Machine Learning for Gait Classification

Xingchen Wang

Universität Bremen 2017

Machine Learning for Gait Classification

Vom Fachbereich für Physik und Elektrotechnik
der Universität Bremen

zur Erlangung des akademischen Grades

Doktor–Ingenieur (Dr.-Ing.)

genehmigte Dissertation

von

M.Sc. Xingchen Wang

wohnhaft in Bremen

Referent: Prof. Dr.-Ing. Axel Gräser
Korreferent: Prof. Dr.-Ing. Udo Frese

Eingereicht am: 28.09.2017
Tag des Promotionskolloquiums: 19.12.2017

Acknowledgements

First of all I would like to thank Prof. Gräser for offering me a research position at the Institute of Automation (IAT) and his valuable support, encouragement, and guidance throughout my research work. I would like to thank also Prof. Frese for accepting to be the second reviewer.

Further I would like to thank Dr. Danijela Ristic-Durrant for her constant support and constructive feedback during my whole research period at the IAT. Also, I would like to thank Ms. Maria Kyrarini for all the brainstorming and discussions for some of my publications.

I would also like to thank Prof. Spranger, for providing all the support and knowledge with his expertise in neurological diseases, as well as for arranging all the experiments with patients.

During my stay at the IAT, I had the opportunity to supervise nine students for their master projects and theses. I would like to thank them for their work on data collection and experiments. Additionally, I would like to thank all the participants of the experiments conducted in this thesis.

My deepest gratitude goes to my parents, for their vital material and spiritual support during the past four years, without whom the thesis cannot be completed.

Kurzfassung

Maschinelles Lernen ist ein mächtiges Werkzeug, um Vorhersagen zu machen und wurde in den letzten Jahrzehnten oft zur Lösung verschiedener Klassifizierungsprobleme eingesetzt. Als eine der wichtigsten Anwendungen des maschinellen Lernens konzentriert sich die Gangart-Klassifikation auf die Unterscheidung verschiedener Gangmuster, indem sie die Qualität des Gangs von Individuen untersucht und kategorisiert. Die am meisten untersuchten Gangmusterklassen sind die normalen Gangmuster von gesunden Menschen, die keine Gangbehinderung durch eine Krankheit oder eine Verletzung haben, und die pathologische Gangart von Patienten mit Krankheiten, die Gangstörungen verursachen, wie z.B. neurodegenerative Erkrankungen (engl. neurodegenerative diseases (NDDs)). Es gab bedeutende Forschungsarbeiten, die versuchten, die Gangart-Klassifizierungsprobleme mit Hilfe fortschrittlicher maschineller Lerntechniken zu lösen, da die Ergebnisse für die frühzeitige Erkennung von NDDs und für die Überwachung des Gangrehabilitationsfortschritts vorteilhaft sein können. Trotz der enormen Entwicklung auf dem Gebiet der Ganganalyse und -klassifizierung gibt es immer noch eine Reihe von Herausforderungen für die weitere Forschung. Eine Herausforderung ist die Optimierung von angewandten Maschinenlernstrategien, um bessere Klassifizierungsergebnisse zu erzielen. Eine weitere Herausforderung besteht darin, Gangart-Klassifizierungsprobleme zu lösen, auch wenn nur begrenzte Daten verfügbar sind. Weiterhin ist eine Herausforderung die Entwicklung von maschinellen Lernmethoden, die präzisere Ergebnisse liefern können, um das Niveau der Gangart oder der Gangstörung zu bewerten, im Gegensatz zu einer einfachen Klassifikation des Gangmusters als gesunder oder pathologischer Gang.

Der Schwerpunkt dieser Arbeit liegt auf der Entwicklung, Umsetzung und Bewertung einer neuartigen und zuverlässigen Lösung für komplexe Gangarten-Klassifizierungsprobleme unter Bewältigung der aktuellen Herausforderungen. Diese Lösung wird als ein Klassifikations-Framework vorgestellt, das auf verschiedene Arten von Gangsignalen angewendet werden kann, wie z. B. den Signalen der Gelenkwinkel der unteren Gliedmaßen, der Beschleunigungen des Rumpfes und der Schrittintervalle. Das entwickelte Framework beinhaltet eine hybride Lösung, die zwei Klassifikatoren kombiniert, um die Klassifizierungsleistung zu verbessern. Um eine große Anzahl von Proben für das Training der Modelle bereitzustellen, wurde eine Methode zur Generierung von Proben entwickelt, das die Gangsignale in kleinere Fragmente segmentieren kann. Die Klassifizierung erfolgt zunächst auf der Stichprobenebene. Anschließend werden die Ergebnisse verwendet, um die Ergebnisse der Subjekt-Ebene mit einem Mehrheitsentscheidungsschema zu generieren. Neben den Klassenbezeichnungen wird ein Vertrauenswert berechnet, um das Niveau der Gangart zu interpretieren.

