to other neurodegenerative pathologies such as Alzheimer's disease. Additionally, symptoms are known to differ from one patient to another [5]. This disease is not yet fully understood and to date there are no documented cures that slow or stop the progression of PD [19]. Medical treatments have been devised to suppress and reduce PD related symptoms. However, these treatments and medications exhibit diverse responses in PD patients and can result in numerous side-effects [19]. Thus, constant supervision and close monitoring of a patient's dosage and response to medication is essential.

2.2 Clinical Assessment and Diagnosis of PD

Parkinson's Disease is diagnosed based on clinical observations of its symptoms where medical practitioners evaluate patient symptoms against known PD characteristic traits [6].

This is achieved using standardised biomarker tests such as the Unified Parkinson Disease Rating Scale (UPDRS) and the Hoehn and Yahr (H&Y) scale tests [6, 7]. These tests are performed by healthcare professionals, giving a score to a set of clinical observed symptom of the patient and then summing up these scores. Additionally, the mobility and gait of a PD patient is assessed by the Timed-Up-and-Go (TUG) protocol [20]. The subsequent subsections detail these PD diagnostic and monitoring scales in greater detail.

2.2.1 Unified Parkinson Disease Rating Scale

UPDRS aims to quantify motor and non-motor symptoms of PD in a patient through a long and comprehensive question assessment where four subscales are given a rating between 0 and 4 with the maximum rating indicating severe PD in the patient and a minimum rating signifying a lack of PD symptoms present in the patient [6, 7]. The four UPDRS subscales include [6, 7]:

1. A non-motor symptoms assessment where the medical examiner evaluates the cognitive function and mood of the patient,
2. A self-assessment of the patient's daily living activities where the medical examiner evaluates the patient's perception of the disease's impact on their day-to-day life,
3. A motor examination where the medical practitioner assesses the motor manifestations of PD in the patient, and
4. A prior motor complication assessment where the medical examiner performs a historical assessment of a patient's prior motor difficulties.

The main disadvantage of the UPDRS scale is that it is a long and comprehensive non-linear test that requires a health care professional, usually a doctor, to collect subjective patient history and utilise their personal experience and knowledge to diagnose and monitor the patient. This can introduce personal bias effects and produce inconsistencies between tests [8]. This coupled with the fact that the UPDRS scale is time-consuming to perform makes this test inadequate to evaluate the short-term treatment efficiencies in this research.

2.2.2 Hoehn and Yahr Scale

The H&Y scale utilises a singular PD evaluation performed by a healthcare professional to determine the severity of the disease [6]. This scale measures features of motor impairment to

suggest the degree of disease progression by rating a patient between 0 and 5 [7]. The following list illustrates the range of severities determined by the H&Y scale [6, 7]:

0. Healthy Patient, no signs of PD.
1. Unilateral disease where the patient displays minor disability symptoms.
2. Bilateral disease where the patient does not experience a problem with their balance.
3. Mild or moderate bilateral disease where the patient is physically independent but experiences slight instability.
4. Severe disability, however, the patient is still able to stand and walk independently
5. Wheelchair-bound where the patient requires full assistance to perform day to day activities.

This scale offers healthcare professionals alternative assistance when evaluating the severity of PD in a patient [7]. The main limitations of this test include: (1) A doctor is required to perform the H&Y test. (2) Patients rated with the same H&Y score may experience completely different symptoms with ranging severities [7]. (3) Patients H&Y scores may change across an array of doctors. However, owing to the fact that this scale scores patient mobility neglecting subjective patient history, it may be useful to evaluate the short-term treatment efficacies in this research.

2.2.3 Timed-Up-and-Go Protocol

The TUG protocol is regarded as the standard for early assessment as a potential indicator for PD [20]. This protocol is used to measure the static and dynamic balance of a subject as well as determine their fall risk and general mobility [20, 21]. This test has been shown to assist clinicians in quantifying the severity of PD in an individual [6] and to evaluate whether specific treatments are improving patient health and well-being [7].

The TUG protocol starts with a subject sitting on a chair. The subject is then required to stand up from the chair walk a set distance, turn around and return to a seated position on the chair. The medical examiner records the time of each part of this protocol as well as the total protocol time. The TUG test, therefore, provides quantitative values i.e. the recorded times, to compare stages of PD and treatment efficacies [20, 21].

The main limitation of this test is that its results can be skewed based on human error when recording TUG times using a stopwatch. It is also a crude measure of gait, as it measures only the accumulated getting up and walking speeds of the patient [20, 21]. Moreover, the TUG test, like the aforementioned scales are performed in medical clinics. However, owing to the fact that this scale only measures patient mobility neglecting subjective patient history, this methodology still may be useful to evaluate short-term treatment efficacies.

The limitations of the abovementioned current PD diagnostic and monitoring tools emphasise the need for an automated and objective tool and/or system that could reliably and objectively monitor a patient's mobility (or gait) and consequently infer their PD diagnosis, PD stage, and their treatment efficacy.

2.3 Parkinson's Disease and Gait Relationship

Gait is defined as a walking pattern [9]. Thus, gait analysis can be used to measure the biomechanical irregularities in movement disorder patients i.e. to assess the manner in which the subject walks (or runs if possible) [10]. A healthy person's gait is repetitive in nature. If gait is observed to vary or becomes asymmetrical it may indicate a physical injury or a neurological disorder such as Parkinson's Disease [10]. Symptoms such as impaired balance, postural instability, and muscle rigidity as well as bradykinesia, ataxia, and dyskinesia are known to affect the gait of PD patient's [5, 18, 4]. Therefore, monitoring a PD patient's gait can assist in evaluating disease severity and treatment efficacies [11, 12].

Numerous sources in the literature described gait abnormalities as a common consequence of PD. Hence, it can be inferred that the more severe a person's Parkinson's Disease the lesser their level of gait motor functions. Previous research supports this understanding showcasing how PD patients can exhibit gait abnormalities such as reduced overall velocity, smaller stride lengths, increased total number of steps (during a test), increased stride times as well as a higher variability and inconsistency between aforementioned measurements [22, 23].

2.4 Analysis of Gait

2.4.1 Gait

Gait is defined as the cyclical pattern of walking [9]. One complete human gait cycle is measured as the time recorded between two consecutive footfalls of the same foot [10, 24]. In other words, the time that is taken between two successive left-foot or right-foot heel-strikes on the ground [10, 24]. Gait cycle is divided into two phases, the time in which the foot is on the ground called the stance phase, and the time in which the foot is in the air (or moving forward when a person is unable to lift their foot off the ground) called swing phase [24, 25]. Figure 2.1 illustrates the human gait cycle in terms of its stance phase and swing phase.

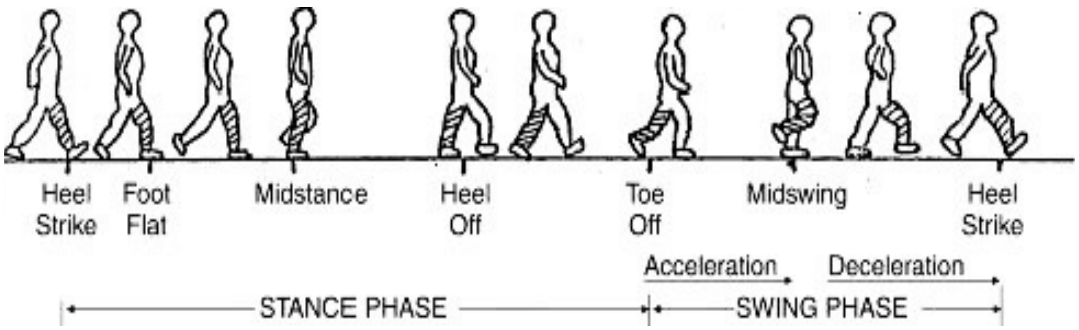

Figure 2.1: The Human gait cycle, in terms of stance and swing phase, where one complete gait cycle is measured as the time recorded between two consecutive footfalls of the same foot [24]

2.4.2 Gait Abnormalities

Owing to the fact that human gait is repetitive, gait abnormality is defined as non-cyclic movement [10]. This occurs when a person does not walk in the usual way as a result of physical injury or neurological disorders such as Parkinson's Disease [10]. Gait abnormalities can be broken down into five categories [26]:

1. Spastic gait where a person drags their feet while walking.
2. Scissors gait where a person's legs cross and sometimes hit each other while walking.
3. Steppage gait where a person's toes point towards the ground while walking.
4. Waddling gait where a person sways from side to side while walking.
5. Propulsive gait where a person walks with his or her head and neck pushed forward.

Thus, a patient suffering from these gait abnormalities, including PD patient, may struggle to walk normally during daily activities and clinical walking tests [26].

2.5 Gait Analysis Systems

Although Parkinson's Disease is identified through clinical observations [5], numerous methods have been developed that can assist in the quantification of disease diagnosis, severity, and treatment efficacy. These systems measure a subject's gait or walking parameters through optical or kinematic sensors in order to draw conclusions [27].

2.5.1 Optical Motion Capture Systems

This type of system aims to measure gait using cameras and optical sensors to compute the spatiotemporal parameters of a subject [27]. Such systems include but are not limited to Optogait, Vicon, Codamotion and Xsens [28, 29, 30, 31]. Within a controlled test environment consisting of a treadmill or set walkway, these systems capture and record the real-time motion of a subject using their respective sensors to generate a precise quantitative gait model. This model is then processed and conclusions regarding patient walking patterns are drawn [28, 29, 30, 31].

Although optical motion capture systems have shown high accuracy in determining gait, they are expensive, require a dedicated testing environment and are complicated to set up [28, 29, 30, 31]. Moreover, patients suffering from more severe forms of PD may require a walking aid. Therefore, an optical motion capture system is deemed as an inappropriate gait monitoring solution for this research.

2.5.2 Instrumented Walkway Systems

Instrumented walkway systems are used to acquire quantitative gait and balance measurements using the pressure that a patient's foot exerts while performing a walking protocol on an instrumented mat [32, 33]. Available in a range of lengths and widths, the GaitRite, GaitMat II and Zeno Walkway systems measure the spatial and temporal parameters of gait [33, 34, 35]. This is achieved by using thousands of individual force sensors placed along each mat [32, 33].

Instrumented walkway systems are portable, cost-effective and relatively simple to set up [33]. However, this type of system has a length constraint and can only be used under specific conditions. Thus, it cannot measure a patient's gait cycle during daily activities. Given these limitations and the fact that severely affected PD patients may require a walker, this approach is not suitable for this study.

2.5.3 Wearable Systems

Wearable technology has been proposed and implemented as an alternative gait measurement system. Some wearable systems in research include but are not limited to an instrumented shoe gait analysis system as suggested by Bamberg *et al* [13], a smart-watch gait recognition technology as proposed by Xu *et al* [14], and an ear-worn gait measurement system as described by Li *et al* [15]. These types of systems utilise sensors such as accelerometers, gyroscopes, magnetometers, force sensors and sonars that are placed on a person's body to compute the spatiotemporal parameters of the subject [13, 14, 15]. Thereafter, using signal processing, a subject's gait parameters in the form of velocities, accelerations, stride lengths, stride times, total walk times and cadence are able to be extracted and conclusions regarding the patient can be made [13, 14, 15].

Although wearable gait analysis technology can be used as a potential solution in gait measurement it does impose limitations on a patient's lifestyle. As described in [16], wearable devices are often uncomfortable and therefore, impact a patient's use of the product. Moreover, patients that suffer from PD may have hand tremors which make it difficult to strap the sensors themselves, consequently requiring assistance or in more severe cases the additional aid of a walker. Therefore, this form of analysis tool is unsuitable as a gait monitoring solution for this research.

2.5.4 Instrumented Walker Systems

An instrumented walker is a standard walker that has been equipped with sensors that record the kinematic data of a subject during use [16]. This type of device has been designed to allow a patient the freedom to move around without the stress of attaching wearable sensors or the setting up of complex machinery. It allows a patient to move freely as they would with the stability of a walker while recording information pertaining to the way in which they walk. This data is then processed using signal processing techniques to evaluate a subject's spatial and temporal gait parameters [16].

Alwan *et al,* in [36] discuss the design of an assisted walker that utilises force sensors in order to calculate the moments exerted by a subject on the walker's handles. Subsequently, from these results, subject footfalls and gait parameters can be extrapolated [36].

Moreover, [16] utilise the Personal Aid for Mobility and Health Monitoring II (PAMMII) instrumented walker (as shown in Figure 2.2) to extract gait parameters and features pertaining to subjects using low-cost distance encoders, force sensors, and an accelerometer. This device was successfully used to diagnose PD patients from controls with sensitivity and specificity scores ranging between 85% and 91% [16].

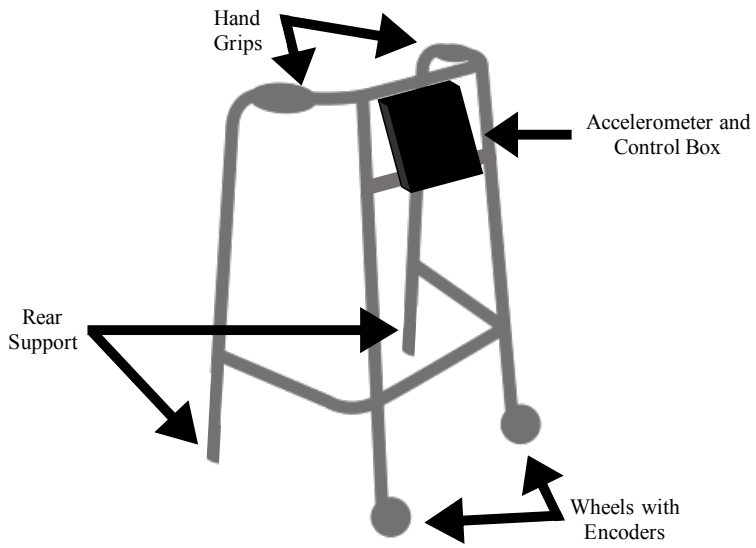

Figure 2.2: Diagram of the PAMMII instrumented walker, highlighting the placement of its sensors and control box

Owing to the success of the abovementioned PAMMII study as well as the fact that the PAMMII system is an affordable and easy-to-use solution that requires no professional labour to measure the gait of a PD patient, in both a clinic and home setting on a daily basis or as needed, the PAMMII instrumented walker is seen as a suitable gait monitoring solution for this research.

2.6 Discrimination Systems Based on Kinematic Signals

Previous studies (as described in Section 2.5 [28, 29, 30, 31, 32, 33, 34, 13, 14, 15].) have shown that signals obtained from kinematic sensors demonstrated various degrees of reliability when discriminating gait anomalies of PD patients.

Particularly, recent studies have gone into the classification and differentiation between PD patients and controls or discrimination between PD patients in different stages of the disease using signals acquired from walker mounted sensors, applying both machine learning and statistical approaches (albeit to a lesser extent) [37, 38]. However, up until now (to the knowledge of the author), data captured from walker mounted sensors has *not* been used to assess subtle changes like the effects of treatment on a PD patient.

Research has shown this type of accurate classification task necessitates extensive signal processing, feature extraction, and feature selection to classify results and produce discrimination sensitivity [25, 12, 27, 16, 39]. Figure 2.3 illustrates how signal processing research typically address similar discrimination problems of raw data, (kinematic or other). Variations of this diagram exist and have been used extensively for many data types, including medical data, and in PD research [25, 12, 27, 16, 39]. The following subsections explain each section of the flow diagram in more detail.

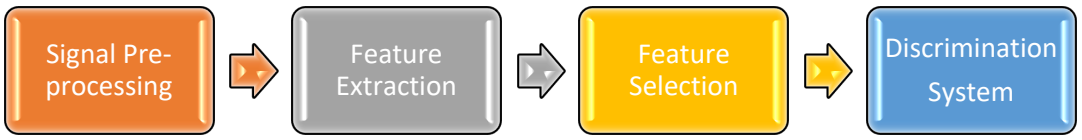

Figure 2.3: Flow diagram of a typical discrimination system research task

2.6.1 Signal Pre-processing

Signal pre-processing is a set of processes that are necessary for most real-world signals to improve the quality of the signals before further processing is applied [40]. Among these processes are noise reduction, raw signals discrepancies correction, segmentation of the signal to sub-parts and transformations. Effective signal pre-processing allows for an accurate feature extraction and more comprehensive classification [40]. Pre-processing is particularly important when the signals are acquired by low-cost sensors, like the ones used in instrumented walkers for gait disorders analysis [16]. Specifically, previous studies proposed and tested numerous types of filters that were able to improve gait signals acquired from PD patients. [23, 32, 36, 41].

2.6.1.1 Signal Filtering

Multiple types of filtering techniques in the context of kinematic signals acquired from PD patients have been explored. Among the prevalent ones are the Finite Impulse Response (FIR) filter, the Infinite Impulse Response (IIR) filters, wavelet de-noising and Empirical Mode Decomposition (EMD).

In literature, a low pass filter is regularly applied to eliminate the high-frequency noise of a person's gait signal. This stems from the fact that human muscle movement has a frequency lower than 16Hz and step frequency smaller than 3Hz [42]. PD patients' symptoms produce additional noise which requires particular filtering.

Dai *et al* investigate filtering raw accelerometer and gyroscope signals using both high-pass FIR and IIR filters with a cut-off frequency of 3.25Hz to detect tremors in PD patients. These filters were implemented in order to reduce the influence of dyskinesia on the measured signals [43]. Moreover, [44] filters raw accelerometer data using a low-pass IIR filter as well as a high-pass IIR filter with cut-off frequencies of 3Hz and 1Hz respectively to estimate the spatiotemporal parameters of a person. This filtering was performed to reduce the effect of bradykinesia, dyskinesia, and the noise of a person's body movement [44].

Another previous research implemented and tested a 4th order low-pass Butterworth filter with cut-off frequency of 2.5Hz to reduce the high-frequency noise of the accelerometer data PD patients and control subjects [16]. This study also employed wavelet de-noising to filter accelerometer signals [16].

Lastly, Ballestoros *et al* and Nunes *et al* used EMD to filter an acceleration signal similar to the kinematic signals of this research task signal [39, 45]. This method decomposes the signal into a number of zero-mean bandlimited oscillatory components, called Intrinsic Mode Functions (IMF). Each IMF is analysed in terms of its data variation where IMF signals with lower variations are removed. The remaining IMF levels are then summed creating a filtered version of the original signal [39, 45]. It is important to note that the signal noise is found in the initial IMF levels such as the first and second levels [39, 45].

2.6.1.2 3-Axis Signal Rotation

An essential component of accelerometer data pre-processing is rotational correction [16, 46, 47]. This is performed to align the accelerometer signal anteroposterior to the direction of movement i.e. ensuring the X-axis and Y-axis of the accelerometer signal is in line with the direction of movement and Z-axis is facing upwards [16, 46]. This must be done as accelerometer measurements include the effects of gravity ($9.81 m/s^2$) therefore, data must be rotated in order to separate a person's acceleration from gravity so that only true movement quantities remain [16, 46]. Favre *et al* applied a rotational correction using a quaternion rotation matrix to adjust accelerometer data to be anteroposterior to the direction of movement [46]. Moreover [48], compared a Kalman filter, Tilt Kalman filter, and the Madgwick algorithm to estimate z-axis orientation and correct rotation, in real time.

2.6.1.3 Footfalls Detection

Owing to the rationale discussed in Section 2.3, the detection of footfalls or when a foot touches the ground is seen as a fundamental requirement for this research task. It is necessary to assist the feature extraction methods in isolating key information pertaining to a subject's gait performance such as stride lengths, stride velocities and stride accelerations etc. [10, 23, 27, 13, 49] (further described in Section 2.6.2).

Gait disorders studies and in particularly PD gait studies used accelerometer signals as the basis to derive gait features that can characterise the difference between patients and control subjects [10, 23, 27, 13, 49]. Byan *et al* used peak detection of a rotated and filtered on-body accelerometer signal as illustrated in Figure 2.4 to identify subject footfalls. As shown, in an ideal, noiseless signal, the peaks of the accelerometer signal represent when a subject's foot touches the ground of footfall [50].

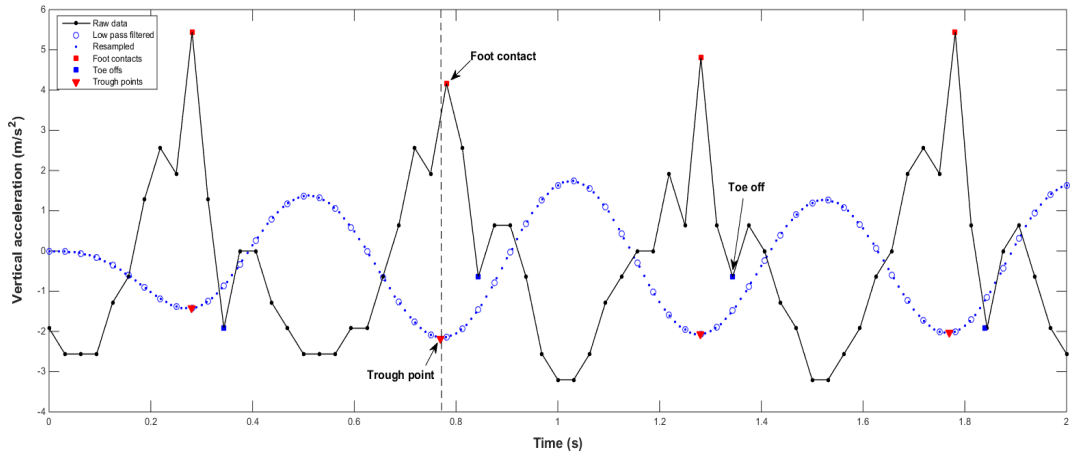

Figure 2.4: Accelerometer based footfalls detection from [50] where red stars on each peak of the signal indicate footsteps

2.6.2 Feature Extraction

Features are defined as mathematical representations of a signal's characteristics. In other words, a feature portrays the distinctive and distinguishing traits of a signal [25, 23]. In gait analysis, features are extracted to quantify how a person walks.

Previous gait analysis studies proposed numerous features of gait such as stride length and stride [10, 23, 27, 13, 49]. Moreover, these studies of gait disorders and particularly PD gait used accelerometer signals and the extracted features to characterise the difference between patients and control subjects [10, 23, 27, 13, 49]. Table 2.1 illustrates examples of features derived in these studies.

Table 2.1: Example table of derived features, in terms of gait and sensor features, found in literature

Derived Features	
Gait Features	***Sensor Features***
Step velocity, step time and step length [27, 13, 16]	3-axes accelerometer signal cross correlation [36, 41]
Standard deviation of step velocity, step time and step length [27, 13, 16]	Accelerometer signal, max, mean and median frequency and bandwidth [51]
Turn time and turn velocity [41]	Kurtosis [49, 51]
Cadence (steps per minute) [51]	Harmonic ratio [23, 27, 51]
Total number of steps taken [12, 44]	Step Regularity and Asymmetry [51, 52]

2.6.3 Feature Selection

Feature extraction often results in a large number of features, or a high dimensional feature space, which may contain redundant and extraneous information [51]. Additionally, research has shown that discrimination systems, such as machine learning algorithms and statistical measures are adversely affected by a large feature space where irrelevant and redundant data is present [53]. Feature selection techniques reduce the size of a feature space by projecting the original high dimensional feature set onto a new smaller dimensional feature space or by selecting relevant feature subsets from the extracted features [44, 51, 54]. Feature selection methods were used to differentiate between PD patients and controls and were found to improve the discrimination system performance, in terms of accuracy and efficiency [53]. The proceeding subsections detail the two types of feature selection methods that exist namely; projection methods and criteria-based feature selection.

2.6.3.1 Projection-Based Feature Selection

Of the surveyed literature, the most common projection-based feature selection method used is Principal Component Analysis (PCA) [16, 39, 55]. Previous studies used PCA to reduce the high-dimensionality of the derived gait characteristic feature-space with the aim of differentiating between PD patients and controls [16, 39, 55].

PCA achieves this by orthogonally transforming the input feature-space in a way that maximises the variability of the output feature-space [56]. The output feature-space is mapped to maximise variance using a linear combination of the original features. Consequently, the feature-space output of the PCA algorithm consists of a smaller set of derived variables, called principal components [56]. The following subsection outlines how the PCA algorithm is performed.

2.6.3.1.1 Principal Component Analysis

Firstly, the original feature space must be normalised. This is achieved using Equation 1 [57].

Equation 1: PCA Normalisation formula [57]

$$Normalised\ X_i = \sum_{i=1}^{n} \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

where,

- X_i is a single extracted feature value
- X_{max} and X_{min} are the maximum and minimum values of the extracted feature across all subjects

Thereafter, the covariance of the features is calculated as shown in Equation 2. This is performed to measure how the variables move together [56].

Equation 2: PCA Covariance formula [56]

$$X = X^T \times X$$

where,

- X is the matrix of the feature-space

Once completed, the data is rotated to fit a new axis that maximises variance [56]. This is achieved by calculating the eigenvectors and eigenvalues of the covariance matrix. Subsequently, the feature-space matrix (X) is projected onto the new axis using Equation 3 to obtain the principle components of the PCA transformation.

Equation 3: Principal component calculation [56]

$$PC = X^T \times E_{largest}$$

where,

- X is the matrix of the feature-space
- $E_{largest}$ is the matrix of eigenvectors that corresponds to the largest eigenvalue

2.6.3.2 Criteria-Based Feature Selection

Unlike projection-based feature selection, criteria-based feature selection performs dimensionality reduction without creating new feature variables through transformation.

Subsequently, clinicians may still analyse the PD patient's derived data results to comprehend system responses [58]. According to literature, two types of criteria-based feature selection methods are used to assess PD patients derived gait characteristics and features namely: wrapper and filter techniques [58].

2.6.3.2.1 Wrapper Feature Selection Method

Wrapper feature selection methods identify a variety of features subsets based on relevance and correlation to the discrimination objective as stipulated by a professional i.e. in the case of this PD research, this would be as specified by a healthcare professional or neurologist [58].