Um die Gangart-Klassifizierungsleistungen deutlich zu verbessern, werden in dieser Arbeit auch neuartige Merkmalsextraktionsmethoden unter Verwendung statistischer Methoden sowie maschineller Lernansätze vorgeschlagen. Gaußsche Mischverteilungsmodelle (GMM), Regressionen nach der Methode der kleinsten Quadrate und k-nächste Nachbarn (kNN) werden eingesetzt, um zusätzliche signifikante Merkmale bereitzustellen. Vielversprechende Klassifikationsergebnisse werden mit dem vorgeschlagenen Framework und den extrahierten Merkmalen erreicht. Das Framework wird letztlich auf das Management von Patienten und deren Rehabilitationen angewendet und in vielen klinischen Szenarien auf seine Anwendbarkeit hin untersucht, wie die Bewertung der Medikamentenwirkung auf Patienten, die an der Parkinson'schen Krankheit (engl. Parkinson's disease (PD)) leiden, und die langfristige Gangüberwachung von Patienten mit einer hereditären spastischen Paraplegie (HSP) durch Physiotherapie.

Abstract

Machine learning is a powerful tool for making predictions and has been widely used for solving various classification problems in last decades. As one of important applications of machine learning, gait classification focuses on distinguishing different gait patterns by investigating the quality of gait of individuals and categorizing them as belonging to particular classes. The most studied gait pattern classes are the normal gait patterns of healthy people, i.e., gait of people who do not have any gait disability caused by an illness or an injury, and the pathological gait of patients suffering from illnesses which cause gait disorders such as neurodegenerative diseases (NDDs). There has been significant research work trying to solve the gait classification problems using advanced machine learning techniques, as the results may be beneficial for the early detection of underlined NDDs and for the monitoring of the gait rehabilitation progress. Despite the huge development in the field of gait analysis and classification, there are still a number of challenges open to further research. One challenge is the optimization of applied machine learning strategies to achieve better classification results. Another challenge is to solve gait classification problems even in the case when only limited amount of data are available. Further, a challenge is the development of machine learning-based methods that could provide more precise results to evaluate the level of gait quality or gait disorder, in contrast of just classifying gait pattern as belonging to healthy or pathological gait.

The focus of this thesis is on the development, implementation and evaluation of a novel and reliable solution for the complex gait classification problems by addressing the current challenges. This solution is presented as a classification framework that can be applied to different types of gait signals, such as lower-limbs joint angle signals, trunk acceleration signals, and stride interval signals. Developed framework incorporates a hybrid solution which combines two models to enhance the classification performance. In order to provide a large number of samples for training the models, a sample generation method is developed which could segments the gait signals into smaller fragments. Classification is firstly performed on the data sample level, and then the results are utilized to generate the subject-level results using a majority voting scheme. Besides the class labels, a confidence score is computed to interpret the level of gait quality.

In order to significantly improve the gait classification performances, in this thesis a novel feature extraction methods are also proposed using statistical methods, as well as machine learning approaches. Gaussian mixture model (GMM), least square regression, and k-nearest neighbors (kNN) are employed to provide additional significant features. Promising classification results are achieved using the proposed framework and the extracted features. The framework is ultimately applied to the management of patients and their rehabilitation, and is proved to be feasible in many clinical scenarios, such as the evaluation of medication effect on Parkinson's disease (PD) patients' gait, the long-term gait monitoring of the hereditary spastic paraplegia (HSP) patient under physical therapy.

Table of Contents

Kurzfassung i
Abstract iii
1. Introduction 1
1.1 Background 1
1.2 Problem Statement 2
1.3 Motivation 3
1.4 Contributions 4
1.5 Related Publications of the Author 5
1.6 Related Theses and Projects Supervised by the Author 6
1.7 Thesis Overview 6
2. Machine Leaning for Classification 9
2.1 Machine Learning and Classification 9
2.2 Feature Extraction Techniques 13
2.2.1 Fast Fourier Transform 13
2.2.2 Discrete Wavelet Transform 14
2.3 Feature Selection and Dimension Reduction 15
2.3.1 Student's t-test 16
2.3.2 Principle Component Analysis 17
2.4 Supervised Learning Models 19
2.4.1 Artificial Neural Network 19
2.4.2 Support Vector Machine 21
3. Machine Learning for Gait Analysis and Classification 23
3.1 Basic Science of Human Gait 23
3.1.1 Gait and Gait Analysis 23
3.1.2 Gait Cycle and Parameters 25
3.1.3 Normal and Pathological Gait 28
3.2 Gait Measurement Systems 29
3.2.1 Vision-based Systems 30
3.2.2 Wearable Sensor-based Systems 30
3.2.3 Comparison of Gait Measurement Systems 33
3.3 Related Work 34
3.3.1 Current Directions 34
3.3.2 State-of-the-art Approaches and Their Limitations 35
3.4 Novel Machine Learning Framework for Gait Classification 40
3.4.1 Term Explanation 40
3.4.2 Highlights 42
3.5 Conclusions 45
4. Gait Classification for Joint Angle Signals 47
4.1 Related Work 47
4.1.1 Joint Angle Signals 48
4.1.2 State-of-the-art and Limitations 48
4.2 Gait Classification Using Variability Features and Shape Features 51
4.2.1 Data Pre-processing 52
4.2.2 Gait Cycle Pairing 55
4.2.3 Variability Features Extraction 56
4.2.4 Shape Features Extraction 61
4.2.5 Feature Analysis and Classification 64
4.3 Experimental Results 66
4.3.1 Experiment 66
4.3.2 Feature Analysis 67
4.3.3 Classification Results 71
4.4 Applications to Patient Management and Rehabilitation 76
4.4.1 Classification of Simulated Impaired Gait 76
4.4.2 Evaluation of Medication Effect 78

4.4.3 Long-term Gait Monitoring 81
4.4.4 Application to Gait Rehabilitation System 82
4.5 Conclusions 84
5. Gait Classification for Trunk Acceleration Signals 85
5.1 Related Work 85
5.1.1 Trunk Acceleration Signals 86
5.1.2 State-of-the-art and Limitations 86
5.2 Gait Classification Using Time-domain, Frequency-domain and Contour Features 87
5.2.1 Data Pre-processing 88
5.2.2 Sliding Window for Sample Generation 91
5.2.3 Time and Frequency Domain Feature Extraction 92
5.2.4 Contour Features Extraction 95
5.2.5 Feature Analysis and Classification 96
5.3 Experimental Results 97
5.3.1 Experiment 97
5.3.2 Feature Analysis 98
5.3.3 Classification Results 102
5.4 Applications to Patient Management and Rehabilitation 105
5.5 Conclusions 107
6. Gait Classification for Stride Interval Signals 109
6.1 Related Work 109
6.1.1 Stride Interval Signals 109
6.1.2 State-of-the-art and Limitations 111
6.2 Gait Classification Using Statistical and Likelihood Features 112
6.2.1 Data Pre-processing 113
6.2.2 Sliding Window and Sample Generation 113
6.2.3 Statistical Features Extraction 114
6.2.4 Likelihood Features Extraction 115
6.2.5 Feature Analysis and Classification 117
6.3 Experimental Results 119
6.3.1 Database 119
6.3.2 Feature Analysis 119
6.3.3 Classification Results 121
6.4 Conclusions 126
7. Conclusions and Outlook 127
7.1 Thesis Summary 127
7.2 Main Contributions 128
7.3 Outlook 129

Abbreviations 131
References 133
List of Figures 139
List of Tables 141

1. Introduction

This thesis explores the topic of gait classification by developing a machine learning framework for solving various classification problems using different types of gait data. Gait patterns are reflections of the characteristics and quality of human walking, which might be influenced by certain neurodegenerative diseases (NDDs), such as Parkinson's disease (PD). By performing classifications, the gait patterns of different diseases can be distinguished from normal gait patterns for supporting the early detection and rehabilitation of those diseases. As a powerful tool, machine learning (ML) techniques have become popular solutions in the field of gait classification and have been widely utilized by biomedical engineers. The ultimate goal of this thesis is to overcome the technical limitations of previous research work, and to develop a novel classification framework using machine learning and advanced feature extraction techniques. This framework can be used by engineers and clinicians to classify, evaluate, understand, and monitor the gait performances of healthy people, as well as patients suffering from NDDs, by providing additional information obtained from gait classification results.

1.1 Background

Machine learning is a major field of computer science aiming at giving "computers the ability to learn without being explicitly programmed", according to Arthur Samuel who offered this definition in 1959, machine learning explores and studies construction of algorithms that can learn and make predictions on data [1]. Even though machine learning is nowadays applied in a wide range of fields such as bioinformatics, computer vision, and medical diagnosis, to use machine learning effectively is still a challenge. This is because it is usually difficult to find an optimal machine learning model, and there is often insufficient amount of training data available in many practical scenarios.

As one of the most popular and important biomedical research areas, human gait analysis has drawn more and more attentions because it can be used for the early diagnosis and rehabilitation monitoring of related NDDs (e.g. PD and polyneuropathy (PNP)), which may cause severe gait disorders. Similar to most of the biomedical research problems, one major research interest lies in the classification of gait patterns.