2.6.3.2.2 Filter Feature Selection Method

Contrasting to the aforementioned technique, the filter selection method does not rely on human selection. This method uses statistical methods to evaluate the significance of a subset of features. Statistical methods aim to summarise and quantify groups in the data in terms of statistical measures such as means, variances, standard deviations, standard mean errors and confidence intervals to test hypotheses [59]. Consequently, this method may provide objective feature subsets as opposed to the wrapper technique [58].

From literature, statistical methods employed in studies that analyse subjects' gait and movement characteristics include but are not limited to; T-tests, paired T-tests, ANOVA tests and post-hoc Bonferroni tests [59, 60, 61, 38]. These tests assess the likelihood that measured changes in patient condition are not random on an individual feature basis i.e. the statistical significance (p-value) of the changes in each feature, as well as the probability that measured mean changes in features are not random across subjects. Specifically, these types methods have been applied in their respective studies to discriminate treatment effects in terms of statistical differences in medication responses of pre-defined subject groups [59, 61]. Additionally, Huang *et al* utilise the widely used Global Statistical Test (GST) as well as multivariate techniques to assess and calculate the overall medication impact of clinically relevant outcomes for long-term PD patients [38].

2.6.4 Discrimination Systems

Classification and discrimination tasks often employ machine learning. The goal of a machine learning algorithm is to classify data in terms of the different patterns found. Consequently, it

may be applied to this research task's problem domain in the form of a supervised and unsupervised learning algorithm [62].

2.6.4.1 Unsupervised Learning Algorithm

Unsupervised learning algorithms attempt to model the fundamental distributions and characteristics of data in order to determine results. Thus, this type of algorithm does not label data and effectively develops the input to output mapping on its own [62]. Usually, this type of machine learning algorithm is implemented when data labels are not explicitly stated as it allows the machine to decide on what features are important. In terms of this PD research, an unsupervised learning algorithm may be useful as it can use the underlying fundamental characteristics of derived features to classify a patient and does not rely on the potentially subjective views of clinicians and pre-assigned data labels.

From literature, the most common unsupervised learning algorithm in PD assessment follows in the form of the K-Means Clustering method [63].

2.6.4.1.1 K-Means Clustering

The K-Means Clustering uses vector quantisation to partition features into n-clusters that categorise the dataset without predefined knowledge about features [63].

A previous study implemented the K-Means Clustering algorithm to assess whether a patient suffers from PD. Results show a sensitivity and specificity of 85% and 91% respectively [16]. Similarly, another study used K-Means Clustering to infer if a person has PD and showcases results of 89% and 79% for sensitivity and specificity respectively [58].

2.6.4.2 Supervised Learning Algorithm

Within a supervised learning algorithm, the aim is to approximate a mapping function that can determine an output based on a specific input. In terms of PD classification, this is achieved using pre-assigned labels of features that are deemed important by feature selection methods to train the implemented algorithm through iterative measures, to predict data specific results [62].

Literature has shown four supervised learning techniques that have been used extensively in different types of PD classification namely: Support Vector Machines (SVM), Artificial Neural Networks (ANN), Naïve Bayes and Bagged Decision Trees (BDT) [63].

2.6.4.2.1 Support Vector Machines

SVM is classified as a non-probabilistic binary linear classifier although research has shown it can be extended to multi-class classification algorithms [63]. SVMs map feature spaces using kernels to higher dimensional spaces using non-linear decision boundaries. Thus, SVM allows features to be linearly separable [53].

In the field of PD vs non-PD classification, Tsanas *et al* utilised SVMs in conjunction with random forests to classify PD patients noninvasively. This was achieved using only ten dysphonia features, showcasing an overall classification accuracy of 99% [37]. Similarly, Das *et al* used an SVM classifier to differentiate mild and severe symptoms in PD patients. Test results for this experiment illustrated an average accuracy of 90% [64].

2.6.4.2.2 Artificial Neural Networks

An ANN is a type of connectionist system that has been designed based on biological neural networks such as animal brains [63]. ANN aims to develop a system that learns from its training examples. It achieves this by extrapolating rules from training data to recognise and classify trends in new input data [63].

A previous study discussed the design and implementation of a Radial Basis Function Neural Network. This system was used to recognise and predict the onset of the PD tremors in human subjects. This system showcased accuracies of 89% [65]. Additionally, another research illustrated an accuracy of 96% using an ANN to classify if a patient has PD [66].

2.6.4.2.3 Naïve Bayes

This type of classifier is a probabilistic classifier that stems from Bayes' theorem and has strong assumptions regarding its features [63]. Thus, the Naïve Bayes method assumes that a classification problem can be modelled probabilistically with its data approximated by a priori probability distribution. Consequently, for particular models, the Naïve Bayes algorithm can seem to oversimplify the problem [63].

Tripoliti *et al* used the Naïve Bayes algorithm to detect freezing of gait events in patients that suffer from Parkinson's Disease. This system showed an accuracy of 96% [67].

2.6.4.2.4 Decision Trees

A Decision Tree (DT) is a non-parametric supervised learning method [68]. It is a predictive process that uses a branching method to create a set of rules that classifies data into categories based on the importance of the given feature set [68, 69]. It achieves this by using a set of if-then-else decision rules to classify data. Subsequently, it partitions a dataset into smaller and smaller subsets to arrive at a decision [68].

A DT comprises of three types of nodes namely [68, 69].

- A root node that corresponds to the best predictor of the class,
- A decision node which is defined as a node that has two or more nodes attached to it and,
- An end node which represents the verdict of the DT

Figure 2.5 illustrates how a DT splits a dataset with predictors smaller than the criterion being placed on the left and those larger than the criterion being placed on the right [68, 69]. The DT splits the feature-space using the Gini Index as depicted in Equation 4. The Gini Index is a measure of the impurity in the features. In other words, it is a measure of the probability that a randomly selected feature from the dataset is falsely predicted if it is randomly labelled according to the distribution of labels in the subset [68, 69].

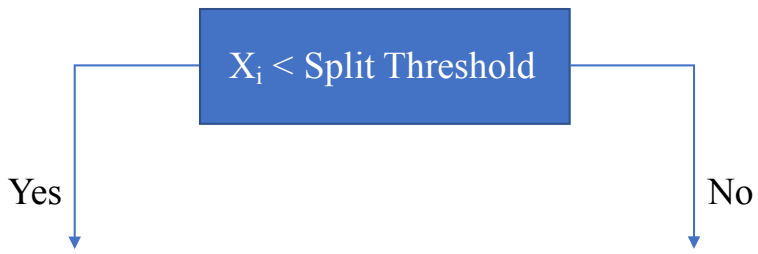

Figure 2.5: Example of a Decision Tree split

Equation 4: Gini Index formula [68, 69]

$$Gini(X_i) = 1 - \sum_{i=1}^{n} (p_i)^2$$

where,

- n is the number of features
- p_i is the relative frequency of each feature

2.6.4.2.5 Bagged Decision Trees (Bagging)

This type of machine learning algorithm combines multiple decision trees, reducing data variance when predicting an outcome [68]. As mentioned in the previous subsection, a decision tree is defined as a predictive process that uses a branching method to create a set of rules that classifies data into categories based on the importance of the given feature set [68].

As shown in Figure 2.6, the BDT algorithm expands on this. It divides the input data into n-subsets using uniform sampling with replacement. Each decision tree considers the most important features across its data subset (sample n) and then creates a list of decisions which best partitions the data sample [68, 70]. Afterwards, the results of all decision trees are equally weighted and averaged to create the final bagged decision tree output classification. Consequently, no features are ever discarded, and the BDT algorithm is considered a more advanced version of the decision tree method as it overcomes individual decision tree overfitting issues [68, 70].

Wu *et al* discussed the analysis of voice patterns to detect PD in patients utilising Bagging Ensemble Machine Learning Methods. This system produced an accuracy of 90.8% with a sensitivity of 98.0% [71]. Another study employed a Bagging algorithm to classify patients with PD from controls with a successful detection rate of 98.3% [72].

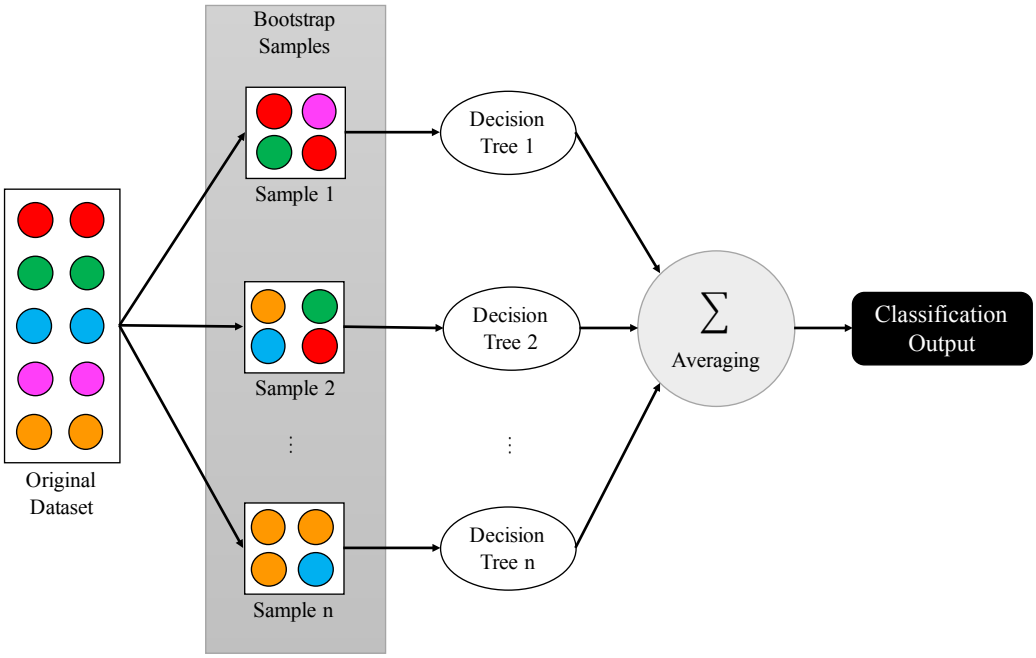

Figure 2.6: Bagged Decision Tree Algorithm overview, highlighting how bootstrap aggregating is used to resample data with replacement

The abovementioned supervised and unsupervised machine learning algorithms maintain similar detection rates for their particular classification tasks. However, owing to the fact that data for this research task is pre-labelled in terms of three classification groups (treatment patients, non-treatment patients, and controls) and H&Y clinical scores (0, 2, 3) respectively, a supervised machine learning approach is deemed more appropriate for this research task.

Moreover, the pre-acquired database is limited in size and data outliers are unknown. With this in mind, a BDT classifier is seen as simpler to implement and train when compared to SVM, ANN, and Naïve Bayes classifiers to generate a robust classification output.

PD is a neurodegenerative disease. It impairs a patient's motor movements compromising their independence and flexibility [4]. These symptoms culminate in problems with the manner of a patient's walking or gait. PD diagnostic and assessment scales exist in the form of the UPDRS, H&Y, and TUG tests. Limitations of these scales necessitate the presence of a healthcare professional to utilise their personal experience and knowledge to diagnose and monitor the patient which can introduce personal bias effects. Technologies have been developed to negate these deficiencies and have demonstrated various degrees of reliability when discriminating gait anomalies of PD patients from control groups. However, they have not been utilised to date (to the knowledge of the author) to evaluate the effects of short-term

treatments on PD patients. Nevertheless, research has shown this type of classification task requires extensive signal processing, feature extraction, and feature selection to classify results and produce a successful discrimination accuracy [25, 12, 27, 16, 39].

Chapter 3. Methodology

3.1 High Level Overview

Figure 3.1 provides a block diagram overview of the signal processing system that is used to answer the research question. Each block of this diagram will be detailed in the subsequent sections.

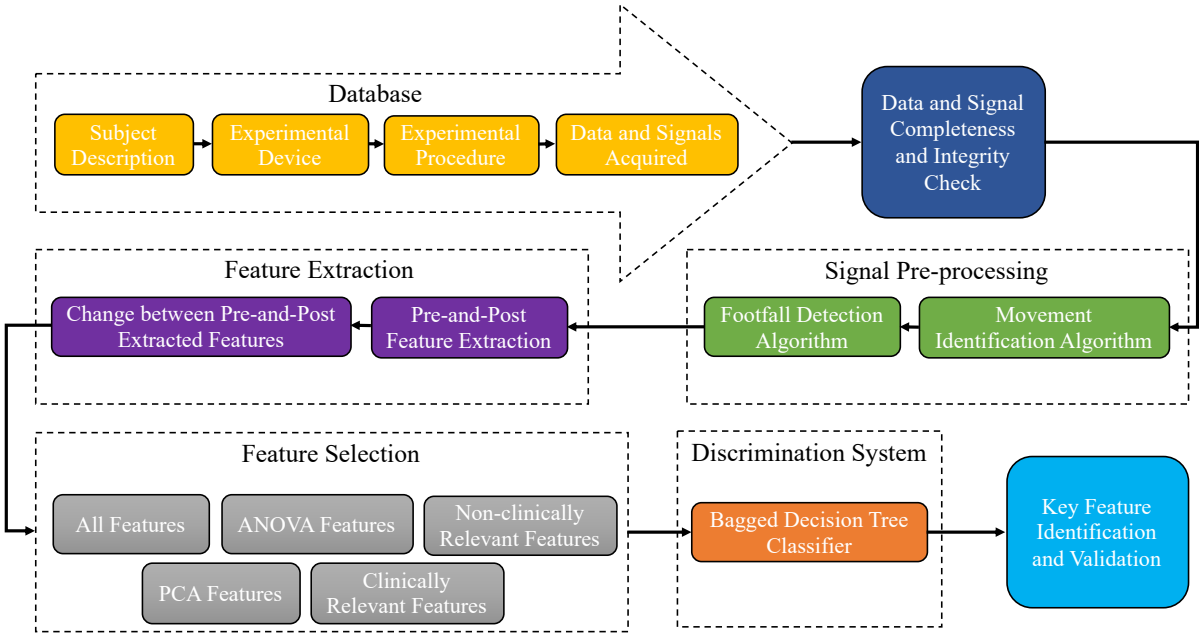

Figure 3.1: Block diagram describing the methodology implemented to answer the research question

3.2 Database

3.2.1 Subject Description

This research aims to assess the extent that features extracted from signals acquired by walker mounted sensors can identify changes in PD patient gait, due to treatment. This requires a suitably sized database which includes *two* tests – before and after treatment, from PD patients who received medication (treatment patients) as well as *two* tests from PD patients who

received no medication (non-treatment patients) and control subjects. To meet this criterion, this study reuses a database that has been acquired at the Rambam Medical Centre in Israel. This database was recorded using the PAMMII instrumented walker owing to rationale discussed throughout Section 2.5 and Section 3.2.2.

The existing database consists of 37 treatment patients, 6 non-treatment patients and 11 controls where each subject performed *two* 3m and 10m TUG walking tests (as outlined in Section 3.2.3). For the treatment patients' case, the tests were performed before the start of the treatment and after the end of the treatment, which took a total 2-3 hours. The non-treatment patients and the controls were likewise tested *twice*, where the two tests were completed 3 hours apart. In each patient test session, a healthcare professional assessed and recorded the H&Y PD clinical severity scores and the control subjects were assigned an H&Y 0 severity score.

This PAMMII instrumented walker experiment was performed in accordance with the ethical standards of the Rambam Medical Centre (Ethics Number: 0108-13-RMB) under the Helsinki Declaration. Furthermore, ethics for the secondary use of this data has been granted by the Human Research Ethics Committee (HREC) of the University of the Witwatersrand (Clearance Certificate Number: M180203) with ethic clearance documentation found in Appendix A.

3.2.2 Experimental Device

The PAMMII instrumented walker was used to record each subjects' kinematic data for this research. As described in Section 2.5.4, this device offers an affordable, user-friendly solution for kinematic signals acquisition, and has shown good results in the diagnosis of PD [16].

The sensors used for this research are:

- Two digital distance encoders
- An ADXL335 tri-axes accelerometer

These sensors are controlled by an Arduino Nano V3 microcontroller with a sampling rate of 21.5Hz to measure the trajectory a user has walked (encoders) as well as a user's acceleration (accelerometers) with a range of ±16g and resolution of 0.1g along each axis (X, Y, and Z axes).

Figure 3.2 illustrates the location of all sensors and the controller. Moreover, the microcontroller writes all recorded signal data to an SD card in the form of a CSV file. The data recorded is detailed in Section 3.2.4.

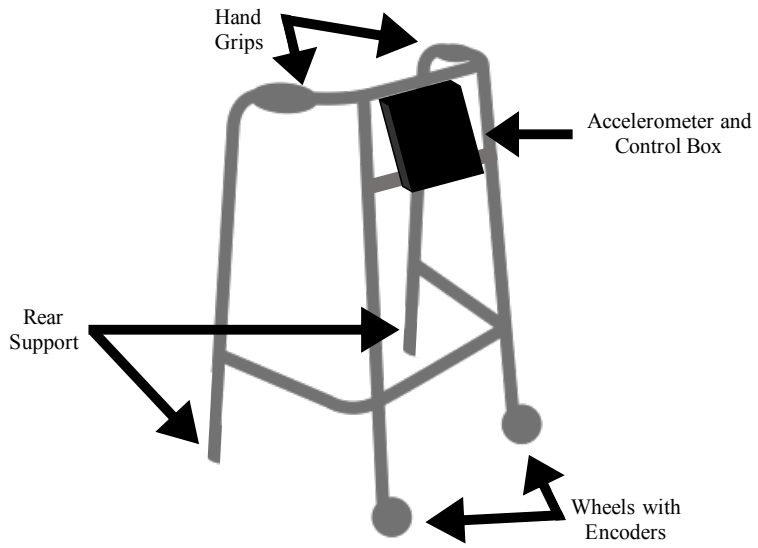

Figure 3.2: PAMMII schematic diagram, highlighting the placement of its sensors and control box

3.2.3 Experimental Procedure

The experimental procedure was performed for each subject group (treatment patients, non-treatment patients, and controls) *twice* using the TUG protocol (as outlined in Section 2.2.3) for both a 3m and 10m walking test: In the treatment patients, the first testing was performed before the treatment, and the second after treatment in the form of intravenous Amantadine-Sulphate IV, In the non-treatment patients and controls case, the first testing was performed and after a delay (approximately 3 hours) simulating treatment time, the second walking test was performed.

3.2.4 Data and Signals Acquired

The PAMMII system recorded the following signals during each subjects' TUG tests:

- A time vector,
- Three acceleration signals (X, Y and Z axes), and
- Two distance measurements.

Additionally, the clinician who oversaw the TUG tests recorded the following clinical information pertaining to each subject in the trial:

- The subject's age,
- The subject's gender, and
- The subject's H&Y clinical score.

These data signals and the subjects' respective clinical information was recorded for the 54 subjects in the dataset (37 treatment patients, 6 non-treatment patients, and 11 controls).

3.3 Data and Signal Completeness and Integrity Check

As usual, in research which re-uses information, especially data that was acquired in a busy hospital setting, data must be authenticated and primed such that accurate and reliable analysis is performed. Thus, the following data and signal completeness and integrity checks have been implemented to ensure the dataset is not missing information.

1. Ensured that each subject has performed a 3m and 10m TUG test pre-and-post treatment.
2. Ensured that data within the raw signals for each subject is meaningful and *not* only noise.
3. Ensured that each subject has their respective clinical details recorded.

3.4 Signal Pre-processing

As gait is defined as the way in which a person walks (Section 2.4.1), it is realised that determining when a subject's foot touches the ground while walking is an essential component prior to feature extraction [10, 23, 27, 13, 49]. Therefore, the aim of signal pre-processing for this research task consists of preparing data such that it can be used for feature extraction in terms of subject gait movement i.e. determining when a subject is walking, determining when a subject is turning/stationary, and determining a subjects' footsteps while walking. With this goal in mind, numerous pre-processing techniques have been implemented.

As discussed in Section 2.5.4, the PAMMII system employs low-cost hardware and sensors. Consequently, on occasion, a sample is missed. In order to counteract this, data interpolation is applied to all raw data. This being said, algorithms are still needed to establish when a

subject is walking and turning/stationary as well as when a subject's foot touches the ground while walking for gait extraction. Therefore, a Movement Identification Algorithm and a Footfall Detection Algorithm have been created. The study described in this proceeding subsections was performed by the author of this thesis, in collaboration, and published [73]. Furthermore, it is important to note that all signal filtering methods described in these subsections were performed on the respective raw data signals to derive all the features extracted as outlined in Section 3.5.

3.4.1 Movement Identification Algorithm

Typically, a movement is composed of several stages, which are different in properties (i.e. straight line vs. circular) and/or separated by no-motion periods. Using the TUG protocol, it is seen that the subjects have two phases of straight-line walking that is separated by a turning phase. This being said, these walking patterns are seen to vary across subjects. Some subjects complete the protocol without a break in a timely manner while other subjects stop during the test for short or long periods as needed. Potentially, these periods could be used as an indicator of PD highlighting the symptoms/physical limitations of the patient (e.g. reduced coordination, postural instability and/or impaired balance). The succeeding subsections explain how this algorithm differentiates between these stages

3.4.1.1 Encoder Signal Preparation

This algorithm utilises the recorded data from the left and right distance encoders as well as its corresponding time stamps. Firstly, all acquired signals are transformed into usable SI units. Therefore, each encoder value is converted from distance travelled in encoder ticks to meters using Equation 5 where values are found from the encoder datasheet and the corresponding time stamps are converted from milliseconds to seconds.

Equation 5: Encoder Ticks to Meters Conversion Ratio Formula

$$Conversion\ Ratio = \frac{\pi * (Wheel\ Diameter\ in\ meters)}{(Number\ of\ Ticks\ per\ Revolution)}$$

Thereafter, the average distance is calculated from the left and right distance encoders and the velocity of the left, right and average distance encoder signals is calculated using numerical differentiation.

3.4.1.2 Signal Filtering

The distance and derived velocity signals are filtered using a Zero-phase Moving Average digital filter to smooth the input data and diminish phase distortion that can result from filtering.

3.4.1.3 Movement Stage Selection

The implemented movement selection algorithm aims to identify the walking periods of a subject and consequently infer when a subject is turning. It identifies this walking region using the distance and velocity plots for the left, right and average encoder signals. Subsequently, logic is used to incorporate the results of all outputs such that a point is only deemed a walking period if all the outputs agree. The implemented walking period selection logic works by traversing through the distance and velocity signals for each plot (left, right and average) in one-second intervals (or windows) until the end of the signal. The flow diagram in Figure 3.3 illustrates the applied logic of this algorithm.

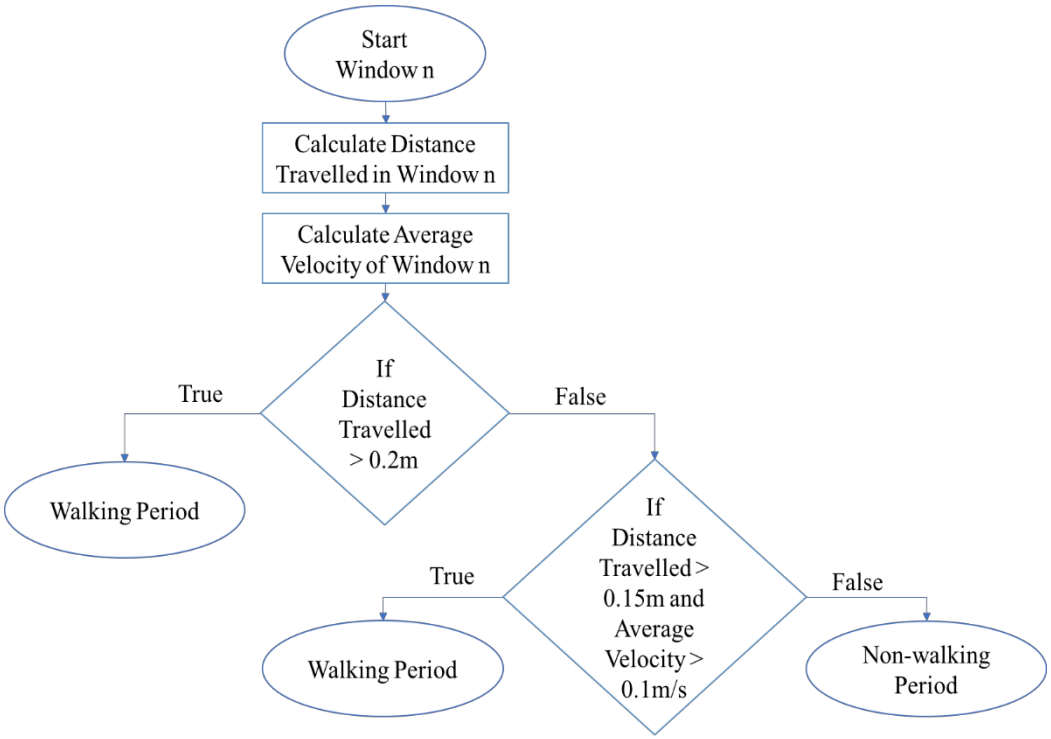

Figure 3.3: Flow diagram describing how the Movement Stage Identification Algorithm determines movement and non-movement periods

3.4.2 Footfall Detection Algorithm

A basic event in the gait cycle, which decomposes it into elementary sections is the footfall. As discussed throughout Section 2.3 there is a relationship between PD and gait. Moreover, gait analysis systems as detailed in Section 2.5 see the footfall as a necessity to calculate reliable gait features. Wearable sensors can provide footfall detection more readily, sensing the "thump" of the foot. For off-body sensors, like the instrumented walker, this task is more challenging as additional noise is added into the system from the 3rd party apparatus that is used to record the measurement.

As outlined in Section 2.6.1.3, once an accelerometer signal has been filtered, it is possible to identify footfalls using the peak detection. This algorithm uses this logic to predict when a subject's foot touches the ground while walking *only*.