Like most pattern recognition problems, gait classification concerns the quantification and interpretation of gait patterns of people, particularly the patients suffering from NDDs. The main purposes and applications of gait analysis and classification are two-folded: the early diagnosis and the rehabilitation monitoring. The early detection aims to predict the probability of incidence of NDDs on healthy people who may potentially

suffer NDDs, and to further prevent them from suffering or progressing with NDDs by assessing the gait quality and comparing with the normal gait patterns; while the rehabilitation monitoring aims to continuously assess the gait performance's changes of patients who have been already diagnosed with NDDs during the rehabilitation period by measuring and evaluating the gait quality and comparing against the healthy reference pattern, or their own walking performances of past medical history.

1.2 Problem Statement

Various machine learning strategies have been applied to the classification of gait patterns as belonging to particular classes, for example "healthy" and "pathological (PT)" (gait pattern impaired by NDDs). Traditional procedures of solving a gait classification problem can be generalized as Fig. 1.1. Given the original gait data, which can be of various types, such as kinematic parameters (e.g., hip and knee joint angles), or kinetic parameters (e.g. ground reaction force (GRF)), a variety of gait features are extracted after pre-processing when necessary. The features are then statistically analyzed, and the significant features are selected and serve as the inputs for the training of the classifier using machine learning. The output of the classification is the predicted label indicating the membership of the test subject to each pre-defined class.

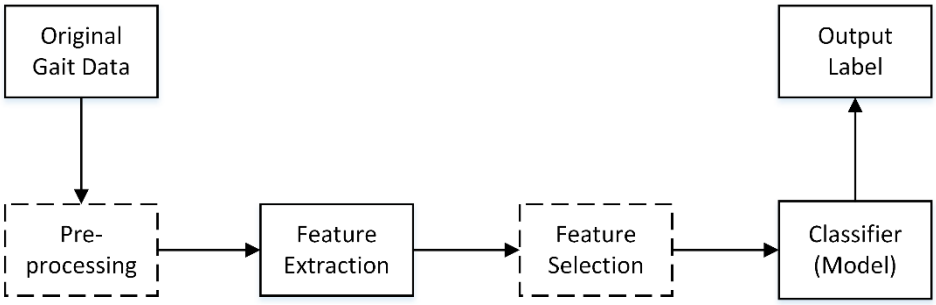


Figure 1.1. Traditional procedures of gait classification.

Several limitations of using machine learning for gait classification are becoming more evident along with the growing needs for more advanced applications and requirements. The state-of-the-art approaches and their shortages are discussed in depth in Section 3.3. The main limitations can be shortly summarized as follows:

1. The classification accuracy needs further improvement. So far most of gait analysis studies are devoted to figure out the most effective classifiers and features for conducting the classification, but rarely try to combine different classifiers or models to achieve higher classification accuracy with the so-called hybrid systems.
2. The number of subjects is usually limited. Like most of studies that focusing on human motion's analysis, the number of subjects involved in the experiments is usually limited, hence, a direct classification on "subject" level

can be infeasible when using machine learning, which in principle requires a large number of samples to train the models.

3. Binary classification result is no longer sufficient. More and more modern applications of gait analysis require a more precise classification outcome with practical significance. In other words, instead of just knowing if the subject's gait is "healthy" or "pathological", it is also necessary to know how "healthy" or "pathological" the gait is, in order to assess the gait quality.
4. Methods depend highly on the type of data. Most previously proposed gait classification methods commonly focus on studying one type of gait data, and may not be suitable for other types of data. Hence, it would be beneficial to develop a general framework based on the basic characteristic of gait, which can be applied to different types of gait data and practical scenarios.

1.3 Motivation

On account of the limitations mentioned above and the rapid advancement of machine learning, this thesis aims at contributing to the field of machine learning based gait classification by proposing a novel solution for solving gait classification problems while overcoming the existing limitations. This solution is presented as a framework which utilizes machine learning as a powerful tool and major component, and develops novel feature extraction algorithms to enhance the classification performances.

The development of the gait classification framework is motivated by the following facts: 1) the gait related signals are semi-periodic due to the semi-periodic behavior of human walking, with one step being considered as a fundamental period. Therefore, it is beneficial to perform gait segmentation to cut long gait signals into shorter incidences, and by doing so, feature extraction and classification are able to be performed on a larger scale of data; 2) the usage of more than one statistical or machine learning models may potentially boost the classification performance; 3) more precise classification results are needed for future applications. Instead of only consider particular discrete values (e.g. 1, 0) as classification output, real-valued number (e.g. 0.13, 0.96) can be more efficient and precise to interpret the level of gait quality; 4) when a walking trial of a subject segmented into a large number of samples, the "level" of gait quality of this trial can be determined by observing the percentage of samples being classified as "heathy"/ "pathological".

Besides the machine learning and feature extraction techniques employed and developed for gait classification, this thesis is also dedicated to broaden the applications fields of gait classification. At the end of chapter 4 and 5, a number of experimental case studies are conducted using data collected from patients suffering from different NDDs, for the monitoring of gait performances, assessment of gait rehabilitation progress, and evaluation of medication effect.

1.4 Contributions

Three types of gait data are studied in this dissertation, which are associated with three different patterns collected from three parts of human body during walking. They are the joint angle signals collected from the sagittal plane of the hip and knee joints, the trunk acceleration signals collected from the back of the waist, and the stride interval signals collected under the feet, respectively. The hip/knee joint angle is one of the most important kinematic parameters that associates with the relative movement of bones, and can reflects the variability (stability) of gait; the trunk acceleration signals are promising representations of walking balance (symmetry) based on the coordination between body parts and their resultant at the center of mass (CoM); and the stride interval signals are good measures of the rhythm of walking in time domain. Those three most important gait data types are studied for classification using the proposed classification framework, and the classification results are applied to four application scenarios for the management of patients and their gait rehabilitation, which are: application in classifying simulated impaired gait; application in evaluating the medication effect on pathological gait; application in long-term monitoring of pathological gait during physical therapy; and the application in patients equipped with robotic rehabilitation system.

With respect to the current directions of machine learning and the state-of-the-art gait classification methods, which is comprehensively discussed in Section 3.3, this thesis contributes to the community of researchers and end-user of machine learning-based gait classifications with the following research achievements:

1. Development of a machine learning based classification framework as a novel gait classification solution. The framework is to have the following advantages:
    - It contains a system which combines two different models to enhance the classification performance. One model is utilized for extracting additional model-fitting features.
    - It is a general framework which can be applied to different types of gait signals, i.e., joint angle signals, trunk acceleration signals, and stride interval signals.
    - It is able to provide an additional confidence score, which can be used as an indicator of the level of gait quality.
    - It is able to yield promising classification results even for a small number of subjects.
    - It contains a post-processing scheme, allowing a more precise classification result.
2. Validation and application of the framework using three types of gait data.
    - Validation and application using hip and knee joint angle signals.
    - Validation and application using trunk acceleration signals.
    - Validation and application using stride interval signals.
3. Improved gait segmentation methods.
    - Improved peak detection-based gait cycle and step segmentation for generating samples (observations) for training and testing of classifiers.

4. Novel sample generation methods.
    - Gait cycle pairing method for sample generation of joint angle signals.
    - Sliding window approach for sample generation of trunk acceleration and stride interval signals.
5. Novel feature extraction algorithms.
    - Machine learning algorithms are utilized for extracting additional features.
    - Distance functions for gait variability features extraction of joint angle signals.
    - Gaussian mixture model (GMM) for shape features extraction of joint angle signals
    - Least square regression method for contour features extraction of trunk acceleration signals.
    - K-nearest neighbors (kNN) for likelihood features extraction of stride interval signals.
6. Post-processing scheme and majority voting (MV).
    - Classification is performed on sample-level, and then the result is used for generating subject-level result with post-processing procedures.
    - MV is utilized to compute the subject-level results in post-processing.
7. Confidence score as an indicator of gait quality.
    - The score is a real-valued number, which can precisely indicate the quality of gait.
8. Practical applications of the framework in patient management and rehabilitation.
    - Experimental study on assessing simulated impaired gait.
    - Case study for monitoring of the medication effect on gait of PD patients.
    - Long-term monitoring of the physical treatment effect on hereditary spastic paraplegia (HSP) patient.
    - Evaluation of gait quality and its changes in subjects equipped with robotic gait rehabilitation system.

1.5 Related Publications of the Author

The following publications have been produced over the course of this thesis, and they are the basis of this thesis.

- S. Natarajan, **X. Wang**, M. Spranger, and A. Gräser, "Reha@home-a vision based markerless gait analysis system for rehabilitation at home", *IEEE 13th IASTED International Conference on Biomedical Engineering*. Innsbruck, Austria, Feb. 2017
- **X. Wang**, D. Ristic-Durrant, M. Spranger and A. Gräser, "Gait Assessment System Based on Novel Gait Variability Measures" *ICORR 2017- 15th IEEE International Conference on Rehabilitation Robotics*, London, 2017.
- M. Kyrarini, S. Naeem, **X. Wang**, and A. Gräser, "Skill Robot Library: Intelligent Path Planning Framework for Object Manipulation" *The 25th European Signal Processing Conference (EUSIPCO)*. 2017.
- J. Shuo, **X. Wang**, M. Kyrarini, and A. Gräser, "A Robust Algorithm for Gait Cycle Segmentation", *The 25th European Signal Processing Conference (EUSIPCO)*. 2017.