3.4.2.1 Accelerometer Signal Preparation

This algorithm utilises the recorded data from the accelerometer signals (X-axis, Y-axis, and Z-axis), its corresponding time stamps and the determined walking periods from the Movement Identification Algorithm to determine footfalls. Firstly, the timestamps are converted from milliseconds to seconds. The Rodriguez rotation matrix is used to align the accelerometer data points to be anteroposterior to the direction of movement (as outlined in Section 2.6.1.2). This is achieved by calculating the normalised cross-product between each accelerometer data point and the required axis ($a = [0,0,1]$) to work out the fixed unit vector of rotation ($w = [w_x, w_y, w_z]$) [47]. Afterwards, the Rodriguez rotation matrix is calculated using Equation 6.

Equation 6: Rodriguez Rotation Matrix formula [47]

$$R_{\widetilde{w}}(\theta) = I + \widetilde{w}sin(\theta) + \widetilde{w}^2(1 - \cos(\theta))$$

where,

- I denotes a 3x3 Identity Matrix
- $\widetilde{w}$ denotes the antisymmetric matrix with entries $\widetilde{w} = \begin{pmatrix} 0 & -w_z & w_y \\ w_z & 0 & w_x \\ -w_y & w_x & 0 \end{pmatrix}$
- θ denotes the normalised dot-product between each accelerometer data point and the required axis

Subsequently, the gravity component is removed from the aligned accelerometer data.

3.4.2.2 Filtering and Footfall Detection

The PAMMII recorded signals are non-stationary time-series signals. Therefore, owing to the basis of Empirical Mode Decomposition (EMD) that a signal is composed of a superposition of fast and slow oscillatory components [39, 45], the accelerometer signals are filtered using EMD.

The mathematical representation of the EMD decomposed signal is shown in Equation 7. This filtering technique iteratively decomposes a non-stationary signal into a number of zero-mean bandlimited oscillatory components, called Intrinsic Mode Functions (IMFs). These IMFs are sorted such that each proceeding IMF level consists of lower frequency oscillations.

Equation 7: Mathematical representation of an EMD signal [39, 45]

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

where,

- $x(t)$ is the original signal,
- $c(t)$ is the IMF decomposition of the signal, and
- $r(t)$ is the remaining signal

Figure 3.4 illustrates how each IMF is calculated. Initially, the maxima and minima of the input signal are determined using cubic splines to produce a signal envelope. Next, the mean of these envelopes is calculated and subtracted from the input signal to obtain a new signal. This is repeated until the IMF conditions are met such that 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 zero. Once these conditions are met, the signal is extracted as the nth order IMF [39, 45]. Afterwards, the remaining signal is taken as the new input. These steps are repeated until a stopping criterion of a standard deviation less than 0.3 is obtained [39, 45].

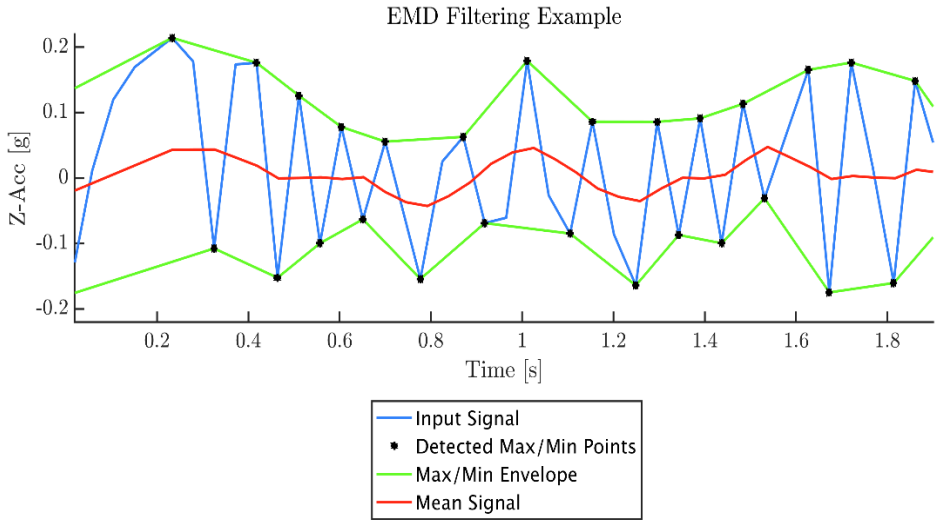

Figure 3.4: Example of how each IMF level is calculated

In order to test which IMF level to reconstruct the accelerometer signals from, the Z-axis accelerometer signal has been reconstructed from various IMF levels to evaluate the accuracy of the filtering technique across the different IMF levels for the footfall detection task. Subsequently, the Z-axis accelerometer signal and its corresponding timestamps (in seconds) were ranged for walking times *only* using outputs of the Movement Identification Algorithm (Section 3.4.1).

Peak detection is then performed on the filtered Z-axis accelerometer signal, where peaks depict predicted footfalls. This is done owing to the logic of Section 2.4 which describes how gait is a cyclic function and the fact that footfalls can be found using the peaks of the filtered accelerometer signal [50].

(The X-axis and Y-axis signals of the accelerometer are filtered in the same manner as the Z-axis accelerometer signal using the EMD technique.)

3.5 Feature Extraction

For this research task, feature extraction is the process of transforming the pre-processed signals (Section 3.4) into a set of features that could highlight distinctive properties in the dataset and provide insight into the data. The performed feature extraction methodology is presented in the succeeding subsections.

3.5.1 Pre-and-Post Feature Extraction

Once a subject's signals (encoder and accelerometer) have been pre-processed such that spatiotemporal gait parameters in the form of straight-line walking, and turning segments are separated, and footsteps have been identified, 178 features are derived from both the 3m and 10m TUG of the 3 subject groups (treatment patients, non-treatment patients, and controls) pre-and-post their respective treatments.

These features include variants of spatial and temporal features (such as step length, step velocity, step acceleration, cadence etc.), frequency related features (such as max frequency, mean frequency, median frequencies, bandwidth etc.) and sensor features (such as accelerometer signal correlations and accelerometer signal covariance etc.) found in literature as defined in Section 2.6.2 as well as statistical measures (such as mean absolute difference, relative mean absolute difference and accelerometer signal correlation etc.) that have been adapted for gait analysis research task. These features are explored, and their usability is assessed. The list of all features extracted is presented in Section 4.4.

3.5.2 Change between Pre-and-Post Extracted Features

Once each subjects' pre-and-post features have been extracted for all TUG data, the differences between every feature for both the 3m and 10m TUG data is calculated for each subject as defined in Equation 8.

Equation 8: Change in features definition

$$\Delta Features = Features_{Pre} - Features_{Post}$$

This change is then normalised across all subjects using Equation 9 for every extracted feature. This normalised feature-set is used for all the following experiments (Section 3.7.5)

Equation 9: Normalisation formula [57]

$$Normalised\ X_i = \sum_{i=1}^{51} \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

where,

- X_i is a single extracted feature value,
- X_{max} and X_{min} are the maximum and minimum values of the extracted feature across all subjects

3.6 Feature Selection

As discussed in Section 2.6.3, feature selection methods aim to reduce data redundancy in the extracted features. Thus, the purpose of a feature selection method is to select attributes from the feature-set that are deemed most relevant for the classification problem.

For this research task, it is seen that the feature extraction stage (Section 3.5) generated a large feature-set of 178 features. This high-dimensional feature-set contains redundant and extraneous information which can negatively impact the performance of the machine learning algorithm as well as create additional data over-fitting issues. Owing to this and the fact that this research aims to determine which features can best discriminate PD patients gait changes due to treatment compared to non-treatment subjects and controls, five feature subsets have been chosen to test the derived features discriminating powers using the applied machine learning algorithm, namely:

1. Using all 178 features that were derived during feature extraction,
2. Using PCA,
3. Using the features that were selected by the 1-way Analysis of Variance (ANOVA) statistical method,
4. Using the traditional spatial and temporal features that were observable to the healthcare professional, and
5. Using the remaining Frequency, Sensor and Statistical (FSS) features that were *not* deemed clinically relevant by a healthcare professional.

These feature subset tests have been selected as they manipulate the derived features subspace in terms of maximising the variability in the feature subset (PCA features) and incorporate

only statistically relevant features (p-value < 0.05) in the feature subset (ANOVA features). Additionally, these feature subsets assess the spatial and temporal features selected by a healthcare professional (Medically Observable features) as well as assess the frequency, sensor and statistical features *not* selected by a healthcare professional (FSS features). Consequently, the discriminating power of features can be evaluated using the applied machine learning algorithm and key features can be identified, as outlined in Section 3.7 and Section 3.8 respectively.

3.7 Discrimination System

The primary aim of this research task aims to determine the extent in which quantitative differentiation between the subject groups (treatment patients, non-treatment patients, and controls) as well as differentiation between PD disease severities can be achieved. This task is deemed a classification type problem. Therefore, a supervised machine learning algorithm is used to answer the research question. The proceeding subsections explain the dataset size considerations employed, the classification algorithm implemented, and the classification experiments performed.

3.7.1 Dataset Size Considerations

Typically, a dataset is larger than the number of derived features. However, for this research like many similar clinical experiments, this is not possible as the number of subjects is relatively small (51 in the case of this study). In order to diminish this issue and prevent overfitting of the applied machine learning algorithm, two methods have been applied:

1. *10-fold Cross-validation* - whereby the dataset is partitioned into 10 random subsets consisting of 40 subjects that are used to train the machine learning algorithm and the remaining 11 subjects are used to test the classifier.
2. *Bootstrap Aggregating* – whereby the machine learning algorithm partitions the training set into 50 subsets using uniform sampling with replacement, trains a classifier for each subset and averages their results (as outlined in Section 2.6.4.2.4).

The number of samples in the dataset is seen to increase 10-fold from the cross-validation technique and 50 times from the bagging ensemble learning method. Thus, the dataset increases from a maximum of 51 samples to 25 500 samples.

3.7.2 Classification Algorithm

As this task has been deemed a classification problem, supervised machine learning algorithms have been investigated. A Bagged Decision Tree (BDT) Algorithm was used to classify the data. This type of machine learning algorithm uses bootstrap aggregating and multiple decision trees to reduce data variance and overcome decision tree overfitting issues [68]. The BDT algorithm has been selected owing to the success of the classifier as discussed in Section 2.6.4.2.4.

3.7.3 Software Implementation

The BDT classifier was implemented using the Matlab Statistics and Machine Learning Toolbox. In Matlab, 10-fold cross-validation is applied to convert the input dataset into 10 randomly selected sets of which 80% of each set is training data ($D = \{D_1, \dots, D_{10}\}$) and 20% of each set is test data ($T = \{T_1, \dots, T_{10}\}$). The BDT classifier uses each training set (D) as an input and then takes 50 bootstrap resamples of D using uniform sampling with replacement ($B = \{B_1, \dots, B_{30}\}$). Each bootstrap sample (B) is used to generate a decision tree. The results of each decision tree are then equally weighted and averaged. Test data (T) is used to assess the classification rate of the trained BDT classifiers and the classification rates across the 10 cross-validation tests are averaged to output a final BDT classification rating.

3.7.4 Performance Metrics

In order to assess the quality of the implemented BDT classifier, three performance metrics are used. These include the accuracy, sensitivity, and specificity of the classifier output as defined by Equation 10 and Equation 11 and Equation 12 respectively [74].

However, the employed classification experiments (Section 3.7.5) incorporate both 2 and 3 class classification. Therefore, a one versus all approach is used for all calculations where the True Positive (TP), True Negative (TN) False Positive (FP), and False Negative (FN) values are determined for each class against the remaining classes and then averaged.

Equation 10: Accuracy formula [74]

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

Equation 11: Sensitivity formula [74]

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

Equation 12: Specificity formula [74]

$$Specificity = \frac{TN}{TN + FP} \times 100\%$$

where,

- TP is the number of correct class predictions,
- TN is the number of correct non-class predictions,
- FP is the number of incorrect class predictions, and
- FN is the number of incorrect non-class predictions.

3.7.5 Classification Experiments

The following experiments were performed to answer part 1 and 2 of the research question. Details pertaining to these experiments are depicted in Table 3.1, highlighting each experiments purpose, the data used for the experiment, and the applied performance evaluation metrics.

Table 3.1: List of the performed experiments describing their purpose, the data used for each experiment and the respective performance evaluation metrics

Number	Experiment Purpose	Data Used	Evaluation Metrics
1	To determine the optimal number of trees that the BDT algorithm needs that allow for effective classification across the feature selection methods.	Number of Subjects= 51 Data = 3m TUG Data 10m TUG Data	Metric = Tested across all the implemented feature selection methods
2	To assess the system's ability in differentiating between the 3 subject groups (treatment patients, non-treatment patients, and controls).	Number of Subjects = 51 Data = 3m TUG Data, 10m TUG Data	Metric = Confusion Matrix, Accuracy, Sensitivity, Specificity
3	To assess the system's ability in differentiating between the PD patients who were administered medication versus the PD patients who were *not* administered medication.	Number of Subjects = 40 (PD Patients only) Data = 3m TUG Data, 10m TUG Data	Metric = Confusion Matrix, Accuracy, Sensitivity, Specificity
4	To assess the system's ability in differentiating between treatment patients who were assigned H&Y clinical scores of 2 and 3 respectively, by a healthcare professional.	Number of Subjects = 34 (Treatment Group, H&Y 2 and H&Y 3) Data = 3m TUG Data, 10m TUG Data	Metric = Confusion Matrix, Accuracy, Sensitivity, Specificity

3.8 Key Feature Identification and Validation

3.8.1 Key Feature Identification

As stated in part 3 of the research question, one of the aims of this research task is to identify the key features that can discriminate changes in PD patients walking patterns, under medication, compared to control groups. In order to identify these important features, the feature subset that yielded the best classifier performance: the PCA method was considered.

As discussed in Section 2.6.3.1.1, the PCA method produced a matrix of principal components using the covariance matrix of the original dataset and the original dataset transposed [57]. These principal components are the linear combinations of the original features that describe the variance in the dataset [57]. The produced eigenvector matrix corresponds to the correlation coefficients of the features and the principal components [75]. In other words, the eigenvector matrix correlates and weights an individual feature's contribution to a principal component.

The PCA method extracts the maximum variance within the original dataset and places it within the first principal component [57]. Moreover, an absolute correlation coefficient depicts a very high correlation above 0.9, a high correlation above 0.7, a moderate correlation above 0.5, a low correlation above 0.3, and no correlation below 0.3 [76]. Therefore, to assess the features that best contributed to the first principal component, a threshold was applied on the absolute correlation coefficient values: the 43 features that produced at least a moderate absolute correlation coefficient (greater than 0.5) were deemed the most important features of the PCA method and denoted as the "key features" from this point forward.

3.8.2 Key Feature Validation

The discriminating powers of the key features must be evaluated. Therefore, the key features are used as an input into the BDT classifier and its accuracy, sensitivity, and specificity, are compared against the results of Experiment 1.

Chapter 4. Results

4.1 Database

54 subjects participated in the clinical trial where,

- 37 subjects are PD patients who were administered a treatment,
- 6 subjects are PD patients who were *not* administered a treatment, and
- 11 subjects are healthy controls.

Most of the treatment patients' data was already collected when this research started. The control groups were collected more slowly, mainly due to the fact that many patients did not want to volunteer for an experiment where they needed to stay at the clinic for 3 hours without getting any treatment. The hope that more non-treatment patients will be recruited by the time of writing did not materialise, therefore the results, analysis, and conclusions are hereby presented, based on the dataset described above. Care was taken through the analysis to use computational methods which could minimize the effects of the small number of subjects and the imbalance between the classes, as detailed in the methods section.

4.2 Data/Signal Completeness and Integrity Check

All 54 subjects should have followed the same testing procedure, as described in Section 3.2.3 and their signals, as well as their clinical information, logged correctly and fully, as described in Section 3.2.4.

Table 4.1 depicts the results of the integrity checks in terms of acquired PAMMII signals and recorded clinical information. These results show that all clinical data has been recorded accurately and that the time vector has also been logged correctly. However, analysis has shown that both distance encoders (left and right) did *not* record data for both the 3m and 10m TUG test of 1 male subject before he was administered medication and that the accelerometer

(X, Y, and Z) did *not* record data for both the 3m and 10m TUG test of 2 female subjects after they were administered medication. These sensor malfunctions appear random as the age of the 3 subjects ranges between 61-82 and their respective H&Y clinical scores differ. Owing to this, *only* these 3 treatment patients who were administered medication have been removed from both the 3m and 10m test databases and the remaining 94% of the test database, 51 subjects in total, is deemed usable for this research task.

Table 4.1: Percentage of correctly logged clinical information and sensor data

	Sensor Data [%]						**Clinical Data [%] (as applicable)**		
	Distance Encoder		**Accelerometer**						
Test	***Left***	***Right***	***X***	***Y***	***Z***	***Time***	***H&Y***	***Age***	***Gender***
3m Pre	98	98	100	100	100	100	100	100	100
3m Post	100	100	96	96	96	100	100	100	100
10m Pre	98	98	100	100	100	100	100	100	100
10m Post	100	100	96	96	96	100	100	100	100
Database Accuracy	98	98	96	96	96	100	100	100	100

Additionally, Table 4.2 and
Table 4.3 depict the subject demographics and the distribution of H&Y clinical scores in terms of treatment patients.

Table 4.2: Subject demographics

Demographics	**Treatment Patients**	**Non-treatment Patients**	**Controls**
Number	n=34	n=6	n=11
Age (Years)	61±5	62±3	59±7
Male/Female	22 M/12 F	4 M/2 F	4 M/7 F

Table 4.3: Distribution of the treatment patients within the dataset in terms of their H&Y clinical scores

H&Y Clinical Score	**Treatment Patients**
2	n=20
3	n=14

4.3 Signal Pre-processing

4.3.1 Movement Identification Algorithm

4.3.1.1 Signal Filtering

The results of the applied Zero-Phase Lag filter are shown in Figure 4.1 and Figure 4.2. As can be seen, the respective shapes of each distance and velocity signal have been preserved and noise has been reduced.

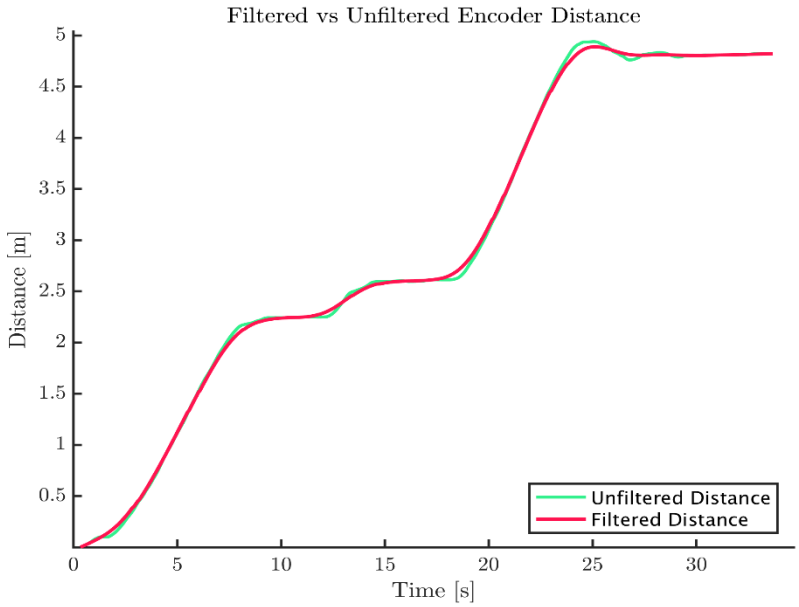

Figure 4.1: Example of a filtered vs. unfiltered distance encoder signal

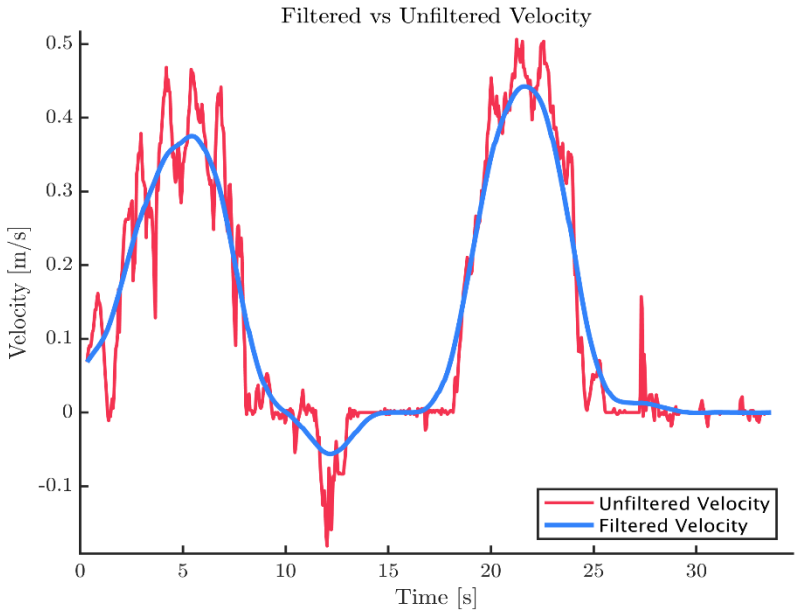

Figure 4.2: Example of a filtered vs. unfiltered derived velocity signal

4.3.1.2 Movement Stage Selection Output

The output of the Movement Identification algorithm identifies when a subject is walking in a straight-line or turning and discards stationary periods at the beginning and end of the TUG test. Figure 4.3 illustrates this highlighting the movement and turning periods by red and blue respectively.

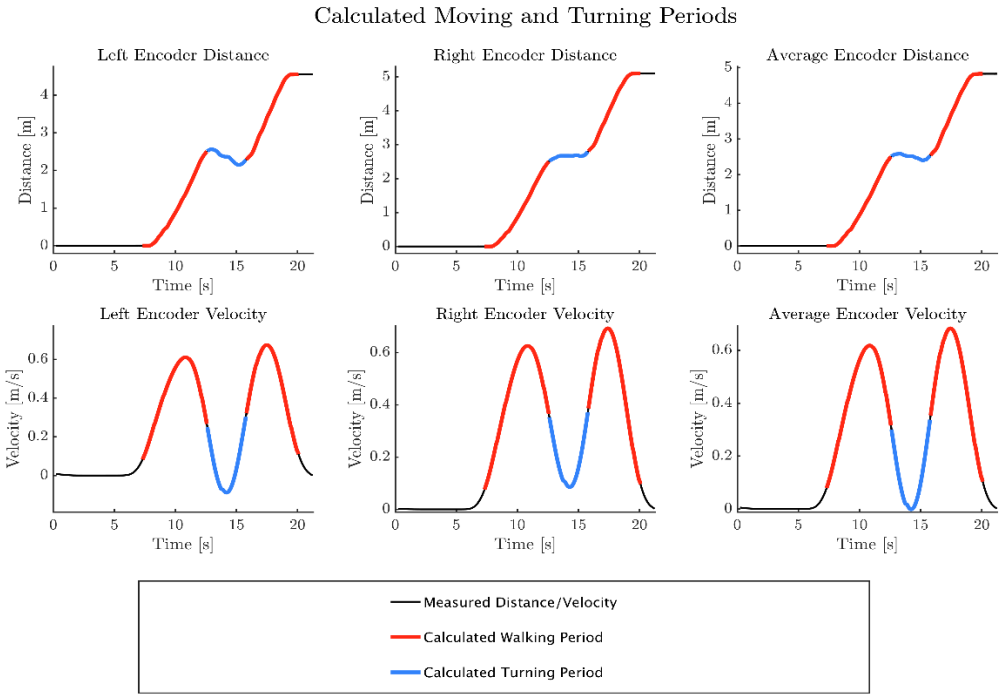

Figure 4.3: Movement Stage Identification Algorithm output results, highlighting the segmentation between straight line walking and turning periods

4.3.2 Footfall Detection Algorithm

4.3.2.1 Signal Filtering

The accelerometer signal is filtered using Empirical Mode Decomposition. The signal is reconstructed from IMF level 3 owing to logic discussed in Section 2.6.1.1 and Section 4.3.2.2. The results of the EMD filtering is shown in Figure 4.4 showing how the shape of the signal has been preserved and noise has been reduced.

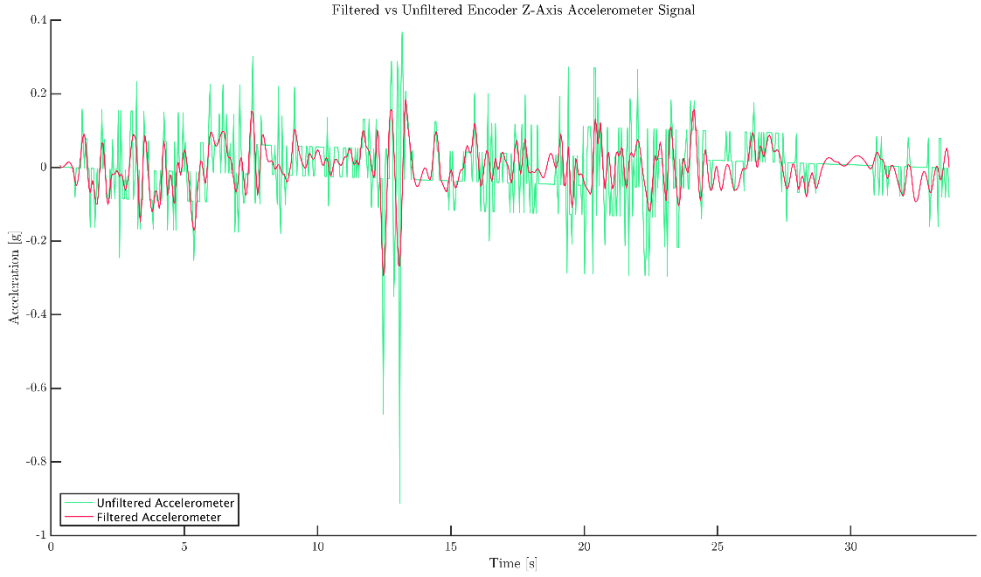

Figure 4.4: Example of an EMD filtered vs. unfiltered Z-axis accelerometer signal

4.3.2.2 Footfall Detection Output and Accuracy Evaluation

The Footfalls Detection Algorithm's accuracy has been validated against video evidence of four subjects walking patterns which simulate normal walking, PD shaking, foot-dragging and irregular walking in order to ensure that the derived footfalls detection algorithm has a satisfactory accuracy. Moreover, the Z-axis accelerometer signal has been reconstructed from a range of IMF levels to ensure that EMD filtering sensitivity is optimal. Figure 4.5 illustrates how walking footfalls have been validated where a footfall is regarded correctly predicted if it falls within the video verified range.

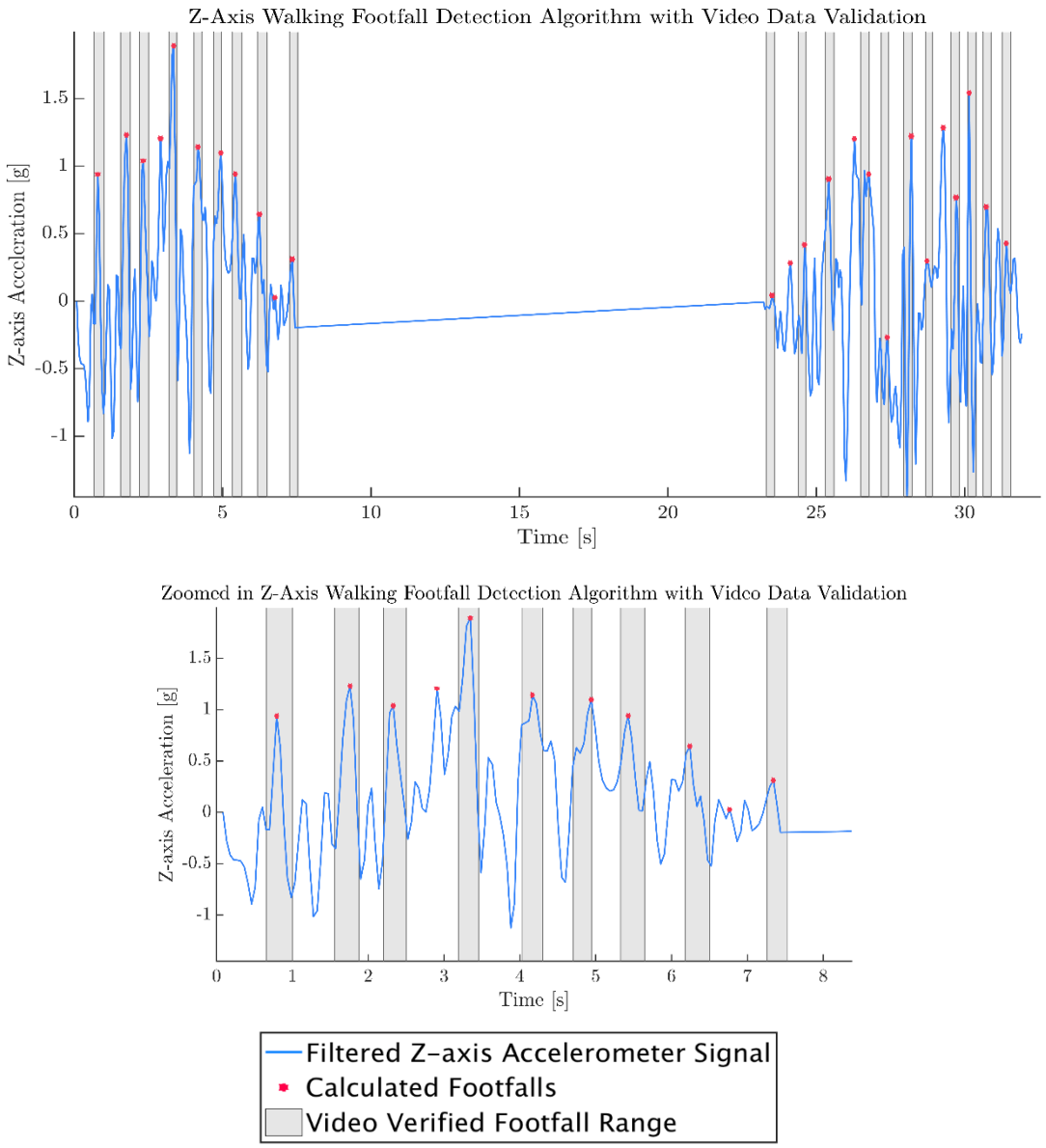

Figure 4.5: Example of Straight-line walking footfalls detection validation, for the 4 tests performed, for IMF levels 1-4

Figure 4.6 depicts the number of video-verified footsteps and algorithm predicted footsteps and compares it against correctly predicted footsteps across different IMF levels (1-4). Additionally, Figure 4.7 portrays the average accuracy of each IMF level across all four tests. Filtering the signal from IMF level 3 is shown to have the highest accuracy of 86%. Thus, it is reasoned that IMF level 1 and IMF level 2 contain the signal's high-frequency noise and IMF level 4 and higher levels contain key characteristics pertaining to the signal. This corroborates findings in [39, 45].

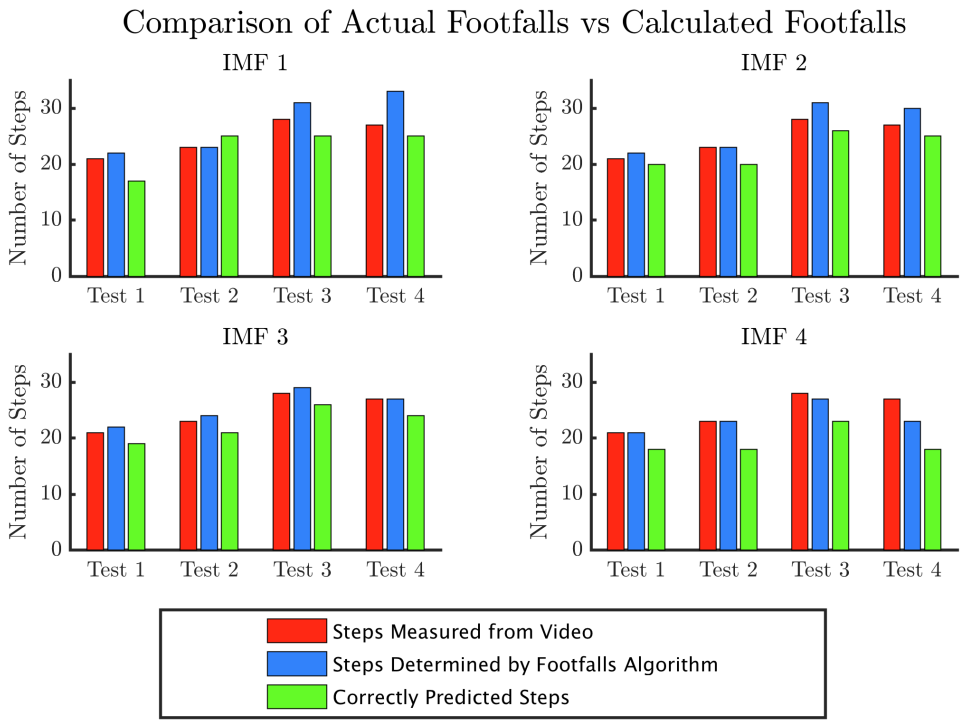

Figure 4.6: Number of correctly predicted footfalls, for the 4 tests performed, for IMF levels 1-4

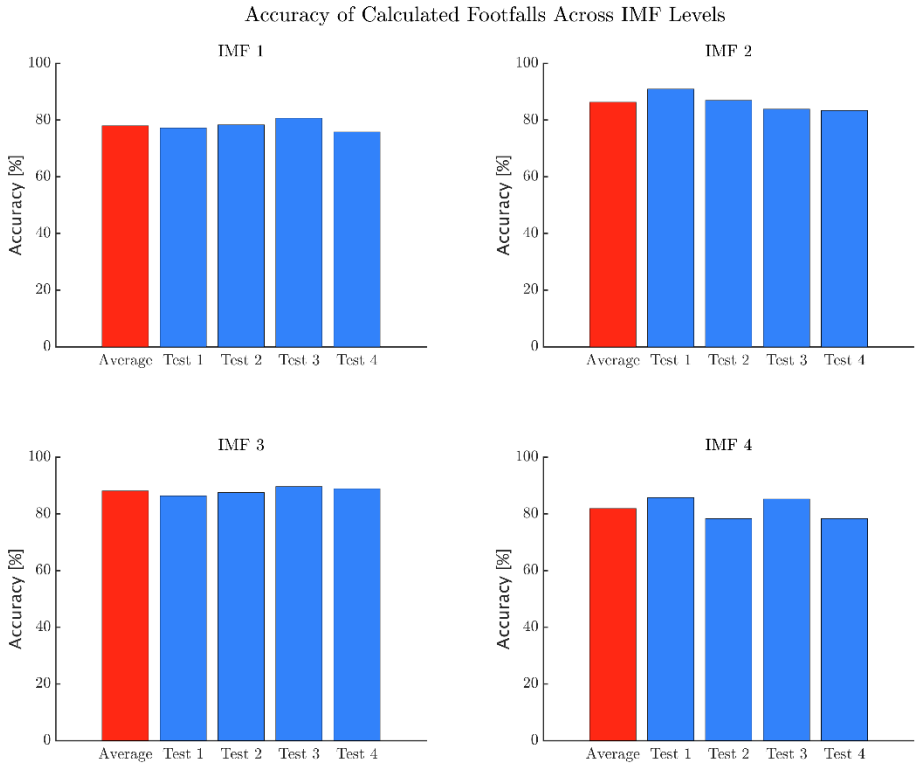

Figure 4.7: Footfall Detection accuracy, for the 4 tests performed and an average accuracy, for IMF levels 1-4

4.4 Feature Extraction

Table 4.4 and Table 4.5 depict the lists of all 178 features in terms of straight-line walking features and turning features that have been extracted.

Table 4.4: List of the features extracted from straight Line walking periods of subjects' TUG data.

	Feature	Feature Description
Time/Step	Total TUG Test Time	Total time taken to complete the TUG test in s
	Total Walking Time	Time duration spent walking during the TUG test in s
	Mean Step Time	Average time taken between steps in s
	SD Step Time	Standard Deviation of time taken between steps in s
	Variance Step Time	Variance of time taken between steps in s
	MAD Step Time	Mean Absolute Difference of time taken between steps in s
	RMAD Step Time	Relative Mean Absolute Difference of time taken between steps in s
	Number of Steps	Total number of steps taken during the walking stages of the TUG test
	Cadence	Ratio of the number of steps taken per minute
Distance	Mean Step Length	Average distance travelled between steps in m
	SD Step Length	Standard Deviation of distance travelled between steps in m
	Variance Step Length	Variance of distance travelled between steps in m
	MAD Step Length	Mean Absolute Difference of distance travelled between steps in m
	RMAD Step Length	Relative Mean Absolute Difference of distance travelled between steps in m
	Encoder Difference	Difference between distance covered by each encoder in m
Velocity	Mean Step Velocity	Average velocity between steps in m/s
	SD Step Velocity	Standard Deviation of velocity between steps in m/s
	Variance Step Velocity	Variance of velocity between steps in m/s
	MAD Step Velocity	Mean Absolute Difference of velocity between steps in m/s
	RMAD Step Velocity	Relative Mean Absolute Difference of velocity between steps in m/s
	Mean Walking Velocity	Average velocity travelled during the walking stages of the TUG Test in m/s
Acceleration	Mean Step Acceleration	Average acceleration between steps in m/s^2
	SD Step Acceleration	Standard Deviation of acceleration between steps in m/s^2
	Variance Step Acceleration	Variance of acceleration between steps in m/s^2
	MAD Step Acceleration	Mean Absolute Difference of acceleration between steps in m/s^2
	RMAD Step Acceleration	Relative Mean Absolute Difference of acceleration between steps in m/s^2

Sensor	Mean Walking Accelerometer Signal	Average of the Accelerometer Signal during walking stages of the TUG in g
	SD Walking Accelerometer Signal	Standard Deviation of the Accelerometer Signal during the walking stages of the TUG in g
	Variance Walking Accelerometer Signal	Variance of the Accelerometer Signal during the walking stages of the TUG in g
	MAD Walking Accelerometer Signal	Mean Absolute Difference of the Accelerometer Signal the during walking stages of the TUG in g
	RMAD Walking Accelerometer Signal	Relative Mean Absolute Difference of the Accelerometer Signal the during walking stages of the TUG in g
	Skewness Walking Accelerometer Signal	Degree of asymmetry in the accelerometer signal distribution
	Kurtosis Walking Accelerometer Signal	Extent that the accelerometer signal distribution lies to the left of the mean signal amplitude
	X, Y Walking Accelerometer Signal Correlation	Measure of mutual relationship between X and Y accelerometer signals
	X, Z Walking Accelerometer Signal Correlation	Measure of mutual relationship between X and Z accelerometer signals
	Y, Z Walking Accelerometer Signal Correlation	Measure of mutual relationship between Y and Z accelerometer signals
	X, Y Walking Accelerometer Signal Covariance	Measure of directional relationship between X and Y accelerometer signals
	X, Z Walking Accelerometer Signal Covariance	Measure of directional relationship between X and Z accelerometer signals
	Y, Z Walking Accelerometer Signal Covariance	Measure of directional relationship between Y and Z accelerometer signals
	Walking Power	Measure of physical power exerted by subject while walking. Measured as the integral of the accelerometer signal
	Step Regularity	Measure of regularity corresponding to peak values and lags of the auto-correlated accelerometer signal
	Step Asymmetry	Measure of difference between two successive steps using an auto-correlated accelerometer signal
	Accelerometer Signal Maximum Frequency	Maximum frequency of the accelerometer signal in Hz
	Accelerometer Signal Mean Frequency	Mean frequency of the accelerometer signal in Hz
	Accelerometer Signal Median Frequency	Median frequency of the accelerometer signal in Hz
	Accelerometer Signal Bandwidth	The difference between the largest and smallest frequencies in the accelerometer signal in Hz
	Accelerometer Signal Average Power	Mean of the total Power Spectral Density estimate of the accelerometer signal W/Hz

Table 4.5: List of the features extracted from turning periods of subjects' TUG data

	Feature	Feature Description
Time	Total Turning Time	Time duration spent turning during the TUG test in s
Velocity	Turning Velocity	Average velocity travelled while a subject is turning in m/s
	SD Turning Velocity	Standard Deviation of velocity travelled while a subject is turning in m/s
	Variance Turning Velocity	Variance of velocity travelled while a subject is turning in m/s
	MAD Turning Velocity	Mean Absolute Difference of velocity travelled while a subject is turning in m/s
	RMAD Turning Velocity	Relative Mean Absolute Difference of velocity travelled while a subject is turning in m/s
Acceleration	Turning Acceleration	Average acceleration travelled while a subject is turning in m/s^2
	SD Turning Acceleration	Standard Deviation of acceleration travelled while a subject is turning in m/s^2
	Variance Turning Acceleration	Variance of acceleration travelled while a subject is turning in m/s^2
	MAD Turning Acceleration	Mean Absolute Difference of acceleration travelled while a subject is turning in m/s^2
	RMAD Turning Acceleration	Relative Mean Absolute Difference of acceleration travelled while a subject is turning in m/s^2
Sensor	Mean Turning Accelerometer Signal	Average of the Accelerometer Signal during the turning stage of the TUG test in g
	SD Turning Accelerometer Signal	Standard Deviation of the Accelerometer Signal during the turning stage of the TUG test in g
	Variance Turning Accelerometer Signal	Variance of the Accelerometer Signal during the turning stage of the TUG test in g
	MAD Turning Accelerometer Signal	Mean Absolute Difference of /the Accelerometer Signal during the turning stage of the TUG test in g
	RMAD Turning Accelerometer Signal	Relative Mean Absolute Difference of the Accelerometer Signal during the turning stage of the TUG test in g
	Skewness Turning Accelerometer Signal	Degree of asymmetry in the accelerometer signal distribution
	Kurtosis Turning Accelerometer Signal	Extent that the accelerometer signal distribution lies to the left of the mean signal amplitude
	X, Y Turning Accelerometer Signal Correlation	Measure of mutual relationship between X and Y accelerometer signals
	X, Z Turning Accelerometer Signal Correlation	Measure of mutual relationship between X and Z accelerometer signals
	Y, Z Turning Accelerometer Signal Correlation	Measure of mutual relationship between Y and Z accelerometer signals
	X, Y Turning Accelerometer Signal Covariance	Measure of directional relationship between X and Y accelerometer signals
	X, Z Turning Accelerometer Signal Covariance	Measure of directional relationship between X and Z accelerometer signals
	Y, Z Turning Accelerometer Signal Covariance	Measure of directional relationship between Y and Z accelerometer signals
	Turning Power	Measure of physical power exerted by subject while turning. Measured as the integral of the accelerometer signal

	Accelerometer Signal Maximum Frequency	Maximum frequency of the accelerometer signal in Hz
	Accelerometer Signal Mean Frequency	Mean frequency of the accelerometer signal in Hz
	Accelerometer Signal Median Frequency	Median frequency of the accelerometer signal in Hz
	Accelerometer Signal Bandwidth	The difference between the largest and smallest frequencies in the accelerometer signal in Hz
	Accelerometer Signal Average Power	Mean of the total Power Spectral Density estimate of the accelerometer signal in W/Hz

4.5 Feature Selection

4.5.1 PCA Selected Features

As discussed in Section 2.6.3.1, PCA reduces data dimensionality by maximising data variance through an orthogonal transformation of the input features. Features were compressed using PCA into 5 principal components for the 3m and 10m TUG data respectively.

4.5.2 1-way ANOVA Selected Features

As discussed in Section 2.6.3.2.2, statistical measures are used to assess the statistical significance (p-value) of the derived features as an objective measurement of feature importance.

For this feature selection method, the 1-way ANOVA statistical test was performed on every feature across all subjects for both the 3m and 10m TUG data to determine which features are statistically significant with a p-value < 0.05. Consequently, the null hypothesis, which stated that the means of each of the groups are equal was rejected and the alternative hypothesis was accepted with a 95% certainty of a statistical difference between the means of the selected features.

Of the 178 derived features, the ANOVA feature selection method determined 43% and 41% of the derived features for the 3m and 10m TUG data respectively, to be statistically significant. Table 4.6 summarises the features that were deemed statistically significant by both sets of TUG data.

Table 4.6: Summary of the statistically significant features selected by the 1-way ANOVA statistical test for the 3m and 10m TUG data

ANOVA Selected Features		
Selected Features	***3m TUG Data***	***10m TUG Data***
Total TUG Test Time	✓	✓
Total Turning Time	✓	✓
Total Walking Time	✓	✓
Total Number of Steps	✓	✓
Mean Step Length	✓	✓
SD Step Length	✓	✓
Variance Step Length	✓	✓
MAD Step Length	✓	✓
RMAD Step Length	✓	×
Encoder Difference	✓	✓
SD Step Velocity	✓	✓
Variance Step Velocity	✓	✓
RMAD Step Velocity	✓	×
Mean Walking Velocity	✓	✓
SD Turning Velocity	✓	✓
MAD Turning Velocity	✓	✓
Variance Step Acceleration	✓	✓
RMAD Step Acceleration	✓	✓
Mean Turning Acceleration	✓	✓
SD Turning Acceleration	✓	✓
Variance Turning Acceleration	✓	✓
MAD Turning Acceleration	✓	✓
Mean Walking Z-axis Accelerometer Signal	✓	✓
SD Walking Z-axis Accelerometer Signal	✓	✓
Variance Walking Z-axis Accelerometer Signal	✓	✓
MAD Walking X, Y, Z-axis Accelerometer Signal	✓	✓
Skewness Walking X, Z-axis Accelerometer Signal	✓	✓
Kurtosis Walking Y, Z-axis Accelerometer Signal	✓	✓
Accelerometer Signal X, Z–axis Max Frequency	×	✓
Accelerometer Signal X, Z–axis Mean Frequency	✓	×
Accelerometer Signal X, Y–axis Median Frequency	✓	✓
Mean Turning Y, Z-axis Accelerometer Signal	✓	✓
SD Turning X-axis Accelerometer Signal	✓	✓
Variance Turning Z-axis Accelerometer Signal	✓	✓
MAD Turning Y, Z-axis Accelerometer Signal	✓	✓
RMAD Turning X, Z-axis Accelerometer Signal	✓	✓
X, Y Turning Accelerometer Signal Correlation	✓	✓
Skewness Turning X, Y-axis Accelerometer Signal	✓	✓
Kurtosis Turning Y-axis Accelerometer Signal	✓	✓
Turning Power	✓	×
Accelerometer Signal Y-axis Median Frequency	✓	✓
Accelerometer Signal X, Y, Z-axis Bandwidth	✓	✓
Accelerometer Signal X, Y, Z-axis Average Power	✓	✓

4.5.3 Medically Observable Features

As discussed in Section 2.6.3.2.1, healthcare professionals were consulted in order to select features that are seen as clinically observable. The term “clinically observable” was chosen

for these features, since these are characteristics which human healthcare examiners observe or calculate (i.e. using a stop-watch) in the traditional, not instrumented TUG test. Table 4.7 depicts the 12 features selected from the entire feature set that have deemed clinically relevant.

Table 4.7: List of the medically observable features selected by a healthcare professional

Medically Observable Selected Features	
Total TUG Test Time	Step Velocity
Total Walking Time	Step Velocity
Total Turning Time	Step Acceleration
Mean Step Time	Number of Steps (during the Test)
Step Time	Cadence (number of steps per min)
Step Length	
Variance Step Length	

4.5.4 FSS Features

It was seen that the healthcare professionals selected only 6.74% of the feature-set, (12 features out of the 178 features). It was reasoned that the remaining features should not be discarded to keep this study impartial and objective. Therefore, the 166 frequency, sensor and statistical features that were *not* selected by the healthcare professionals have been employed as a feature selection method in order to assess the discriminating powers of these features.

4.6 Discrimination System

As discussed in Section 3.7.5 and summarised in Table 3.1, a total of 4 experiments were conducted in order to answer part 1 and 2 of the research question. Table 4.8 summarises the results of these experiments, highlighting the aim of the experiment, the data used for the experiment and the results of the experiment.

In Experiment 2 and Experiment 3, there was a class imbalance noted, especially with the smaller number of non-treatment patients in the dataset. In order to diminish the effects of this imbalance, bootstrap aggregating has been performed on the data with 50 random subsamples taken using sampling with replacement. Moreover, additional evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier. Nevertheless, the possible effects of this imbalance have been noted and are discussed in Chapter 5.

Table 4.8: Summary of the performed experiments describing their purpose, the data used for each experiment and the results of each experiment

Number	Experiment Purpose	Data Used	Experiment Results
1	To determine the optimal number of trees that the BDT algorithm needs that allow for effective classification across the feature selection methods.	Number of Subjects= 51 Data = 3m TUG Data 10m TUG Data	Optimal Number of Trees = 50
2	To assess the system's ability in differentiating between the 3 subject groups (treatment patients, non-treatment patients, and controls).	Number of Subjects = 51 Data = 3m TUG Data, 10m TUG Data	3m TUG Data: Best Feature Selection Method = PCA Accuracy = 92.16% ± 3.92% Sensitivity = 82.35% Specificity = 93.38% 10m TUG Data: Best Feature Selection Method = PCA Accuracy = 90.20% ± 3.86% Sensitivity = 81.37% Specificity = 92.54%
3	To assess the system's ability in differentiating between the PD patients who were administered medication versus the PD patients who were *not* administered medication.	Number of Subjects = 40 (PD Patients only) Data = 3m TUG Data, 10m TUG Data	3m TUG Data: Best Feature Selection Method = PCA Accuracy = 90.00% ± 4.12% Sensitivity = 97.06% Specificity = 50.00% 10m TUG Data: Best Feature Selection Method = PCA Accuracy = 87.50% ± 4.63% Sensitivity = 94.12% Specificity = 50.00%
4	To assess the system's ability in differentiating between treatment patients who were assigned H&Y clinical scores of 2 and 3 respectively, by a healthcare professional.	Number of Subjects = 34 (Treatment Group, H&Y 2 and H&Y 3) Data = 3m TUG Data, 10m TUG Data	3m TUG Data: Best Feature Selection Method = ANOVA Accuracy = 79.41% ± 4.26% Sensitivity = 85.00% Specificity = 73.43% 10m TUG Data: Best Feature Selection Method = ANOVA Accuracy = 76.47% ± 4.57% Sensitivity = 80.00% Specificity = 73.43%

The following subsections describe the results of these experiments in detail. Table 4.8 conveys that the 3m TUG and 10m TUG results are similar for all the tests, and moreover, the 3m TUG results are slightly higher than the 10m ones. Since the 3m is also the test favoured by healthcare professionals (being the original TUG), *only* the results for 3m TUG data will be detailed. The results for the 10m TUG data are found in Appendix B.

4.6.1 Experiment 1 – Finding the Optimal Number of Trees

As with all machine learning tasks, there is a trade-off between classification accuracy and complexity or execution time [68]. In the BDT method, classification performance and execution time are proportional to the number of decision trees within the BDT classifier [68, 70]. Therefore, to make the system both accurate and efficient, an optimal number of trees to classify a class has been calculated.

All 51 subjects' TUG data was used for this experiment using 10-fold cross-validation. Figure 4.8 shows that approximately 50 decision trees are the optimal number of trees needed for bootstrap aggregating to yield the highest classification accuracy of the bagged decision tree algorithm across the five feature selection methods, for the 3m TUG data.