- **X.Wang**, D. Ristic-Durrant, M. Spranger, and A. Gräser, "Novel Measure for Gait Quality Monitoring", *Current Directions in Biomedical Engineering*. Vol.3. Issue. s1. 2017.
- **X.Wang**, D. Ristic-Durrant, M. Spranger, and A. Gräser, "Novel Measure for Gait Quality Monitoring", *TAR 2017: Technically Assisted Rehabilitation*, Berlin, Germany. Mar. 2017.
- **X.Wang**, M. Kyrarini, D. Ristic-Durrant, M. Spranger, and A. Graeser, "Monitoring of Gait Performance Using Dynamic Time Warping on IMU-Sensor Data", *IEEE 2016 International Symposium on Medical Measurements and Applications (MeMeA)*, Benevento, Italy, May 2016, pp. 1–6,
- G. Gao, M. Kyrarini, M. Razavi, **X. Wang**, and A. Graeser, "Comparison of Dynamic Vision Sensor-based and IMU-based Systems for Ankle Joint Angle Gait Analysis" ", *IEEE 2016 International Conference on Frontiers of Signal Processing*, Warsaw, Poland, Oct 2016.
- **X. Wang**, O. Kuzmicheva, M. Spranger, and A. Graeser, "Gait feature analysis of polyneuropathy patients", *IEEE 2015 International Symposium on Medical Measurements and Applications (MeMeA)*, Turin, Italy, May 2015, pp. 58–63.
- M. Kyrarini, **X. Wang**, and A. Graeser, "Comparison of vision-based and sensor-based systems for joint angle gait analysis," *IEEE 2015 International Symposium on Medical Measurements and Applications (MeMeA)*, Turin, Italy, May 2015, pp. 375–379.

1.6 Related Theses and Projects Supervised by the Author

Some of the data collection work, as well as some experimental results that are described and discussed within this thesis have been achieved with the valuable support of students who have completed their master thesis or projects under my individual supervision or joint supervision. To acknowledge the vital contribution of those students to the projects behind this dissertation, references to their works are provided here concisely.

- **Vidyani Parataneni**, "Marker-less Human Gait Analysis Using Kinect Xbox One", Master Project, 2017.
- **Shuo Jiang**, "Human Gait Segmentation and Classification Using eButton", Master Project, 2017.
- **Mohamed Shaltout**, "Gait Phase Detection Using Foot Pressure Sensors and Fuzzy Algorithm", Master Project, 2016.
- **Mohsin Latif**, "Building and Evaluation of a Low-Cost IMU-Based System for Human Gait Analysis", Master Project, 2016.
- **Ammar Najjar**, "Gait Classification of Parkinson's Disease Using Machine Learning Algorithms", Master Thesis, 2015.
- **Jishu Chowdhury**, "Development of a Simple Gait Signature based on Walking Bounding Box and Silhouettes", Master Thesis, 2015.
- **Boping Liu**, "A Marker-based Gait Detection System", Master Project, 2015.
- **Jintao Lu**, "Kinematic Abnormality Detection of Human Gait Based on Fuzzy Logic", Master Project, 2015.
- **Qinyuan Fang**, "Features Analysis of Young and Elderly Healthy Gait Patterns Using Machine Leaning Methods", Master Project, 2015.

1.7 Thesis Overview

The thesis outline is as follows: In Chapter 2, the fundamental knowledge of machine learning for classification is described, including the main procedures and the role of machine learning, the most commonly used signal processing and statistical techniques

for feature extraction and feature selection, as well as the basic theory of some machine learning techniques utilized in this thesis.

In Chapter 3, gait classification as one of essential application and research areas of machine learning is considered, and a novel framework is introduced for solving various gait classification problem which may overcome the existing limitations, after discussing the current direction and the state-of-the-art approaches.

In Chapter 4, the proposed classification framework is applied to lower-limbs joint angle signals. An enhanced gait segmentation method for segmenting the trajectories, a novel gait paring method for generating samples, and four distance functions for extracting the variability features are proposed. GMM is employed for generating novel model-fitting features. The effectiveness of the framework and the procedures are validated with an experimental study involving 58 subjects using the LOSO validation. Additionally, four case studies are described at the end of the chapter to prove the feasibility of the framework on management of patients and rehabilitation.

In Chapter 5, the framework is further validated and applied to human trunk acceleration signals for gait balance analysis. In this chapter, a sliding window approach was developed for the sample generation, and novel contour features are extracted using the least square regression method. At the end of the chapter, a clinical case study that shows the feasibility of the framework for monitoring the medication effect in PD patients is described.

Chapter 6 deals with validation and application of the framework on stride interval signals. The sliding window approach and kNN are utilized for sample generation and machine learning features extraction respectively. Multiclass classification is performed using two strategies.

Chapter 7 summarizes the whole thesis with major findings and points out future directions of research.

2. Machine Leaning for Classification

2.1 Machine Learning and Classification

Machine learning (ML) is a core branch of artificial intelligence that systematically applies algorithms to synthesize the underlying relationship among data and information [2]. ML has already very broad application in web search, stock market prediction, behavior analysis, big data analytics, image processing and more areas. The computational role of machine learning is to generalize the experience trained from examples in order to output an estimated target function or model, so to characterize relationship within large array of data for various problems. One important goal of machine learning model is to accurately predict the correct categories of data for unseen instances. The generalization process requires classifiers to output class labels using discrete or continuous feature vectors or matrixes as input.

The goal of ML is to predict the class memberships of unknown events or scenarios based on past experiences, which is in other words to solve classification or pattern recognition problems. The learning process is essential in generalizing the classification problems by modelling on historical experiences in the form of training dataset, and aims at achieving accurate results on new data and unseen tasks in a form a testing dataset. Some key terminologies are explained below.

Classifier

A classifier is a method that can process a new input sample as an unlabeled instance of a feature vector, and outputs the label of a class to which it belongs. Most of the commonly used classifiers utilize the probability measures (statistical inference) to categorize the optimal label for an input sample.

Confusion Matrix

Confusion matrix is a matrix that visualizes the overall performance of a classifier or a classification algorithm. The matrix shows the results with the predicted classification labels against the actual classification labels in a form of several key measures, such as accuracy (Acc), true positive rate/sensitivity (TPR/Sen), true negative rate/specificity (TNR/Spe), positive predictive value/precision (PPV/Pre), negative predictive value (NPV), and area under the curve (AUC). Illustration of a standard confusion matrix for a binary classification can be seen below.

		Actual Condition	
		Positive	Negative
Predicted Condition	Positive	TP	FP
	Negative	FG	TN

TP = Ture Positive, sample correctly predicted as positive.
TN = Ture Negative, sample correctly predicted as negative.
FP = False Positive, sample incorrectly predicted as positive.
FN = False Negative, sample incorrectly predicted as negative.
Table 2.1 Confusion matrix of a binary classification.

Based on those concepts, several metrics are defined to measure the performance of the classification algorithm more precisely. The formulas for the calculations are listed below.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

$$TPR = Sen = \frac{TP}{TP + FN} \tag{2.2}$$

$$TNR = Spe = \frac{TN}{TN + FP} \tag{2.3}$$

$$PPV = Pre = \frac{TP}{TP + FP} \tag{2.4}$$

$$NPV = \frac{TN}{TN + FN} \tag{2.5}$$

Accuracy (Acc), the rate of correct predictions, is the most important measure of the classification performance, and is commonly estimated from an independent test dataset that was totally unused during the learning/training process. For a dataset contains a limited number of samples, cross-validation are commonly used. Besides, the area under the receiver operating characteristic (ROC) curve (AUC) is also often used as an essential term for measuring the diagnostic ability of a binary classifier as its discrimination threshold is varied. The ROC curve is created by plotting the TPR against the FPR at various threshold settings. The AUC is calculated as the accumulated area under the ROC curve, and can be used for investigating and comparing ML models. The value of the AUC usually lies between 0.5 and 1.0. A value that is between 0.5 and 0.6 is considered as an presentation of a poor classier, while a value lies between 0.9 and 1.0 is regarded as an indicator of an excellent classifier.

Feature Matrix

In machine learning, a feature is an individual measurable property of characteristic of phenomenon being observed [3]. Wisely extract and choose informative, discriminative and independent features are essential steps in performing classification. Efficient extraction and selection of features is known as feature engineering, which requires the

full understanding of the characteristics of data being handled and the comprehensive knowledge of the signal processing and data analytic methods. A collection of features can be called a feature set, which is a subset of the entire feature set being extracted. The feature set can be often formed into a feature matrix for the ease of learning processes, with each row represents a sample/observation and each column represent a certain feature.

Validation

Validation methods are verification techniques that evaluate the generalization ability of a trained classifier/model for new unseen test dataset. Cross-validation is the most commonly employed validation method, of which the k-fold cross-validation and Holdout validation are the two major approaches. For the k-fold cross-validation method, the whole dataset is arbitrarily partitioned into k subsets of equal size; the model is trained for k times, where each iteration uses one of the k subset for testing and the remaining $k-1$ subsets for training. The final accuracy is computed as the average of the k iterations. For the Holdout validation, the dataset is randomly partitioned into training set and test set with a predefined proportion. The size of each of the sets is arbitrary although typically the training set is larger than the test set. The final results are usually aggregated from multiple runs.