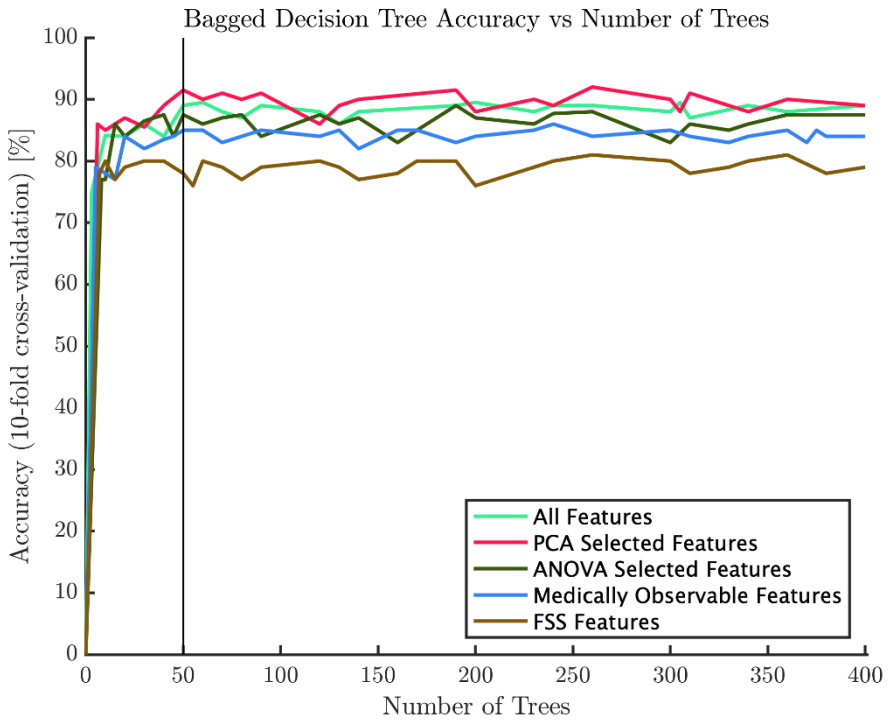

Figure 4.8: Accuracy of the BDT classifier vs. Number of Trees, highlighting the optimal number of trees needed across the implemented feature selection methods, for the 3m TUG data

4.6.2 Experiment 2 – Classification of the 3 Subject Groups

This experiment was performed to determine the discriminating power of the derived features, in differentiating between the subject groups (treatment patients, non-treatment patients, and controls).

All 51 subjects' TUG data was used for this experiment using 10-fold cross-validation and 50 decision trees. It was noted for this experiment that there was a class imbalance with a smaller number of non-treatment patients within the dataset. In order to diminish the effects of this imbalance, the BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier.

Figure 4.9 depicts the performance of the BDT classifier across the five feature selection methods, highlighting the average accuracy of the BDT classifier for each input within a 95% confidence bound.

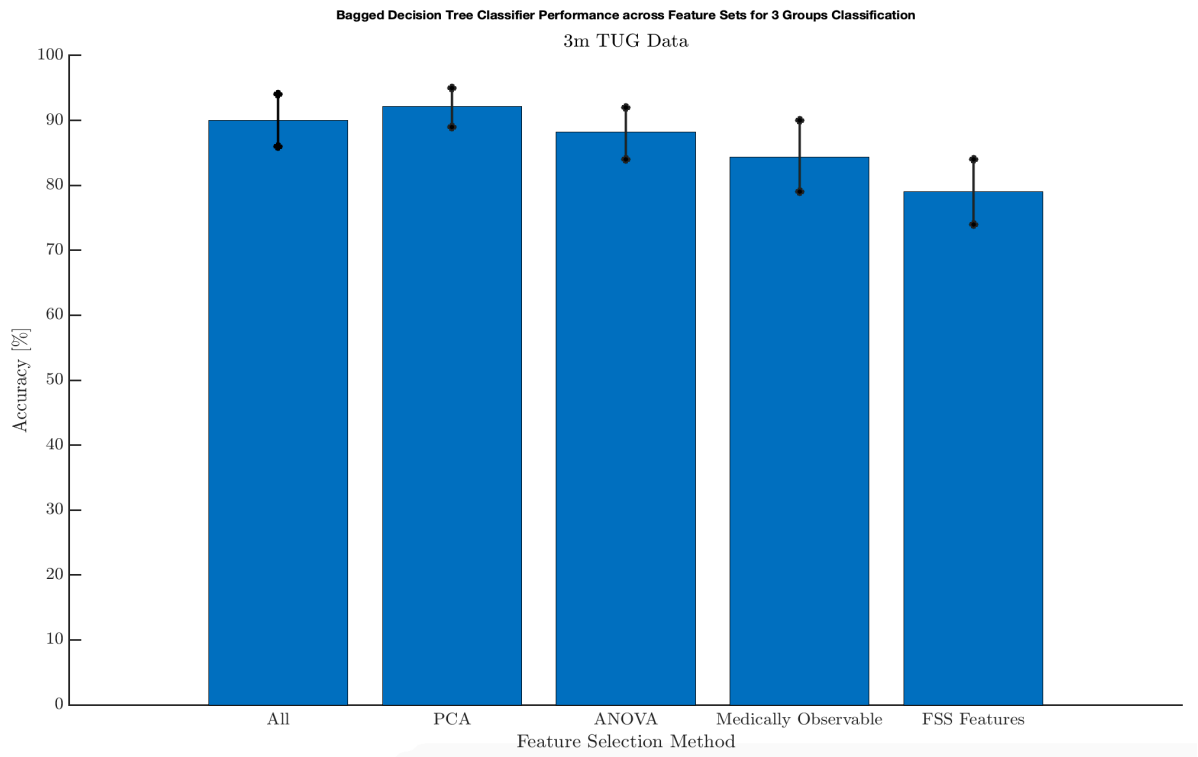

Figure 4.9: Accuracy of the BDT classifier using the five feature selection methods as an input to classify the 3 subject groups, for the 3m TUG data

As seen, using the feature-set selected by PCA as an input results in the best BDT classification performance. Figure 4.10 depicts the confusion matrix of the Bagged Decision Tree classifier using this feature-set, highlighting the nature of the classification errors, in terms of how each class is misclassified.

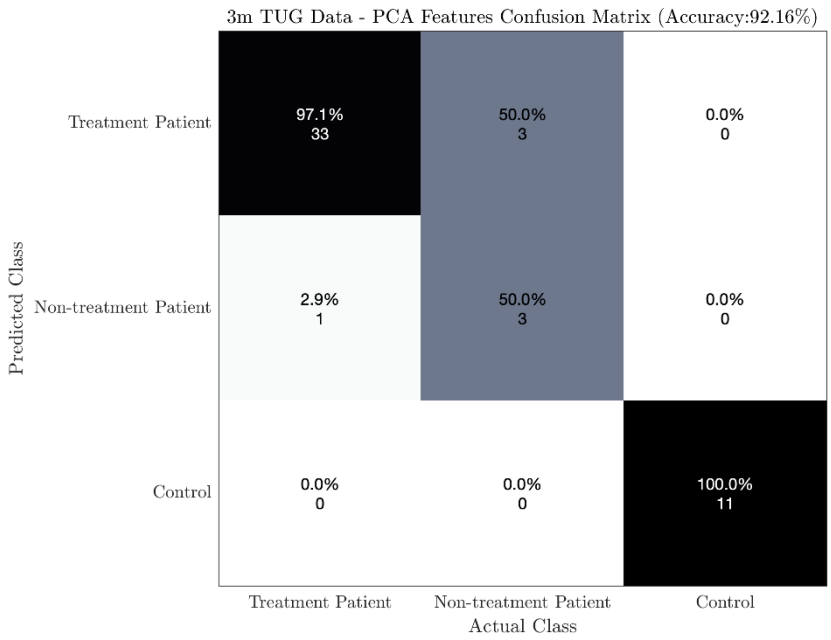

Figure 4.10: Confusion Matrix of the BDT classifier using the feature-set selected by PCA as an input for classification of the 3 Subject Groups. The actual classes (columns) are the subjects divided into their real subject groups (treatment patients, non-treatment patients, and controls) and the predicted classes (rows) are the BDT classifier output for each subject into the 3 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 3m TUG data

Table 4.9 expresses the individual sensitivity and specificity of each class (subject group) as well as the average sensitivity and specificity and accuracy of the BDT classifier using the feature-set selected by PCA as its input.

Table 4.9: Classification results for the BDT classifier using the feature-set selected by PCA as its input for classification of the 3 subject groups, for the 3m TUG data

	Class			
	Treatment Patients	**Non-treatment Patients**	**Controls**	**Average**
Sensitivity [%]	97.06	50.00	100.00	82.35
Specificity [%]	82.35	97.78	100.00	93.38

Figure 4.9 shows that using the PCA feature selection method as an input results in the best classification performance of the BDT classifier with the highest accuracy of 92.16% ± 3.92%, a sensitivity of 82.35% and a specificity of 93.38%, for the 3m TUG data. Further conclusions from this experiment are discussed in Chapter 5.

4.6.3 Experiment 3 – Classification between Treatment Patients and Non-treatment Patients

This experiment was performed to further assess the distinguishing power of the derived features, in differentiating and discriminating between the PD patients who were administered medication and the PD patients who were *not* administered medication.

Hence, all controls were removed from this experiment. The remaining 40 patients' TUG data was used for this experiment using 10-fold cross-validation and 50 decision trees. It was noted for this experiment that the class imbalance still exists, with a much smaller number of non-treatment patients compared to the treatment patients. In order to diminish the effects of this imbalance, the BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier.

Figure 4.11 shows the average accuracy of the BDT classifier using the five feature selection methods as an input to classify between treatment patients and non-treatment patients within a 95% confidence bound.

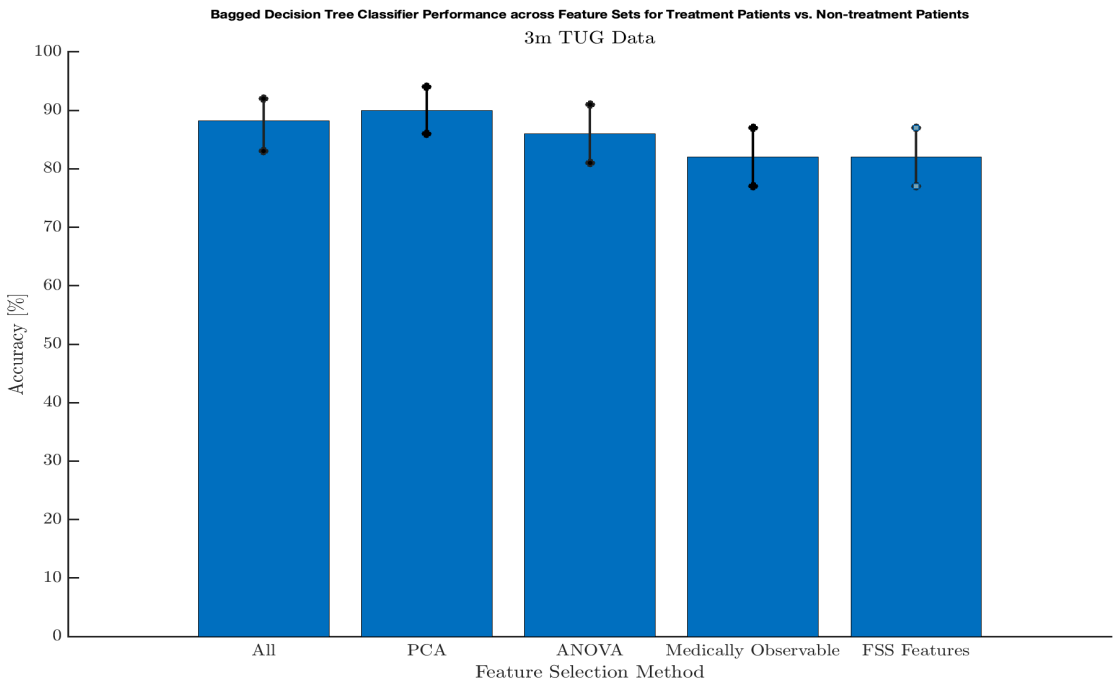

Figure 4.11: Accuracy of the BDT classifier using the five feature selection methods as an input to classify treatment patients vs. non-treatment patients, for the 3m TUG data

As seen, like Experiment 2, using the feature-set selected by PCA as an input results in the best BDT classification accuracy. Figure 4.12 depicts the confusion matrix of the BDT classifier using this feature-set as its input. This confusion matrix displays the best performance of the BDT classifier for this experiment and in doing so highlights the nature of how the two classes are misclassified.

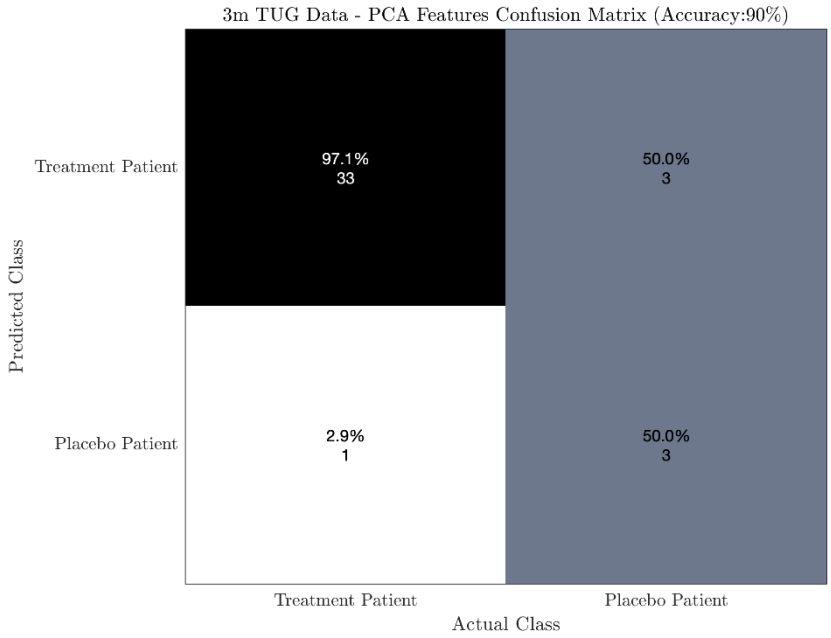

Figure 4.12: Confusion Matrix of the Bagged Decision Tree Algorithm using the feature-set selected by PCA as an input for classification of treatment patients vs. non-treatment patients. The actual classes (columns) are the subjects divided into treatment patients and non-treatment patients and the predicted classes (rows) are the BDT classifier output of each subject into these 2 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 3m TUG data

Table 4.10 expresses the accuracy, sensitivity and specificity of the BDT classifier using the feature-set selected by PCA as it input, highlighting the detection rate and false detection rate of the treatment patients and the non-treatment patients.

Table 4.10: Classification results for the BDT classifier using the feature-set selected by PCA as its input for classification of treatment patients vs. non-treatment patients, for the 3m TUG data

Accuracy [%]	90.00 ± 4.12
Sensitivity [%]	97.06
Specificity [%]	50.00

Figure 4.11 demonstrates that using the PCA feature selection method as an input results in the best classification performance of the BDT classifier with the highest accuracy of 90.00% ± 4,12%, a sensitivity of 97.06% and a specificity of 50.00%, for the 3m TUG data. Further conclusions from this experiment are discussed in Chapter 5.

4.6.4 Experiment 4 – Classification between Treatment Patients H&Y 2 and H&Y 3 Clinical Scores

This experiment was performed to assess the distinguishing power of the derived features, in differentiating and discriminating between *only* the treatment patients with H&Y clinical scores of 2 and 3.

Hence, all non-treatment patients and controls were removed from this experiment. The remaining 34 subjects' TUG data was used for this experiment using 10-fold cross-validation and 50 decision trees. The BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier.

Figure 4.13 depicts the average accuracy of the BDT classifier using the five feature selection methods as an input to classify treatment patients in terms of their H&Y 2 and H&Y 3 clinical scores within a 95% confidence bound.

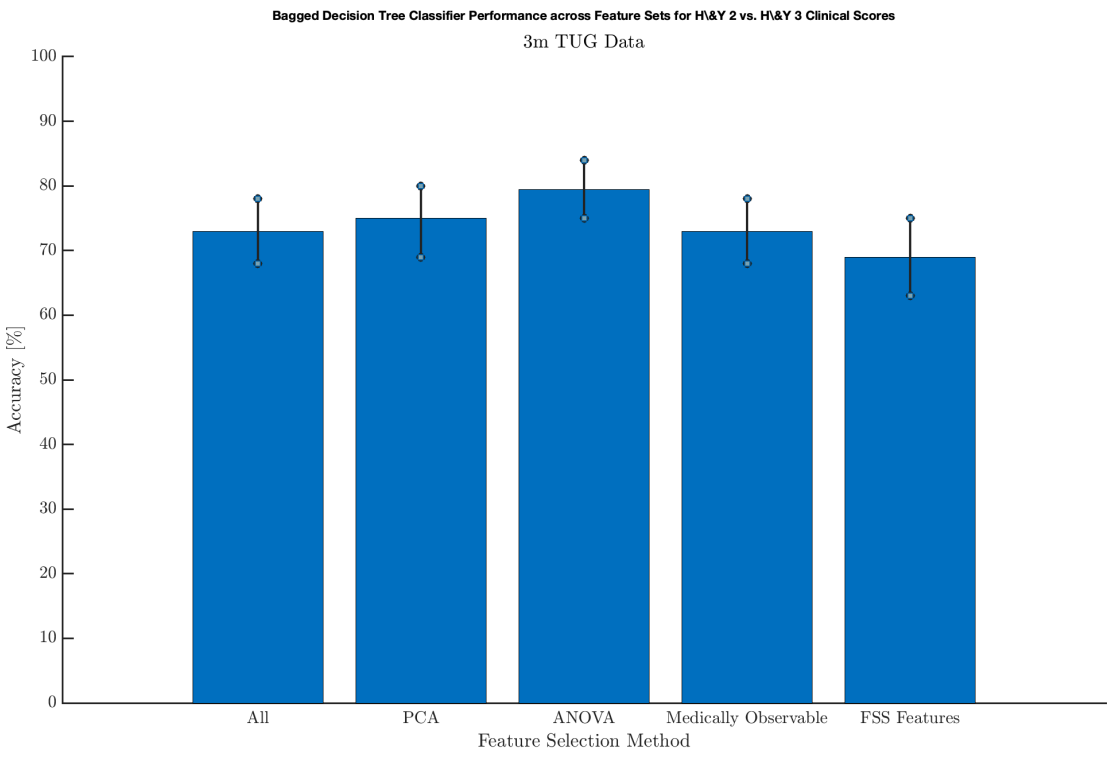

Figure 4.13: Accuracy of the BDT classifier using the five feature selection methods as an input to classify treatment patients in terms of their H&Y 2 and H&Y 3 clinical scores, for the 3m TUG data

Unlike Experiment 2 and Experiment 3, using the feature-set selected by the 1-way ANOVA method as an input results in the best BDT classification accuracy. Figure 4.14 shows the confusion matrix of the Bagged Decision Tree classifier using this feature-set, highlighting the nature of the classification errors, in terms of how each class is misclassified.

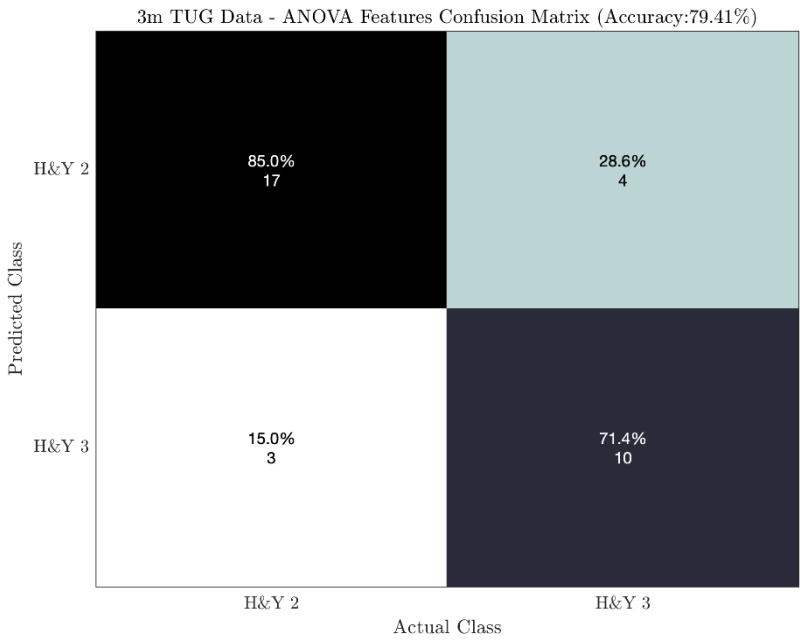

Figure 4.14: Confusion Matrix of the BDT Algorithm using the feature-set selected by ANOVA as an input for classification of the treatment patients according to their H&Y clinical scores. The actual classes (columns) are the subjects divided into their H&Y 2 and H&Y 3 clinical scores and the predicted classes (rows) are the BDT classifier output of each subject into these 2 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 3m TUG data

Table 4.11 expresses the accuracy, sensitivity and specificity of the BDT classifier using the feature-set selected by the 1-way ANOVA method as it input, highlighting the detection rate and false detection rate of the treatment patients according to their H&Y clinical scores.

Table 4.11: Classification results for the BDT classifier using the feature-set selected by ANOVA as its input to classify treatment patients in terms of their H&Y 2 and H&Y 3 clinical scores, for the 3m TUG data

Accuracy [%]	79.41 ± 3.86
Sensitivity [%]	85.00
Specificity [%]	71.43

Figure 4.13 shows that using the 1-way ANOVA feature selection method as an input results in the best classification performance of the BDT classifier with the highest accuracy of

79.41% ± 3.86%, a sensitivity of 85.00% and a specificity of 71.43%, for the 3m TUG data. Further conclusions from this experiment are discussed in Chapter 5.

4.7 Key Feature Identification

The PCA algorithm yielded the best BDT classification performance in terms of accuracy, sensitivity, and specificity for Experiment 1. In order to assess the features that best contributed to the first principal component, a threshold was applied on the absolute correlation coefficient values: the 43 features that produced at least a moderate absolute correlation coefficient (greater than 0.5) were deemed the most important features. Table 4.12 lists the 43 features found, separating straight-line walking features and turning features.

Table 4.12: List of the 43 key features that had a moderate-very high absolute correlation coefficient between the original features and the first principal component

Key Features	
Straight Line Walking Features	***Turning Features***
Total TUG Test Time	Mean Turning Acceleration
Total Walking Time	MAD Turning Acceleration
Mean Step Length	Mean Turning Z-axis Accelerometer Signal
SD Step Length	RMAD Turning Z-axis Accelerometer Signal
Variance Step Length	Skewness Turning X, Y-axis Accelerometer Signal
MAD Step Length	Kurtosis Turning Y-axis Accelerometer Signal
RMAD Step Length	Turning Power
Encoder Difference	Accelerometer Signal Y–axis Median Frequency
SD Step Velocity	Accelerometer X, Y, Z-axis Signal Bandwidth
Variance Step Velocity	
RMAD Step Velocity	
Mean Walking Velocity	
Variance Step Acceleration	
RMAD Step Acceleration	
Mean Walking Y-axis Accelerometer Signal	
SD Walking Z-axis Accelerometer Signal	
Variance Walking Z-axis Accelerometer Signal	
RMAD RMS-Z Accelerometer Signal	
RMS, Z Walking Accelerometer Signal Correlation	
RMS, Z Walking Accelerometer Signal Covariance	
Skewness Walking RMS–Z Accelerometer Signal	
Kurtosis Walking RMS-Z, Z-axis Accelerometer Signal	
Accelerometer Signal X, Z–axis Max Frequency	
Accelerometer Signal X, Z–axis Mean Frequency	
Accelerometer Signal X, Y–axis Median Frequency	
Accelerometer Signal X, Z, RMS Bandwidth	

4.8 Key Feature Validation

In order to validate that the key features of Table 4.12 are able to discriminate between the treatment patients, non-treatment patients and controls, these features are fed into the BDT

classifier and its accuracy, sensitivity, and specificity is compared with the results of Experiment 1.

The 51 subjects' 3m TUG data was used for this validation experiment using 10-fold cross-validation and 50 decision trees. As in Experiment 1, it was noted for this experiment that the class imbalance exists due to the significantly smaller number of non-treatment patients. In order to diminish the effects of this imbalance, the BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to validate the overall performance of the BDT classifier.

Figure 4.15 compares the performance of the BDT classifier using the key features subset as input against the PCA selected feature-set, highlighting the accuracy as well as the average sensitivity and the average specificity of the BDT classifier for each input.

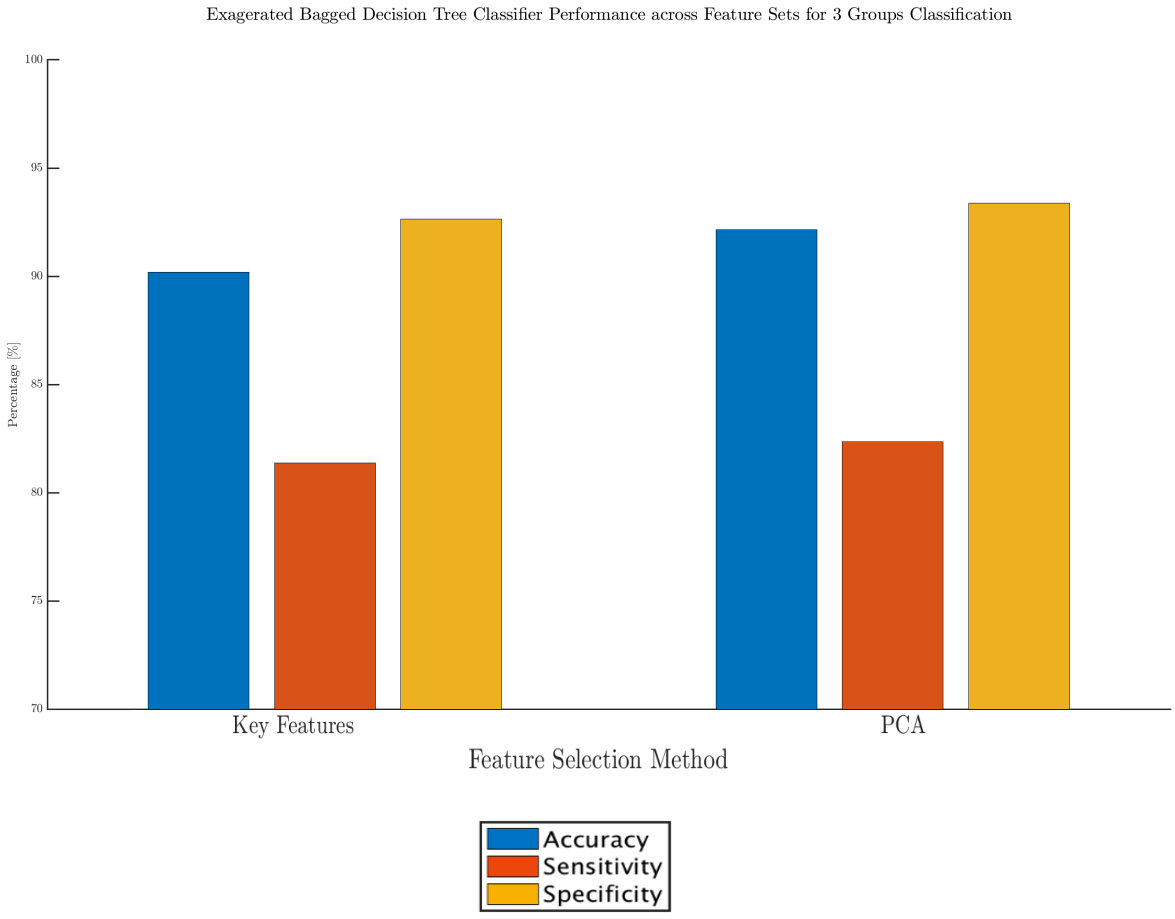

Figure 4.15: Performance comparison of the BDT classifier comparing the key features performance against the PCA selection method in terms of classification of the 3 subject groups, for the 3m TUG data

The figure illustrates that using the key features feature-set as an input yields a slightly reduced performance than the PCA method, but the proportion between the measures is similar. Figure 4.16 depicts the confusion matrices of the BDT classifier using the key features as its input for the 3m TUG data. This highlights the nature of the classification errors, in terms of how each class is misclassified.

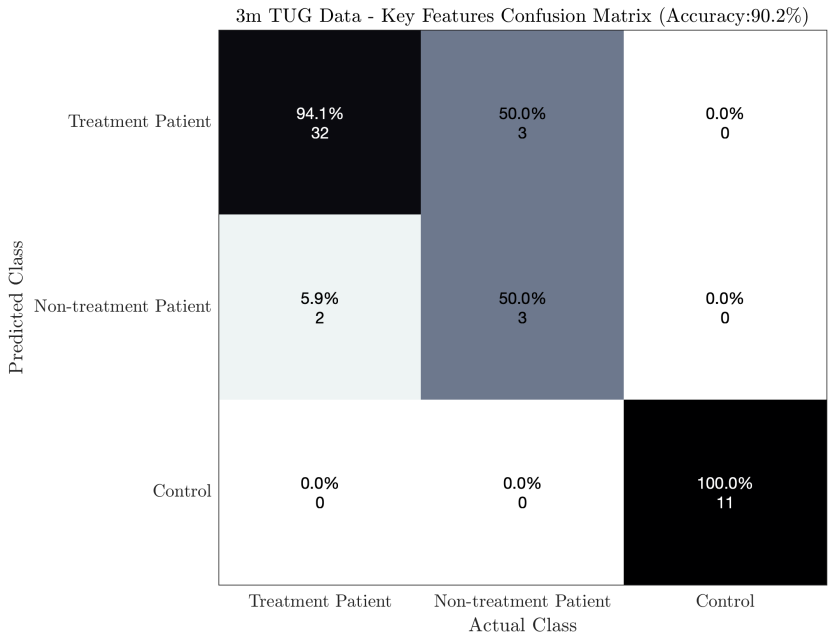

Figure 4.16: Confusion Matrix of the BDT Classifier using the key features as an input for classification of the 3 subject groups. The actual classes (columns) are the subjects divided into their real subject groups (treatment patient, non-treatment patient and control) and the predicted classes (rows) are the BDT classifier output of each subject into the 3 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 3m TUG data

Table 4.13 expresses the sensitivity and specificity of each class (subject group) as well as the average sensitivity and specificity and accuracy of the BDT classifier using the key features feature-set as its input.

Table 4.13: Classification results for the BDT classifier using the key features feature-set as its input for classification of the 3 subject groups, for the 3m TUG data

	Class			
	Treatment Patients	**Non-treatment Patients**	**Controls**	**Average**
Sensitivity [%]	94.12	50.00	100.00	81.37
Specificity [%]	82.35	95.56	100.00	92.64

Table 4.14 compares the accuracy, average sensitivity, and average specificity of the BDT classifier using the key features subset against the same performance measures of the PCA input subset.

Table 4.14: BDT classifier results, comparing the key features against the PCA feature selection method inputs for classification of the 3 subject groups, for the 3m TUG data

	3m TUG Data [%]	
	Key Features	**PCA**
Accuracy	90.20 ± 4.26	92.16 ± 3.92
Sensitivity	81.37	82.35
Specificity	92.64	93.38

In spite of the slightly reduced classification performance using the key features as compared to the PCA, the key features set is a subset of the original features and is useful for further analysis and interpretations of these features in the context of discrimination between the treatment patients, non-treatment patients, and controls. Further conclusions of this experiment are discussed in Chapter 5.

Chapter 5. Discussion and Conclusion

5.1 Key Findings

5.1.1 Research Goals

The primary goal of this research was to answer the following question "To what extent can features extracted from signals acquired by walker mounted sensors identify changes in PD patient gait due to treatment."

This question is translated into three objectives as follows:

1. To determine to what extent combinations of these features provide a classification of the three subject groups.
2. To determine to what extent these features or their combinations identify the severity of the disease according to H&Y clinical scales.
3. To determine which of these features, derived from kinematic signals from walker mounted sensors best discriminate PD patients gait changes due to treatment compared to non-treatment patients and controls.

Discussion of the results that answer this question is outlined in the succeeding subsections.

5.1.1.1 Experiment 2 – Classification of the 3 Subject Groups

The aim of this experiment was to gain insight into how the treatment patients responded to medication compared to non-treatment patients and controls. This was assessed through the BDT classifiers ability to differentiate between the three subject groups. Part 1 of the research question has been answered using a BDT classifier with 50 decision trees and was validated using 10-fold cross-validation.

For this experiment, it was noted that there was a class imbalance, especially with a smaller number of non-treatment patients compared to the other two groups. The BDT classifier was implemented using bootstrap aggregating to mitigate some of the class imbalance effects. Furthermore, additional evaluation metrics were included in this study, in the form of accuracy, sensitivity, specificity and confusion matrices, to validate the overall performance of the classification.

As outlined in Section 4.6.2, using the feature-set selected by PCA resulted in the best classification performance with an accuracy of 92.16% ± 3.92%, sensitivity of 82.35%, and specificity of 93.38% measured, for the 3m TUG data.

3-class classification of gait data in PD was not found in literature. The performance comparison is therefore made to a 2-class (binary) classification results. A previous research analysed signals from wearable sensors, attached to a person's shoe to detect severe PD patients from controls with an accuracy of 91%, a sensitivity of 100% [77]. The current research accuracy (92.16%) is consistent with this finding, while the current research sensitivity (82.35%) is far worse. Naturally, a 3-class classification differs from a 2-class (binary) classification. In the current research, 2 of the 3 classes are PD patients and one is a control group. Therefore, it seems plausible that the 2 classes of PD patient would be less distinguishable, lowering the sensitivity.

This assumption is corroborated by further inspection of the performance of the BDT classifier. The PCA selected feature-space confusion matrix in Figure 4.10 displays the best classification performance in terms of accuracy (92.16%), sensitivity (82.35%), and specificity (93.38%) as well as the type of classification errors. All controls in this experiment were successfully identified and discriminated from treatment patients and non-treatment patients. This finding implies a similar pattern to previous findings, where the PAMMII instrumented discriminated PD patients from controls with sensitivity and specificity scores of 91% and 95% respectively [16].
However, as demonstrated in Table 4.9, the decrease in these average performance measure stems from the result that 50.00% of the non-treatment patients were misclassified as treatment patients. No definite explanation could be found for this finding. Some plausible suggestions follow: A previous study suggested that placebo treatments can have a strong influence on PD

symptoms even if the patient is aware that the treatment is a placebo or for control purposes [78]. Although in the current experiment the non-treatment patients did not get treatment at all, the influence of placebo effects cannot be excluded, as the consultation with one's doctor may have caused some non-treatment patients to feel "treated" [78]. This may have caused the changes seen in some the non-treatment patients. Notably, those few non-treatment patients who agreed to participate in the study and sit for 3 hours between the tests, are a minority of the patients who asked to, which may explain their assumed "changes".

The interpretation can also consider the results of the treatment group, which show high discrimination: 33 of 34 patients in this group were correctly classified as treatment patients. Assuming that the features convey changes before and after the treatment, this result implies that these patients have gone through a change. Amantadine has been shown to improve PD patient condition over time [79]. Naturally, not all patients respond similarly: Previous clinical trials conducted on 36 patients receiving Amantadine showcased efficacious improvements in 64% of patients [79]. Another study exhibited a 60% improvement in terms of dyskinesia across 14 patients who completed the trial [80]. The much higher percentage of assumed "responders" to the treatment in the current study (97.06%) may imply that either the majority of treatment patients are indeed responding to the Amantadine-Sulphate IV medication or that the system is too sensitive, detecting small changes which means that the successful detection rate of the treatment patients is probably exaggerated. The latter argument could also explain the result that 50% of the non-treatment patients also appear to "respond" and suggest that the system may be too sensitive picking up subtle changes across both the treatment patients and non-treatment within the short 3-hour test frame. The psychological and physical factors which affected both groups of PD patients in the 3-hours they spent in the hospital, being surrounded by doctors, nurses, and other patients may have caused the changes in most PD patients regardless of the treatment given.

Another interpretation can be a methodology-related one: regardless of all the efforts taken to diminish and mitigate the effects of the class imbalance within experiment 2, there still may be imbalance effects, which may explain some of the findings. Moreover, only 6 non-treatment patients were employed in this study. This cohort must be increased to provide conclusive evidence regarding these findings. Therefore, future research requires additional subjects (treatment patients, non-treatment patients, and controls) to further validate the methodology and provide clearer insight into the group differences. Particularly, if these results are

improved for a larger more balanced cohort, the system could be used as a marker that can predict if a PD patient will respond to this treatment already in a first session. If no response is indicated, doses may need to be changed, or a different treatment started.

An additional experiment was performed, limiting the method to a 2-class discrimination in order to further assess the system's ability to distinguish between the treatment patients and non-treatment patients, excluding the controls. This is discussed in the succeeding subsection.

5.1.1.2 Experiment 3 – Classification between Treatment Patients and Non-treatment Patients

The aim of this experiment was to re-validate the findings of experiment 2, examining if the classifier's results will be different when the controls were excluded from the data and the classification was reduced to 2-classes only: PD patients who were administered medication (treatment patients) and PD patients who were *not* administered medication (non-treatment patients). The same BDT classifier with 50 decision trees was used for this experiment and validated using 10-fold cross-validation.

It was noted that the class imbalance still exists in this experiment. Therefore, the BDT classifier was implemented again using bootstrap aggregating to mitigate some of the class imbalance effects. Additionally, extra evaluation metrics were included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the classification.

The PCA feature selection method as an input to the classifier yielded again the best classification performance with an accuracy of 90.00% ± 4.12%, sensitivity of 97.06%, and specificity of 50% measured for the 3m TUG data.

To the knowledge of the author, this type of classification experiment, where sensors were used for the analysis and machine learning classification, performed for PD treatment efficacy has not been performed to date. Therefore, a comparison with previous literature could not be accomplished.

Interpreting the best classification results of this experiment, in terms of the highest accuracy, sensitivity, and specificity, the results obtained support the findings of Experiment 2. Both experiments found that the PCA selected feature-space resulted in the best classification performance. In both experiments, 97.06% of the treatment patients were identified successfully whereas only 50% of the non-treatment patients were identified successfully.

Within this experiment, a class imbalance between the 2 groups still existed. Regardless of all the care taken to diminish and mitigate the effects of the class imbalance, there still remained an imbalance. The small class of 6 non-treatment patients may also have produced a bias being too small.

The results of the 2 first experiments, therefore, indicate that more research is needed to validate these initial findings, in order to provide a likely interpretation. The methodology employed in this study should be examined by using a larger cohort of non-treatment patients. Since no objective measure of improvement exist, a patient's self-report, which is the way the treatment was evaluated in former non-instrumented studies, could also be used for validating patients' improvement.

With larger and balanced cohorts, this system may be useful to further explore differences in response in different patients' groups (looking at age, gender, other clinical measures, disease history and more), which may have insights for understanding PD and determining the suitability of Amantadine-Sulphate IV and other treatments. Additionally, all treatment patients received the same dosage Amantadine-Sulphate IV. Certain patients may require higher or lower dosages to improve their condition. Future research could be performed using this system to evaluate this postulation.

5.1.1.3 Experiment 4 – Classification between Treatment Patients H&Y 2 and H&Y 3

The aim of this experiment was to further explore which treatment patients responded to their respective treatments and to examine if this response could be predicted by their H&Y clinical scores. This is assessed through the BDT classifiers ability to differentiate between treatment patients who were assigned an H&Y score of 2 and 3 by a healthcare professional. Part 2 of the research question has been answered using a BDT classifier with 50 decision trees and was validated using 10-fold cross-validation.

As opposed to the previous experiments, where PCA feature selection yielded the best classification performance, in this experiment the 1-way ANOVA selected feature-space as an input, resulted in a marginally better classification performance with an accuracy of 79.41% ± 4.26%, a sensitivity of 85.00%, and a specificity of 73.43%, for the 3m TUG data.

To the knowledge of the author, this type of classification experiment has not been implemented to date where changes in gait before-and-after medication are calculated (as discussed in Section 3.5.2) and classified. This kind of feature extraction (looking at differential values, or changes) removes the inherent differences between individual subjects' gait. Presumably, this emphasises the differences between each subjects' walking tests before-and-after medication rather than the difference between individuals. However, by removing the differences between individuals, it may increase the likeness and resemblance between subjects in the different H&Y subject groups, especially if they responded in a comparable manner, thus reducing the classifier performance.

The 1-way ANOVA selected feature-space confusion matrix shown in Figure 4.14 highlights the best classification performance of this experiment in terms of accuracy (79.41%), sensitivity (85%%), and specificity (73.43%). This confusion matrix highlights the type of classification errors that were obtained. As seen, 15% of the H&Y 2 treatment patients have been misclassified as H&Y 3 and 29% of the H&Y 3 treatment patients have been misclassified as H&Y 2. The only similar variable between the two groups of patients was the dosage of Amantadine-Sulphate IV that was administered to all. Differences in age, gender, and physical condition may prove to affect more the response than 'these patients H&Y severity scale. Future research could incorporate patient groups who receive different dosages of the medication to assess the full effects of the treatment. A larger patient group is needed in order to examine the effects of these variables.

5.1.1.4 Feature Selection Methods

A large feature-set, consisting of 178 features was computed in this research. These features included the standard medically observable, spatiotemporal features (such as step length, step velocity, step acceleration, cadence etc.) as well as frequency related features (such as max frequency, mean frequency, median frequencies, bandwidth etc.), sensor features (such as accelerometer signal correlations and accelerometer signal covariance etc.) and statistical

measures (such as mean absolute differences, relative mean absolute differences and accelerometer signal correlations etc.) to improve the analysis of this study and assess the capabilities of multiple feature selection methods.
As discussed in Section 3.6, four feature selection methods were implemented:

- A PCA feature subset,
- An ANOVA feature subset,
- A Medically Observable (Neurologist selected) spatiotemporal feature subset, and
- An FSS feature subset that included the remaining features not-selected by the neurologist.

The results have shown that using the feature subsets selected by the PCA and the 1-way ANOVA feature selection methods as inputs into the BDT classifier produced the best classification performance in terms of accuracy, sensitivity, and specificity across the experiments. The PCA and the 1-way ANOVA feature subsets included both Medically Observable features and FSS features. This suggests that a combination of the traditional Medically Observable feature subset and the FSS feature subset should be used to better classify treatment patients, non-treatment patients, and controls as well treatment patients according to the H&Y clinical scores.

5.1.1.5 Key Features Identification and Validation

Part 3 of the research question has been answered by interpreting which features contributed the most to the PCA technique. 43 key features were found (listed in Table 4.12) and resulted in a comparable albeit slightly lower accuracy, sensitivity and specificity in the classification compared to the PCA method (illustrated in Figure 4.15). These results suggest that using the subset of the key features is a desirable trade-off, further reducing the feature space while providing the additional capability of interpreting individual features from the original set and producing only slightly lower classification performance.

Furthermore, the key features sub-space is larger than the ones used in previous studies [10, 23, 27, 13, 49]. This stems from the fact that the feature extraction stage (Section 3.5 and Section 4.4) of this research included standard medically observable, spatiotemporal features (such as step length, step velocity, step acceleration, cadence etc.) as well as adapted frequency

related features (such as max frequency, mean frequency, median frequencies, bandwidth etc.), sensor features (such as accelerometer signal correlations and accelerometer signal covariance etc.) and statistical measures (such as mean absolute differences, relative mean absolute differences and accelerometer signal correlations etc.). This approach is unique to this study. The results obtained across all experiments suggest that this approach may be advantageous for future research.

5.1.1.5.1 Interpretation of the Key Features

The key features subset included 43 features listed in Table 4.12.

Previous studies placed emphasis on extracting the spatial and temporal features of gait namely: step velocity, step time and step length with these features being derived from wearable sensors or instrumented walkers [27, 13, 16]. The key features set in this contains features from four feature categories namely: spatiotemporal features (3/43), adapted frequency related features (13/43), sensor features (11/43), and statistical measures (16/43).

Variations of step length and step velocity are found in the list of key features (Table 4.12), which corroborates past studies [27, 13, 16]. However, a significant percentage of additional features: frequency features, sensor feature, and statistical measures were included in this study's key features feature-space. This suggests that the gait analysis of PD patients' response to treatment should be expanded beyond the traditionally used medically observable, time-based features. Additional abstract and mathematical features such as frequency features, sensor feature, and statistical measures should be included in future similar research as these features may yield further insights into PD patients' response to treatment and the general understanding of PD.

5.1.2 Differences between the 3m and 10m TUG Data

The experiments of this study were performed on both the 3m and 10m TUG walking test data. As summarised in Table 4.8, using the 3m and 10m TUG data produced similar discrimination performance in terms of accuracy, sensitivity, and specificity across each experiment. This is corroborated by a previous study where the PAMMII device was used to record a 3m and 10m TUG walking tests of PD patients and controls [16]. Results showed that both TUG walking tests maintained a similar subject classification sensitivity and specificity of 91% and 95% for the 3m test and 96% and 100% for the 10m test in [16]. Thus,

it is reasoned that only *one* TUG test should be considered in future studies. Furthermore, it is preferable to use the 3m TUG test *only* as it is shorter, the original standard TUG test and results in higher classifier performance accuracy.

5.1.3 Conclusions regarding the use of the PAMMII Instrumented Walker

This research task utilised data that was pre-acquired from a trial using the PAMMII instrumented walker (as outlined in Section 3.2). This device was selected for this research owing to the success of a PAMMII study (mentioned in Section 2.5.4) as well as the fact that the PAMMII system is an affordable and easy-to-use solution that requires no professional labour to measure the gait of a PD patient, in both a clinic and home setting on a daily basis or as needed.

This device uses low-cost sensors that are mounted on a 3rd party apparatus. This resulted in a more challenging signal pre-processing task, specifically in establishing when a subject is walking and turning/stationary as well as when a subject's foot touches the ground while walking for gait extraction. The proceeding subsections will offer insight into the pre-processing required.

5.1.3.1 Accurate Movement Stage Identification

In general, a movement is composed of several stages, which are different in properties (i.e. straight line vs. circular) and/or separated by no-motion periods. The TUG walking test protocol used in the PAMMII trial showed that subjects walking patterns can be partitioned into two phases of straight-line walking that are separated by a turning phase.

This being said, these walking patterns are seen to vary across subjects. Some patients complete the protocol without a break in a timely manner while other patients stop during the test for short or long periods as needed. The implemented Movement Stage Identification algorithm differentiated between these stages successfully. This shows the capabilities of the PAMMII system's low-cost distance encoders measurements and its numerically differentiated velocities. However, this algorithm has been built for the pre-acquired 3m and 10m datasets. Consequently, it is reasoned that this algorithm should be tested using other trials and datasets to fine-tune the thresholds of this algorithm as outlined in Figure 3.3.

5.1.3.2 Accurate Footfalls Detection

A basic quantity of walking, which decomposes it into elementary sections is the footfall. As discussed throughout Section 2.3 there is a relationship between PD and gait. Moreover, gait analysis systems as detailed in Section 2.5 see the footfall as a necessity to calculate reliable gait features.

As mentioned in Section 3.4, the implemented Footfalls Detection Algorithm was performed by the author of this thesis, in collaboration, and published [73]. This previous work compared the newly proposed Z-axis accelerometer algorithm (86% accuracy) described in this research against another replicated footfall detection algorithm (69% accuracy) that utilises the difference between force sensors (placed on the instrumented walker's handles) to estimate footfalls [73]. Results from [73] highlight the superiority of the newly proposed Z-axis accelerometer algorithm which incorporates EMD filtering and peak detection to identify footfalls (Section 3.4.2).

The following list summarises the key findings regarding the EMD filtering of the original Z-axis accelerometer signals using the reconstruction of the EMD signal from IMF levels 1-4:

1. **IMF level 1:** Reconstruction from IMF level 1 recombined all components of the original signal. Therefore, no noise is removed. This resulted in a footfalls detection accuracy of 72%.
2. **IMF level 2:** Reconstruction from IMF level 2 removed some components of noise in the signal. This resulted in a footfalls detection accuracy of 82%.
3. **IMF level 3:** Reconstruction from IMF level 3 removed the most components of noise in the signal. This resulted in the best footfalls detection accuracy of 86%.
4. **IMF level 4:** Reconstruction from IMF level 4 removed key characteristics of the original. Therefore, important features and signal information is lost. This resulted in a footfalls detection accuracy of 78%.

Thus, is reasoned that IMF level 1 and IMF level 2 contained the signal's high-frequency noise components and IMF level 4 and higher levels contained key characteristics pertaining to the signal. This corroborates findings in [39, 45]. Ultimately, it was found that when using the EMD filtering technique, reconstructing the signal from IMF level 3 produced the highest

footfall detection accuracy (86%) further illustrated by Figure 4.6 and Figure 4.7. This has been deemed to occur owing to the fact that high-frequency noise and baseline drift was removed from the original signal while key signal characteristics were preserved.

5.2 Key Findings Impact

The proposed methodology can potentially be used to quantify the response to treatment in PD patients using kinematic signals acquired from a user-friendly, low-cost instrumented walker. The results extend previous studies that have shown the viability of these walker mounted sensors as an objective measurement system, provide feasibility to detect changes due to treatment. The results imply a capability to discriminate between PD patients who were administered medication, non-treatment patients and controls and to a lesser degree, between different stages of PD. A possible practical implication of these results is to provide more usable, continuous patient care: Such a device could be used by the patients when waiting for their appointment in the clinic, without needing a healthcare professional to administer the test, or at home, and allow clinicians to continuously monitor patient condition, treatment courses, and disease progression.

5.2.1.1 New Methods to Overcome Noise and Limitations of Low-cost Sensors

As discussed in Section 2.5.4, the PAMMII instrumented walker makes use of noisy, low-cost, off-body walker mounted sensors to record measurements. Compared to previous studies (Section 2.5.3), these properties of the system are significant, as former studies incorporated good quality, expensive sensors that are placed on the person's body which yield more reliable and consistent results but also reduced the affordability and usability of these systems. Using the PAMMII system provided a signal processing challenge. As shown, using signal pre-processing techniques (as discussed in Section 3.4) the limitations of low-cost noisy sensors can be overcome increasing the affordability of potential future solutions.

5.2.2 New Footfalls Detection Algorithm

The accelerometer footfalls detection algorithm utilised in this research correctly predicts footfalls with an 86% accuracy rate using a low-cost accelerometer that is placed on a 3rd party apparatus to the user. A footfall is seen as a basic element of walking and gait analysis. Therefore, this algorithm may be replicated using other instrumented walkers or exo-body device for gait disorder analysis tasks.

5.3 Research Limitations and Future Work Recommendations

The following list presents the limitations of this research and recommends future improvements:

- The movement stage identification algorithm and footfall detection algorithm have been validated using video evidence. However, the subjects used for this validation were controls who simulated PD gait disorders (such as tremors, limping, slow walking etc.). Moreover, only 4 controls were used to validate this algorithm. Therefore, it is recommended that a larger dataset of controls, PD patients and/or patients with gait disorders is used to further validate these algorithms.
- The dataset incorporated 51 subjects in total. This is a substantial number compared to previous studies on the sensor-based classification of PD. However, the groups of treatment patients, non-treatment patients, and controls were different in the distribution of ages and gender. Moreover, the small size of the non-treatment patient group makes this bias even more apparent. Therefore, it is recommended that a larger dataset is used to further validate this classification and enhance system credibility.
- The dataset captures information from subjects over a short-term interval (two short tests 3-hours apart). It is recommended that future trials capture the data over a longer period, to see changes in response over time.
- The H&Y clinical scores were assessed by one clinician. This could have produced a subjectivity into the results obtained. Therefore, it is recommended that future studies utilise multiple healthcare professionals to assign H&Y clinical scores to the PD patients to remove this bias.
- It is recommended that only one of the 3m or 10m TUG walking tests are implemented in future studies as it has been shown that they produce similar results.

5.4 Conclusion

The research included three objectives, designed to answer the research question: “To what extent can features extracted from signals acquired by walker mounted sensors identify changes in PD patient gait due to treatment.”

The objectives included:

1. To determine to what extent combinations of these features provide a classification of the three groups.
2. To determine to what extent these features or their combinations identify the severity of the disease according to H&Y clinical scales.
3. To determine which of these features, derived from kinematic signals from walker mounted sensors best discriminate between PD patients gait changes due to treatment compared to non-treatment subjects and controls.

The first research objective was answered using the feature-set selected by PCA to discriminate between the 3 subject groups: The best classification performance of the BDT classifier resulted from the 3m TUG data with the highest accuracy of 92.16% ± 3.92%, sensitivity of 82.35%, and specificity of 93.38%.

The second research objective was answered using the feature-set selected by the 1-way ANOVA method to discriminate between the treatment patients clinically assigned H&Y scores of H&Y 2 and H&Y 3 respectively: The best classification performance of the BDT classifier resulted from the 3m TUG data with the highest accuracy of 79.41% ± 4.12%, sensitivity of 85.00%, and specificity of 73.43%.

The third research objective was answered by evaluating the feature subsets of the best performing feature selection method (PCA), in terms of accuracy sensitivity and specificity and identifying features that contributed significantly to the feature selection methods: The key features subset reduced dimensionality further while producing similar performance to the PCA subset: an accuracy of 90.20% ± 4.26%, sensitivity of 81.37%, and specificity of 91.64%.

The future improvement in this study incorporated the validation of the methodology and its results using a larger dataset. However, this research demonstrated both the capabilities and the limitations of the signal processing techniques and machine learning algorithms employed on signals acquired from low-cost sensors mounted on a 3rd party apparatus (Walker)in several discrimination experiments using. These quantitative methods could assist healthcare professionals with the diagnosis and monitoring of PD patients.

In conclusion, this study suggested the feasibility of using signals acquired by walker mounted sensors to identify changes in PD patient gait due to treatment and to discriminate between PD stages based on these changes.

References

[1] E. Naqvi, “Parkinson’s Disease Statistics,” 7 August 2017. [Online]. Available: https://parkinsonsnewstoday.com/parkinsons-disease-statistics/. [Accessed 2018 February 13].

[2] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger and F. Seitelberger, “Brain dopamine and the syndromes of Parkinson and Huntington Clinical, morphological and neurochemical correlations,” *Journal of the Neurological Sciences,* vol. 20, no. 4, pp. 415-455, 1973.

[3] A. Gröger, R. Kolb, R. Schäfer and U. Klose, “Dopamine Reduction in the Substantia Nigra of Parkinson's Disease Patients Confirmed by In Vivo Magnetic Resonance Spectroscopic Imaging,” *Plos One,* vol. 10, 2014.

[4] T. Müller, “Drug therapy in patients with Parkinson’s disease,” *Translational Neurodegeneration,* vol. 1, no. 10, 24 May 2012.

[5] A. Mandal, “Parkinson's Disease History,” 24 Hanuary 2013. [Online]. Available: https://www.news-medical.net/health/Parkinsons-Disease-History.aspx. [Accessed 2018 February 13].

[6] J. Perlmutter, “Assessment of Parkinson Disease Manifestations,” *Current Protocols in Neuroscience,* 2009.

[7] K. Chaudhuri, D. Healy and A. Schapira, “Non-motor symptoms of Parkinson’s disease: Diagnosis and management,” *Lancet Neurol,* 2006.

[8] T. P. S. Group, “Levodopa and the Progression of Parkinson's Disease,” *The New England Journal of Medicine,* vol. 351, pp. 2498-2508, 2004.

[9] MedicineNet.com, “Medical Definition of Gait,” 2018. [Online]. Available: https://www.medicinenet.com/script/main/art.asp?articlekey=3533.

[10] S. V. Perumal and R. Sankar, “Gait and tremor assessment for patients with parkinsons disease using wearable sensors,” *ICT Express, special Issue on Emerging Technologies for Medical Diagnostics,* vol. 2, no. 4, p. 168 – 174, 2016.

[11] C. G. e. al, “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, 2016.

[12] M. Plotnik, N. Giladi, Y. Balash, C. Peretz and J. M. Hausdorff, “Is freezing of gait in parkinson’s disease related to asymmetric motor function?,” *Annals of neurology,,* vol. 57, no. 5, pp. 656-663, 2005.

[13] S. Bamberg, A. Benbasat, D. Scarborough, D. Krebs and J. Paradiso, “Gait Analysis Using a Shoe-Integrated Wireless Sensor System,” *IEEE Transactions on Information Technology in Biomedicine,* vol. 12, no. 4, pp. 413-423, 2008.

[14] W. Xu, Y. Shen, Y. Zhang, N. Bergmann and W. Hu, “Gait-Watch: A Context-Aware Authentication System for Smart Watch Based on Gait Recognition,” in *Internet-of-Things Design and Implementation* , Pittsburgh, 2017.

[15] L. Li, L. Atallah, B. Lo and G.-Z. Yang, “Feature Extraction from Ear-Worn Sensor Data for Gait Analysis,” *IEEE,* 2014.

[16] V. Aharonson, I. Schlesinger, A. McDonald, S. Dubowsky and A. Korczyn, “A Practical Measurement of Parkinson’s Patients Gait Using Simple Walker Based Motion Sensing and Data Analysis,” *ASME Journal of Medical Devices,* 2017.

[17] V. Aharonson, I. Schlesinger, A. McDonald, S. Dubowsky and A. Korcyzn, “Monitoring of Parkinson’s patients gait using simple walker based motion sensing and data analysis,” in *Design of Medical Devices Conference (DMD)*, MN, USA, 2017.

[18] A. Fil, R. Cano-de-la-Cuerda, E. M. Hellín, L. Velac, M. Ramiro-González and C. Fernández-de-las-Peñas, “Pain in Parkinson disease: A review of the literature,” *Parkinsonism & Related Disorders,* vol. 19, no. 3, pp. 285-294, March 2013.

[19] C. A. Davie, “A review of Parkinson's disease,” *British Medical Bulletin,* vol. 86, no. 1, pp. 109-127, 2008.

[20] 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, no. 2, pp. 142-148, 1991.

[21] A. Shumway-Cook, S. Brauer and M. Woollacott, “Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test,” *Physical therapy,* vol. 80, no. 9, p. 896–903, 2000.

[22] G. Lewis, W. Byblow and S. Walt, “Stride length regulation in Parkinson's disease: the use of extrinsic, visual cues,” *Brain,* vol. 123, no. 10, p. 2077–2090, 2000.

[23] J. Hausdorff, “Gait dynamics in Parkinson’s disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling,” *Chaos,* vol. 19, no. 2, 2009.

[24] M. Karadsheh, “Gait Cycle,” 2018. [Online]. Available: https://www.orthobullets.com/foot-and-ankle/7001/gait-cycle.

[25] T. Sayeed, “Methods and models in signal processing for gait analysis using waist-worn accelerometer: A contribution to parkinsons disease,” 2015.

[26] J. Fletcher, “What is abnormal gait?,” Medical News Today, 2017. [Online]. Available: https://www.medicalnewstoday.com/articles/320481.php.

[27] N. Mortaza, N. Abu-Osman and N. Mehdikhani, “Are the spatio-temporal parameters of gait capable of distinguishing a faller from a non-faller elderly?,” *European Journal of Physical and Rehabilitation Medicine,* pp. 677-691, 2014.

[28] M. Corporation, “System for Gait Analysis,” Microgate Corporation, 2012. [Online]. Available: http://www.optogait.com. [Accessed 24 February 2018].

[29] Vicon, “Vicon, Camera Systems,” 2016. [Online]. Available: https://www.vicon.com/products/camera-systems. [Accessed 24 February 2018].

[30] Codamotion, “3D Clinical Gait Systems - Reliable & Easy,” 2017. [Online]. Available: http://codamotion.com/index.php/technology/clinical-gait-analysis. [Accessed 24 February 2018].

[31] Xsens, “Gait Analysis with Xsens inertial sensor technology,” 2018. [Online]. Available: https://www.xsens.com/tags/gait-analysis/?gclid=EAIaIQobChMI2dXb47y-2QIVqLztCh1STguWEAAYASAAEgLkp_D_BwE. [Accessed 24 February 2018].

[32] G. Cutlip, R. Mancinelli, C. Huber and F. DiPasquale, “Evaluation of an instrumented walkway for measurement of the kinematic parameters of gait.,” *Gait & posture,* vol. 12, no. 2, pp. 134-138, 2000.

[33] B. Rice, K. Toth and C. Robinson, “Instrumented Walkways,” Feinberg School of Medicine, 2014. [Online]. Available: http://www.nupoc.northwestern.edu/docs/news/instrumented-walkwaysCR.pdf.

[34] GaitRite, “About GaitRite,” 2015. [Online]. Available: https://www.gaitrite.com/about.

[35] ProtoKinetics, “The Zeno Walkway,” ProtoKinetics, 2018. [Online]. Available: http://www.protokinetics.com/zeno-walkway/.

[36] M. Alwan, A. Ledoux, G. Wasson, P. Sheth and C. Huang, “Basic walker-assisted gait characteristics derived from forces and moments exerted on the walker’s handles: Results on normal subjects,” *Medical Engineering and Physics,* vol. 29, p. 380–389, 2006.

[37] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman and L. O. Ramig, “Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease,” *IEEE Transactions on Biomedical Engineering ,* vol. 59, no. 5, pp. 1264-1271, 2012.

[38] P. Huang, C. G. Goetz, R. F. Woolson, B. Tilley, D. Kerr, Y. Palesch, J. Elm, B. Ravina, K. J. Bergmann, K. Kieburtz and T. P. S. Group, “Using Global Statistical Tests in Long-Term Parkinson’s Disease Clinical Trials,” *PubMed,* vol. 24, no. 12, 2010.

[39] J. Ballesteros, C. Urdiales, A. B. Martinez and M. Tirado, “Automatic Assessment of a Rollator-User’s Condition During Rehabilitation Using the i-Walker Platform,” *IEEE Transactions on Neural Systems and Rehabilitatio Engineering,* vol. 21, no. 11, 2017.

[40] B. Lathi, Modern Digital and Analog Communication Systems 3e Osece, Oxford university press, 1998.

[41] 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,” *26th International Conference on Computer Communication and Networks (ICCCN),* pp. 1-5, 2017.

[42] K.-C. Lan and W.-Y. Shih, “Early Diagnosis of Parkinson’s Disease using a Smartphone,” in *International Conference on Mobile Systems and Pervasive Computing*, Taiwan, 2014.

[43] H. Dai, P. Zhang and T. C. Lueth, “Quantitative Assessment of Parkinsonian Tremor Based on an Inertial Measurement Unit,” *Sensors (Basel),* vol. 15, no. 10, 2015.

[44] 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,,” *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* p. 3686–3689, 2008.

[45] J. Nunes and E. Delchelle, “Empirical mode decomposition: Applications on signal and image processing,” *Advances in Adaptive Data Analysis,* vol. 1, no. 1, 2009.

[46] J. Favre, B. Jolles, O. Siegrist and K. Aminian, “Quaternion-based fusion of gyroscopes and accelerometers to improve 3D angle measurement,” *Electronics Letters ,* vol. 42, no. 11, pp. 612 - 614, 2006.

[47] E. Piña, “Rotations with Rodrigues' vector,” *European Journal of Physics ,* vol. 32, no. 5, 2011.

[48] S. Yean, B. S. Lee, C. K. Yeo and C. H. Vun, “Algorithm for 3D orientation estimation based on Kalman Filter and Gradient Descent,” *Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016 IEEE 7th Annual,* 2016.

[49] L. Atallah, O. Aziz, E. Gray, B. Lo and G.-Z. Yang, “An ear-worn sensor for the detection of gait impairment after abdominal surgery,” *Surgical Innovation,* vol. 20, no. 1, pp. 86-94, 2013.

[50] S. Byun, J. W. Han, T. H. Kim and K. W. Kim, “Test-Retest Reliability and Concurrent Validity of a Single Tri-Axial Accelerometer-Based Gait Analysis in Older Adults with Normal Cognition,” *PLoS ONE,* vol. 11, no. 7, 18 July 2016.

[51] D. Jarchi, J. Pope, T. K. Lee, L. Tamjidi, A. Mirzaei and S. Sanei, “A Review on Accelerometry Based Gait Analysis and Emerging Clinical Applications,” *IEEE Reviews in Biomedical Engineering,* no. 99, 2018.

[52] R. Moe-Nilssen and J. L. Helbostad, “Estimation of gait cycle characteristics by trunk accelerometry,” *Journal of Biomechanics,* pp. 121-126, 2003.

[53] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” The University of Waikato, Ph.D. dissertation, New Zealand, 1999.

[54] 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,” *International Workshop on Wearable and Implantable Body Sensor Networks (BSN’06),* 2006.

[55] 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? 45 accelerometers reveal that reduced movement quality while walking is associated with parkinson’s disease, ageing and fall risk,” *6th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* p. 5968–5971, 2014.

[56] I. Jolliffe, “Principal component analysis,” in *International encyclopedia of statistical science*, Berlin, 2011.

[57] Stephanie, “Normalized Data / Normalization,” Statistics How To, 2018. [Online]. Available: http://www.statisticshowto.com/normalized/.

[58] L. Palmerini, “Data mining in clinical practice for the quanti cation of motor impairment in parkinson’s disease,” Ph.D. dissertation, University of Bologna, 2012.

[59] D. Spina, “Statistics in Pharmacology,” *British Journal of Pharmacology,* vol. 152, no. 3, p. 291–293, 2007.

[60] R. Shier, “Statistics,” 2004. [Online]. Available: http://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf.

[61] A. Worker, C. Blain, J. Jarosz, K. R. Chaudhuri, G. J. Barker, S. C. R. Williams, R. G. Brown, P. N. Leigh, F. Dell’Acqua and A. Simmons, “Diffusion Tensor Imaging of Parkinson’s Disease, Multiple System Atrophy and Progressive Supranuclear Palsy: A Tract-Based Spatial Statistics Study,” *PubMed,* vol. 9, no. 11, 2014.

[62] N. M. Nasrabadi, “Pattern Recognition and Machine Learning,” *Journal of Electronic Imaging,* vol. 16, no. 4, 2007.

[63] S. Bind, A. Tiwari and A. Sahani, “A survey of machine learning based approaches for parkinson disease prediction,” *International Journal of Computer Science and Informa- tion Technologies,* vol. 6, no. 2, p. 1648–1655, 2015.

[64] S. Das, L. Trutoiu, A. Murai, D. Alcindor, M. Oh, F. D. l. Torre and J. Hodgins, “Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data,” in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, Boston, 2011.

[65] D. Wu, K. Warwick, Z. Ma, M. Gasson, J. Burgess, S. Pan and T. Aziz, “Prediction of Parkinson's disease tremor onset using a radial basis function neural network based on particle swarm optimization,” *International Journal of Neural Systems,* vol. 20, no. 2, 2012.

[66] H. Manap, N. Tahir and A. Yassin, “Statistical analysis of parkinson disease gait classification using artificial neural network,” *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),* pp. 60-65, 2011.

[67] E. Tripoliti, A. Tzallas, M. Tsipouras, G. Rigasa, P. Bougiaa, M. Leontiou, S. Konitsiotis, M. Chondrogiorgi, S. Tsouli and D. Fotiadis, “Automatic detection of freezing of gait events in patients with Parkinson's disease,” *Computer Methods and Programs in Biomedicine,* vol. 110, no. 1, pp. 12-26, 2013.

[68] Y. Freund and L. Mason, “The alternating decision tree learning algorithm,” in *icml*, 1999.

[69] R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology.,” *IEEE transactions on systems, man, and cybernetics,* vol. 21, no. 3, pp. 660-674, 1991.

[70] S. Dudoit and J. Fridlyand, “Bagging to improve the accuracy of a clustering procedure,” *Bioinformatics,* vol. 19, no. 9, pp. 1090-1099, 2003.

[71] Y. Wu, P. Chen, Y. Yao, X. Ye, Y. Xiao, L. Liao, M. Wu and J. Chen, “Dysphonic Voice Pattern Analysis of Patients in Parkinson's Disease Using Minimum Interclass Probability Risk Feature Selection and Bagging Ensemble Learning Methods,” *Computational and mathematical methods in medicine,* 2017.

[72] N. Fayyazifar and N. Samadiani, “Parkinson's disease detection using ensemble techniques and genetic algorithm,” *Artificial Intelligence and Signal Processing Conference (AISP),* 2017.

[73] N. Seedat, D. Beder, V. Aharonson and S. Dubowsky, “A comparison of footfall detection algorithms from walker mounted sensors data,” in *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Istanbul, 2018.

[74] P. Jiang, H. Wu, W. Wang, W. Ma, X. Sun and Z. Lu, “classification of real and pseudo microRNA precursors using random forest prediction model with combined features,” *Nucleic Acids Research,* vol. 35, no. 2, pp. 339-344, 2007.

[75] P. Kline, An easy guide to factor analysis, Routledge, 2014.

[76] M. Mukaka, “A guide to appropriate use of Correlation coefficient in medical research,” *Malawi Medical Journal,* vol. 24, no. 3, pp. 69-71, 2012.

[77] J. Klucken, J. Barth, P. Kugler, J. Schlachetzki, T. Henze, F. Marxreiter, Z. Kohl, R. Steidl, J. Hornegger, B. Eskofier and J. Winkler, “Unbiased and Mobile Gait Analysis Detects Motor Impairment in Parkinson's Disease,” *PLOSone,* vol. 8, no. 2, 2013.

[78] D. l. F.-F. R, R. TJ, S. V, S. M, C. DB and S. AJ, “Expectation and dopamine release: Mechanism of the placebo effect in Parkinson’s disease,” *Science,* p. 1164–1166, 2000.

[79] H. Sawada, T. Oeda, S. Kuno, M. Nomoto, K. Yamamoto, M. Yamamoto, K. Hisanaga and T. Kawamura, “Amantadine for Dyskinesias in Parkinson's Disease: A Randomized Controlled Trial,” *PLoS One,* vol. 5, no. 12, 2010.

[80] L. V. Metman, P. D. Dotto, P. v. d. Munckhof, J. Fang, M. Mouradian and T. N. Chase, “Amantadine as treatment for dyskinesias and motor fluctuations in Parkinson's disease,” *Neurology,* vol. 50, no. 5, pp. 1323-1326, May 1998.

[81] U. o. R. M. Center, “Parkinson's disease: A looming pandemic,” 13 November 2017. [Online]. Available: https://medicalxpress.com/news/2017-11-parkinson-disease-looming-pandemic.html. [Accessed 2018 February 15].

[82] M. Hariharana, K. Polat and R. Sindhuc, “A new hybrid intelligent system for accurate detection of Parkinson's disease,” *Computer Methods and Programs in Biomedicine,* vol. 113, no. 3, pp. 904-913, 2014.

[83] H. Gill, J. Morris, E. Biden and J. O’Connor, “Optometric methods in biomechanical gait analysis,” in *Optical measurement methods in biomechanics*, Boston, Springer, Boston, MA, pp. 125-153.

[84] M. Clinic, “Parkinson's disease,” 2018. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/diagnosis-treatment/drc-20376062.

[85] Drugs.com, “Drugs Used to Treat Parkinson's Disease,” 2018. [Online]. Available: https://www.drugs.com/condition/parkinson-s-disease.html.

[86] J. Annegarn, M. A. Spruit, H. H. C. M. Savelberg, P. J. B. Willems, C. v. d. Bool, A. M. W. J. Schols, E. F. M. Wouters and K. Meijer, “Differences in Walking Pattern during 6-Min Walk Test between Patients with COPD and Healthy Subjects,” *PLoS ONE,* 2012.

[87] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K. Biglan, E. Dorsey and M. Little, “Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study,” *Parkinsonism and Related Disorders,* vol. 21, no. 5, p. 650–653, 2015.

[88] L. Flavell, “Surgical Treatments for Parkinson’s Disease,” 2018. [Online]. Available: https://parkinsonsnewstoday.com/parkinsons-disease-treatments/surgical-treatments/.

[89] H. Thimbleby, “Technology and the Future of Healthcare,” *Journal of Public Health Research,* vol. 2, no. 3, 2013.

[90] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Patel and A. Haines, “The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis,” *Plos Medicine,* 2013.

[91] J. Benesty, J. Chen, Y. Huang and I. Cohen, “Pearson Correlation Coefficient,” in *Noise Reduction in Speech Processing*, Springer Topics in Signal Processing, 2009, pp. 1-4.

Appendix A. Ethics Clearance Certificate

The following document describes the ethical clearance obtained from the Humans Research Ethics Committee of the University of the Witwatersrand for this research task.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mr David Beder

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M180203

NAME: (Principal Investigator)	Mr David Beder
DEPARTMENT:	School of Electrical and Information Engineering
PROJECT TITLE:	Objective Analysis of a Parkinson's Patients Health Before and After Treatment Based on Signal Processing
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).

David Beder — Principal Investigator Signature

23/02/18 — Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B. 10m TUG Data Results

B.1. Experiment 1 – Finding the Optimal Number of Trees

As with all machine learning tasks, there is a trade-off between classification accuracy and complexity or execution time [68]. In the BDT method, classification performance and execution time are proportional to the number of decision trees within the BDT classifier [68, 70]. Therefore, to make the system both accurate and efficient, an optimal number of trees to classify a class has been calculated.

All 51 subjects' TUG data was used for this experiment using 10-fold cross-validation. Figure B.1 shows that approximately 50 decision trees are the optimal number of trees needed for bootstrap aggregating to yield the highest classification accuracy of the bagged decision tree algorithm across the five feature selection methods, for the 10m TUG data.

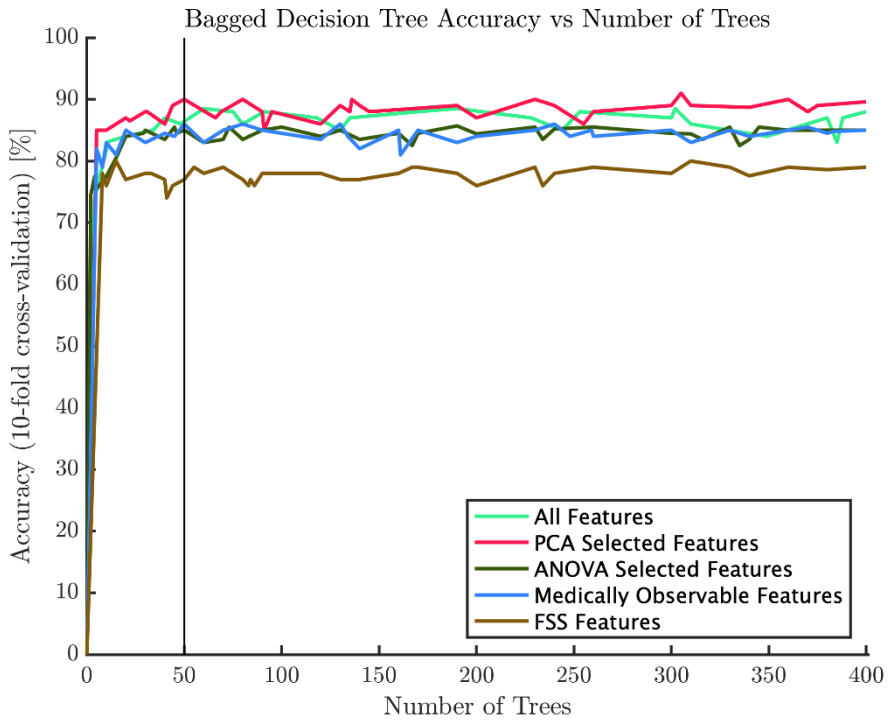

Figure B.1: Accuracy of the Bagged Decision Tree classifier vs. Number of Trees, highlighting the optimal number of trees needed across the implemented feature selection methods, for the 10m TUG data

B.2. Experiment 2 – Classification of the 3 Subject Groups

This experiment was performed to determine the discriminating power of the derived features, in differentiating between the subject groups (treatment patients, non-treatment patients, and controls).

All 51 subjects' TUG data was used for this experiment using 10-fold cross-validation and 50 decision trees. It was noted for this experiment that there was a class imbalance with a smaller number of non-treatment patients within the dataset. In order to diminish the effects of this imbalance, the BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier.

Figure B.2 depicts the performance of the BDT classifier across the five feature selection methods, highlighting the average accuracy of the BDT classifier for each input within a 95% confidence bound.

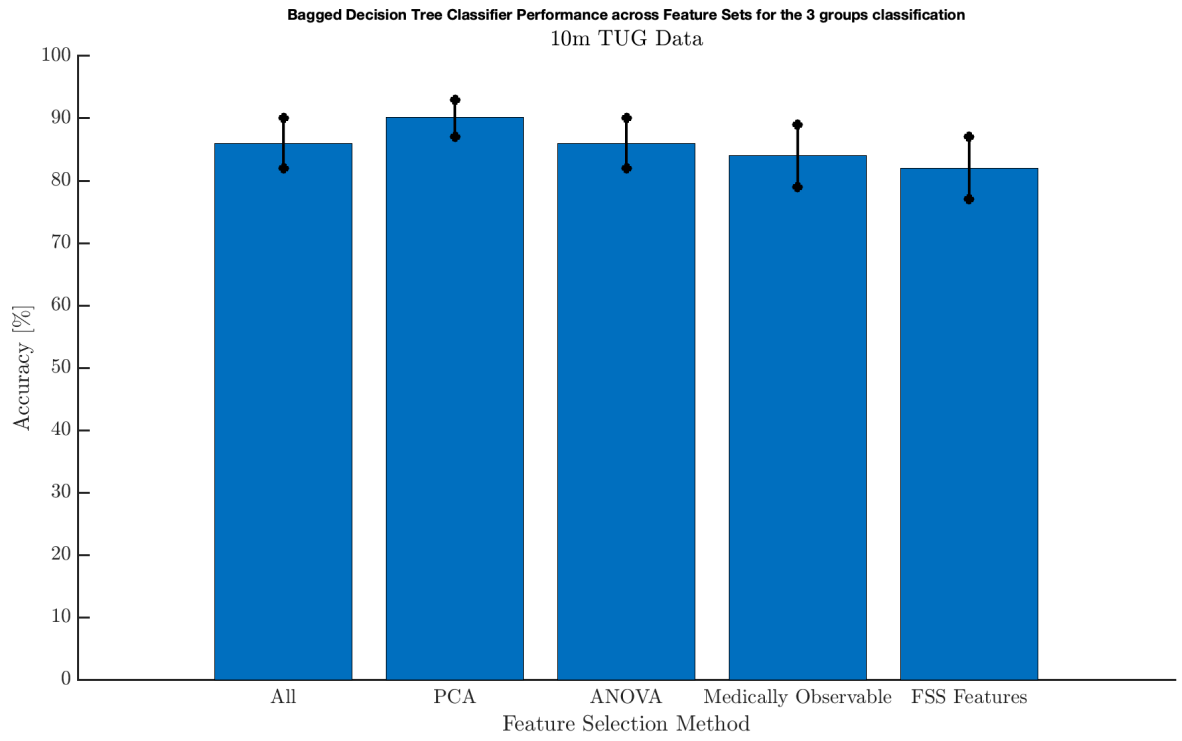

Figure B.2: Average accuracy of the BDT classifier using the five feature selection methods as an input to classify the 3 subject groups, for the 10m TUG data

As seen, using the feature-set selected by PCA as an input results in the best BDT classification performance. Figure B.3 depicts the confusion matrix of the Bagged Decision Tree classifier using this feature-set, highlighting the nature of the classification errors, in terms of how each class is misclassified.

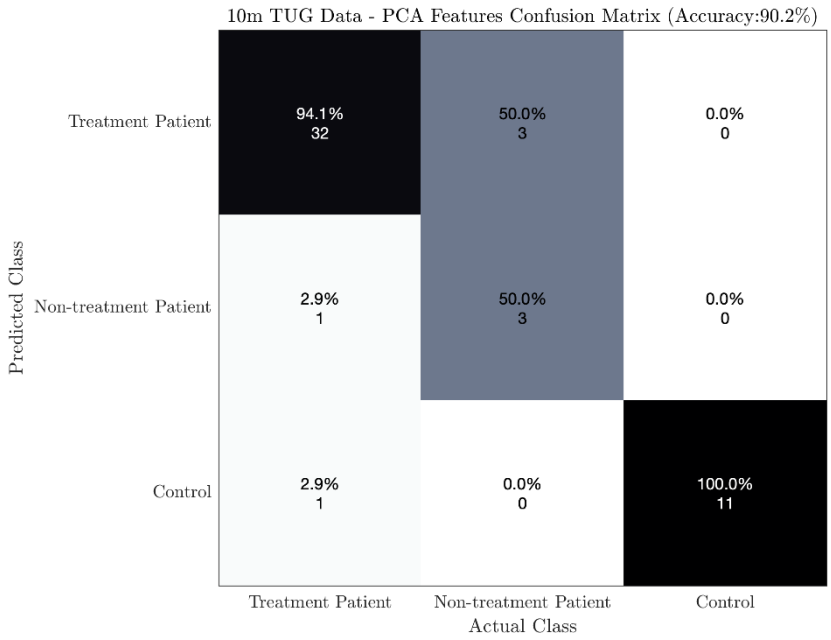

Figure B.3: Confusion Matrix of the BDT classifier using the feature-set selected by PCA as an input for classification of the 3 Subject Groups. The actual classes (columns) are the subjects divided into their real subject groups (treatment patients, non-treatment patients and control) and the predicted classes (rows) are BDT classifier output of each subject into the 3 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 10m TUG data

Table B.1 expresses the individual sensitivity and specificity of each class (subject group) as well as the average sensitivity and specificity and accuracy of the BDT classifier using the feature-set selected by PCA as its input.

Table B.1: Classification results for the BDT classifier using the feature-set selected by PCA as its input for classification of the 3 subject groups, for the 10m TUG data

	Class			
	Treatment Patients	**Non-treatment Patients**	**Controls**	**Average**
Sensitivity [%]	94.12	50.00	100.00	81.37
Specificity [%]	82.35	97.78	97.50	92.54

Figure B.2 shows that using the PCA feature selection method as an input results in the best classification performance of the BDT classifier with the highest accuracy of 90.20% ± 3.86%, a sensitivity of 81.37% and a specificity of 92.54%, for the 10m TUG data.

B.3. Experiment 3 – Classification between Treatment Patients and Non-treatment Patients

This experiment was performed to further assess the distinguishing power of the derived features, in differentiating and discriminating between the PD patients who were administered medication and the PD patients who were *not* administered medication.

Hence, all controls were removed from this experiment. The remaining 40 patients' TUG data was used for this experiment using 10-fold cross-validation and 50 decision trees. It was noted for this experiment that the class imbalance still exists, with a much smaller number of non-treatment patients compared to the treatment patients. In order to diminish the effects of this imbalance, the BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier.

Figure B.4 shows the average accuracy of the BDT classifier using the five feature selection methods as an input to classify between treatment patients and non-treatment patients within a 95% confidence bound.

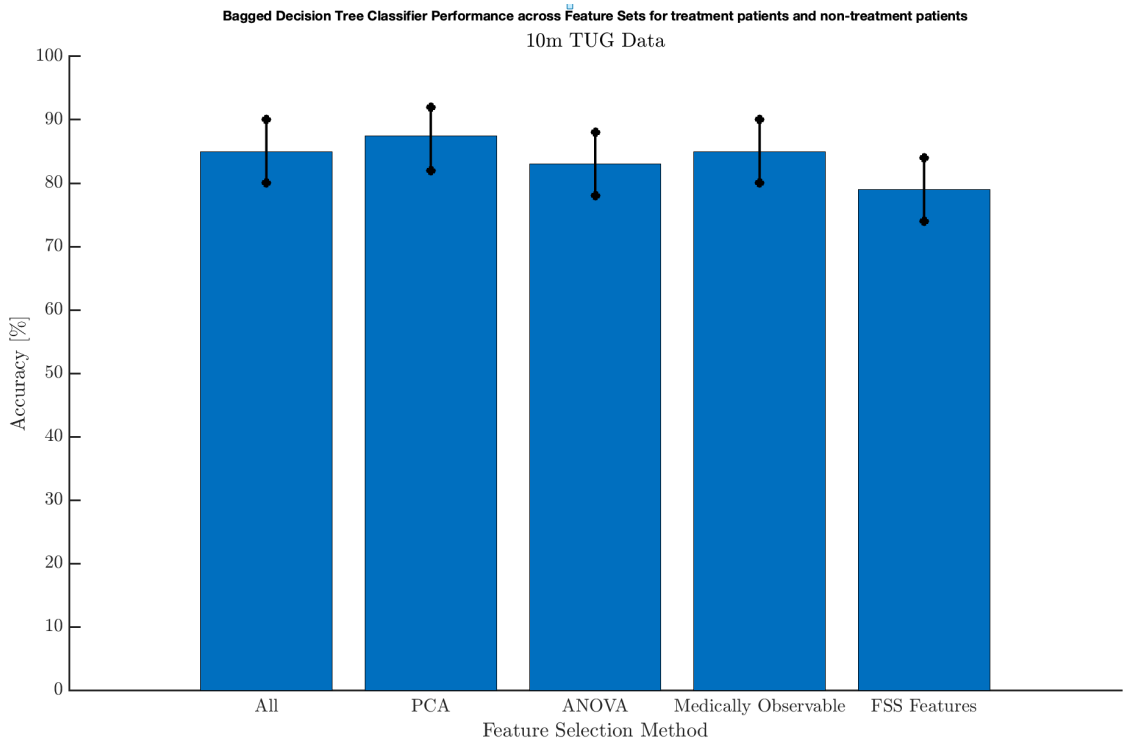

Figure B.4: Accuracy of the BDT classifier using the five feature selection methods as an input to classify treatment patients vs. non-treatment patients, for the 10m TUG data

As seen, like Experiment 2, using the feature-set selected by PCA as an input results in the best BDT classification accuracy. Figure B.5 depicts the confusion matrix of the BDT classifier using this feature-set as its input. This confusion matrix displays the best

performance of the BDT classifier for this experiment and in doing so highlights the nature of how the two classes are misclassified.

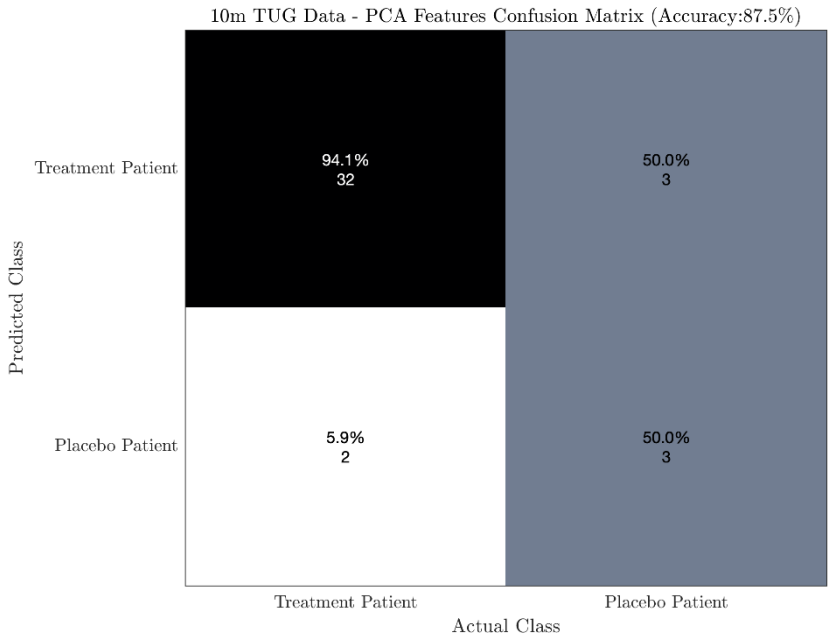

Figure B.5: Confusion Matrix of the Bagged Decision Tree Algorithm using the feature-set selected by PCA as an input for classification of treatment patients vs. non-treatment patients. The actual classes (columns) are the subjects divided into treatment patients and non-treatment patients and the predicted classes (rows) are BDT classifier output of each subject into these 2 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 10m TUG data

Table B.2 expresses the accuracy, sensitivity and specificity of the BDT classifier using the feature-set selected by PCA as it input, highlighting the detection rate and false detection rate of the treatment patients and the non-treatment patients.

Table B.2: Classification results for the BDT classifier using the feature-set selected by PCA as its input for classification of treatment patients vs. non-treatment patients, for the 10m TUG data

Accuracy [%]	87.50 ± 4.63
Sensitivity [%]	94.12
Specificity [%]	50.00

Figure B.4 demonstrate that using the PCA feature selection method as an input results in the best classification performance of the BDT classifier with the highest accuracy of 87.50% ± 4.63%, a sensitivity of 94.12% and a specificity of 50.00%, for the 10m TUG data.

B.4. Experiment 4 – Classification between Treatment Patients H&Y 2 and H&Y 3 Clinical Scores

This experiment was performed to assess the distinguishing power of the derived features, in differentiating and discriminating between *only* the treatment patients with H&Y clinical scores of 2 and 3.

Hence, all non-treatment patients and controls were removed from this experiment. The remaining 34 subjects' TUG data was used for this experiment using 10-fold cross-validation and 50 decision trees. The BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to measure the overall performance of the BDT classifier.

Figure B.6 depicts the accuracy of the BDT classifier using the five feature selection methods as an input to classify treatment patients in terms of their H&Y 2 and H&Y 3 clinical scores within a 95% confidence bound.

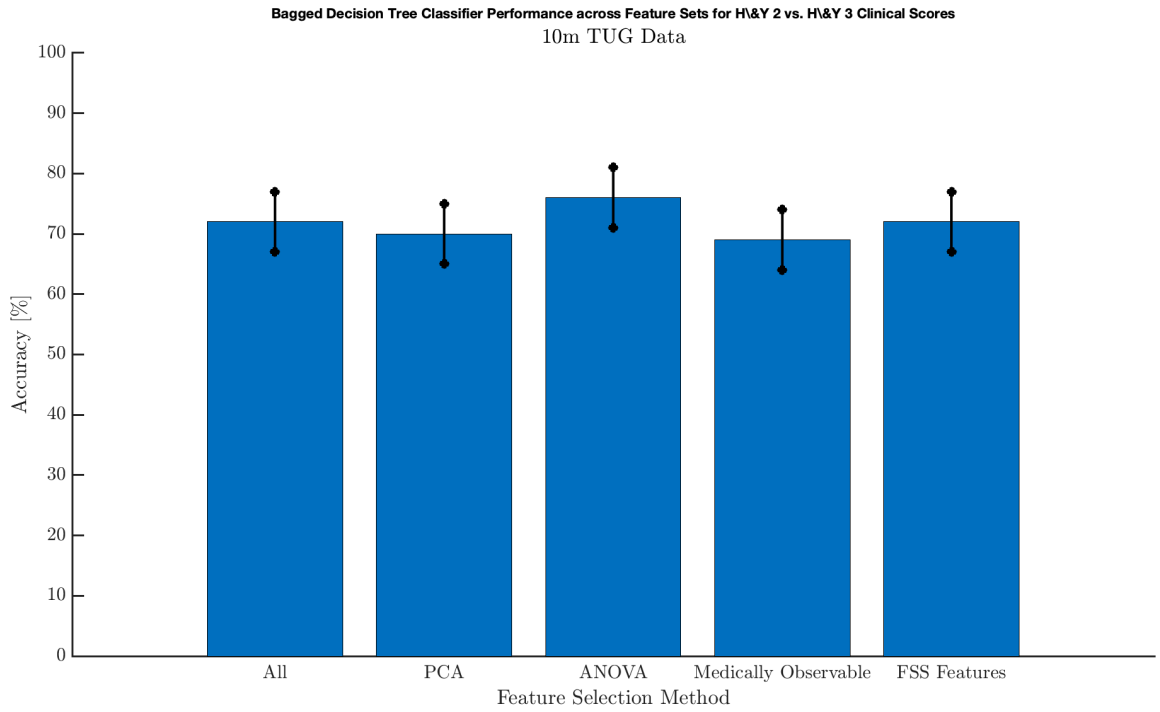

Figure B.6: Accuracy of the BDT classifier using the five feature selection methods as an input to classify PD patients H&Y 2 and H&Y 3 clinical scores, for the 10m TUG data

Unlike Experiment 2 and Experiment 3, using the feature-set selected by the 1-way ANOVA method as an input results in the best BDT classification accuracy. Figure B.7 shows the

confusion matrix of the Bagged Decision Tree classifier using this feature-set, highlighting the nature of the classification errors, in terms of how each class is misclassified.

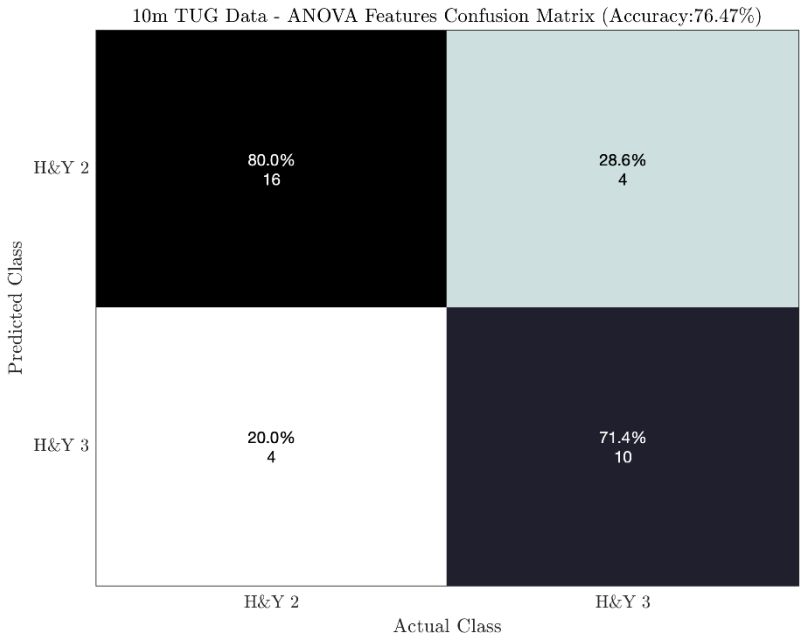

Figure B.7: Confusion Matrix of the BDT Algorithm using the feature-set selected by ANOVA as an input for classification of the treatment patients according to their H&Y clinical scores. The actual classes (columns) are the treatment patients divided into H&Y 2 and H&Y 3 clinical scores and the predicted classes (rows) are BDT classifier output of each subject into these 2 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 10m TUG data

Table B.3 expresses the accuracy, sensitivity and specificity of the BDT classifier using the feature-set selected by the 1-way ANOVA method as it input, highlighting the detection rate and false detection rate of the treatment patients according to their H&Y clinical scores.

Table B.3: Classification results for the BDT classifier using the feature-set selected by ANOVA as its input to classify treatment patients in terms of their H&Y 2 and H&Y 3 clinical scores, for the 10m TUG data

Accuracy [%	76.47 ± 4.57
Sensitivity [%]	80.00
Specificity [%]	71.43

Figure B.6 shows that using the 1-way ANOVA feature selection method as an input results in the best classification performance of the BDT classifier with the highest accuracy of 76.47% ± 4.57%, a sensitivity of 80.00% and a specificity of 71.43%, for the 10m TUG data.

B.5. Key Feature Validation

In order to validate that the key features of Table 4.12 are able to discriminate between the treatment patients, non-treatment patients and controls, these features are fed into the BDT classifier and its accuracy, sensitivity, and specificity is compared with the results of Experiment 1.

The 51 subjects' 10m TUG data was used for this validation experiment using 10-fold cross-validation and 50 decision trees. Similar to Experiment 1, it was noted for this experiment that the class imbalance exists due to the significantly smaller number of non-treatment patients. In order to diminish the effects of this imbalance, the BDT classifier has been implemented using bootstrap aggregating. Consequently, 50 random subsamples have been taken using sampling with replacement. Additionally, extra evaluation metrics have been included in the form of accuracy, sensitivity, specificity and confusion matrices to validate the overall performance of the BDT classifier.

Figure B.8 compares the performance of the BDT classifier using the key features subset as input against the feature-sets selected by PCA method, highlighting the accuracy as well as the average sensitivity and the average specificity of the BDT classifier for each input.

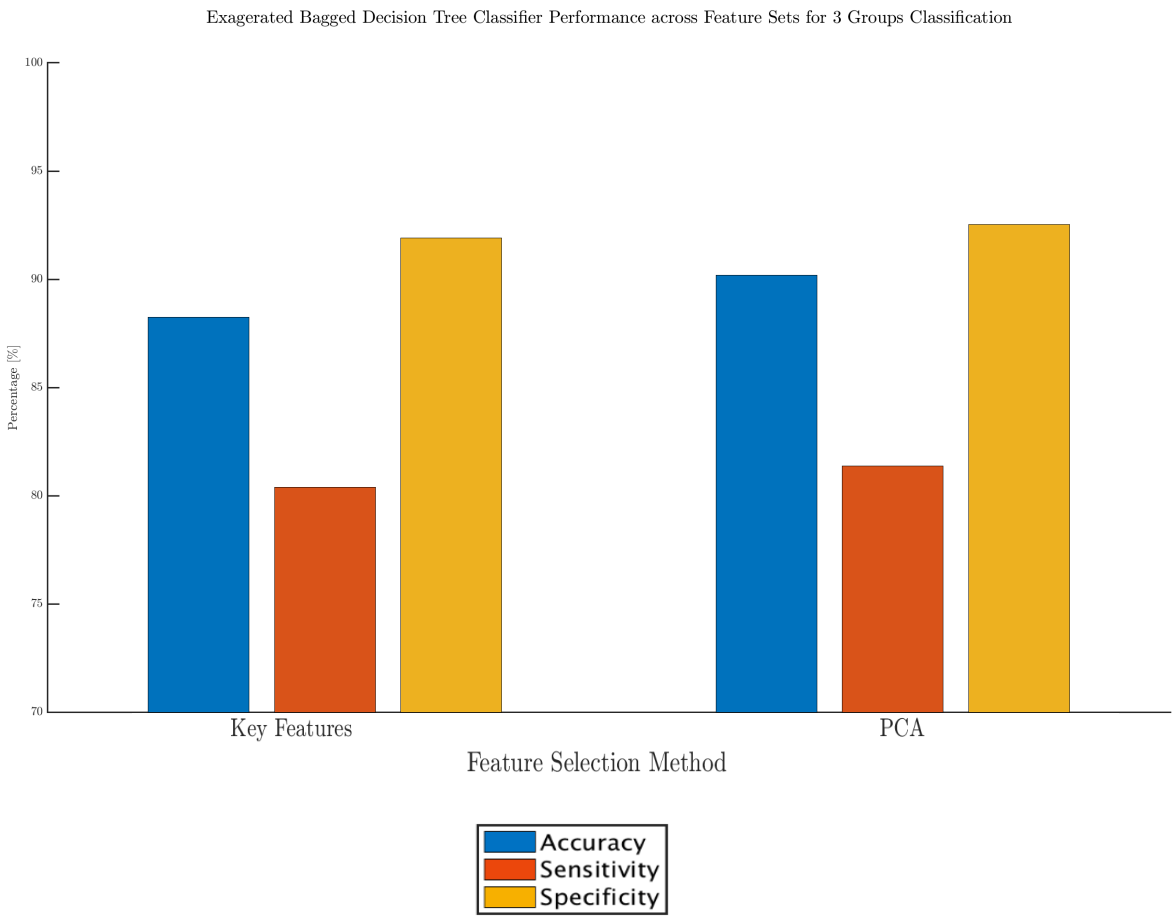

Figure B.8: Performance comparison of the BDT classifier comparing the key features performance against the PCA feature selection method in terms of classification of the 3 subject groups, for the 10m TUG data

The figure illustrates that using the key features feature-set as an input yields a slightly reduced performance than the PCA method, but the proportion between the measures is similar. Figure B.9 depicts the confusion matrices of the BDT classifier using the key features as its input for the 10m TUG data. This highlights the nature of the classification errors, in terms of how each class is misclassified.

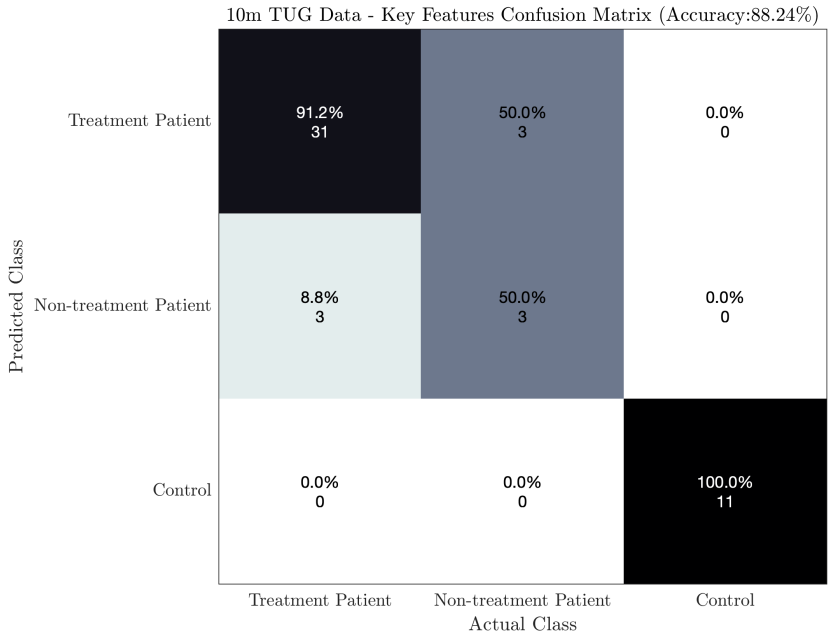

Figure B.9: Confusion Matrix of the BDT Classifier using the key features as an input for classification of the 3 subject groups. The actual classes (columns) are the subjects divided into their real subject groups (treatment patient, non-treatment patient and control) and the predicted classes (rows) are the BDT classifier output of each subject into the 3 groups. Each cell contains both the number and percentage of subjects classified by the BDT algorithm, for the 10m TUG data

Table B.4 expresses the individual sensitivity and specificity of each class (subject group) as well as the average sensitivity and specificity and accuracy of the BDT classifier using the key features feature-set as its input.

Table B.4: Classification results for the BDT classifier using the key features feature-set as its input for classification of the 3 subject groups, for the 3m TUG data

	Class			
	Treatment Patients	**Non-treatment Patients**	**Controls**	**Average**
Sensitivity [%]	91.18	50.00	100.00	80.39
Specificity [%]	82.35	93.33	100.00	91.90

Moreover, Table B.5 compares the accuracy, average sensitivity, and average specificity of the BDT classifier using the key features subset against the same performance measures of the PCA input method.

Table B.5: BDT classifier results, comparing the key features against the PCA selection method inputs for classification of the 3 subject groups, for the 3m TUG data

	10m TUG Data [%]	
	Key Features	**PCA**
Accuracy	88.24 ± 4.23	90.20 ± 3.86
Sensitivity	80.39	81.37
Specificity	91.90	92.54

In spite of the slightly reduced classification performance using the key features as compared to the PCA, the key features set is a subset of the original features and is useful for further analysis and interpretations of these features in the context of discrimination between the treatment patients, non-treatment patients and controls.