Supervised Learning

Supervised learning is a learning mechanism that infers the underlying relationship between the observations and the target class labels that is subject to prediction. Distinguished from unsupervised learning, which are designed to unfold the hidden structures in unlabeled datasets, in which the desired output class labels are unknown, the supervised learning utilizes the labeled training data to synthesize the model functions which aims to generalize the relationship between the feature matrix and the labeled output. The feature matrix and labeled output jointly influence the direction and magnitude of change during training in order to improve the overall performance of the function model by minimizing the error between the desired labels and the real output labels [2]. Overfitting and underfitting are two commonly seen phenomenon in models that are not promisingly trained, where overfitting happens when a model learns the detail and noise in the training set to the extent that it negatively impacts the performance on new data, while underfitting refers to a model that can neither model the training data nor generalize to new data.

The normal process of developing supervised ML algorithms can be decomposed into 6 steps:

1. **Data acquisition**. This step acquires the valuable data that shall be used in the machine learning-based classification problem. Since the quality and amount of data highly influences the performance of classification, it is important to consider the most advance measurement systems for data collection, especially for human motion related classifications, where the events of motions can be easily disturbed by noise.

2. **Pre-processing**. The pre-processing steps manipulate on signals to obtain the signals in required form. Formatting, cleaning, sampling, and normalization are common techniques performed on raw data. Formatting step presents the data in a useable format; cleaning generates smoothed, noise-removed data; sampling outputs the resampled data at regular or adaptive intervals in a manner such that redundancy is minimized without losing important information; and normalization brings data from different dimensions into the same scale [2].
3. **Feature extraction**. This process starts from the initial set of preprocessed raw data and builds derived features intended to be informative and non-redundant, aiming at facilitating the subsequent learning process, and in some cases leading to better human interpretations [4].
4. **Feature selection**. It is a process of selecting a subset of relevant features for the model construction. Four main reasons of performing feature selection are: for the simplification of models to make them easier to interpret; for shortening the training time; for avoiding the curse of dimensionality; and for enhancing the generalization by reducing overfitting.
5. **Train the algorithm**. Select the training and test set from the whole dataset of features, train the algorithm using the corresponding machine learning approach and validate the model.
6. **Test the algorithm**. Evaluate the algorithm to test its effectiveness and performance on new dataset. If the performance of the trained model needs improvement, repeat the previous steps by changing the data streams, tuning the learning configurations, parameters or kernels methods to reach better results [2].

The high-level flow of the supervised learning can be seen below.

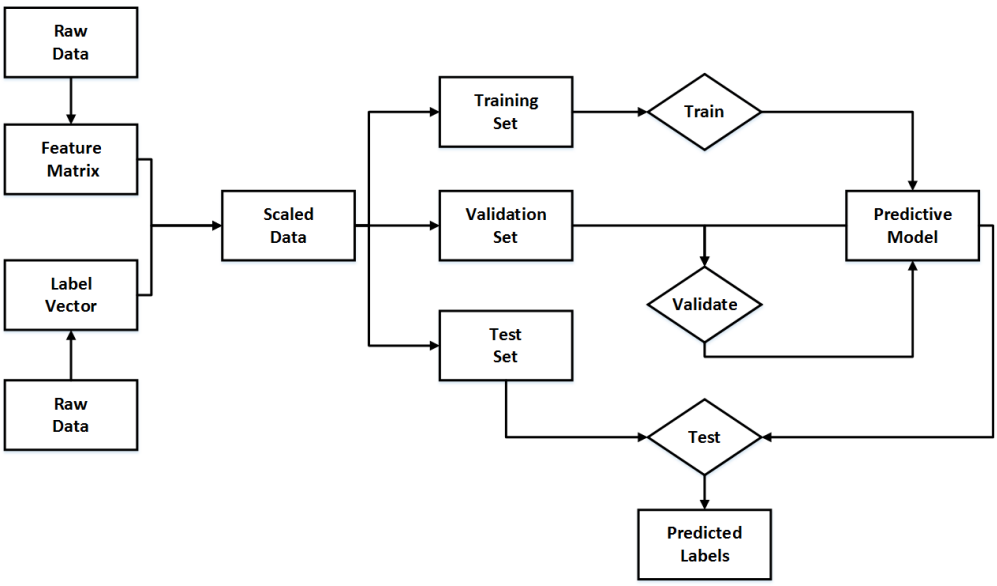


Figure 2.1. High-level flow of supervised learning.

2.2 Feature Extraction Techniques

Feature extraction is a core procedure in the processing of signals for classification. For signals, such as sensor signals collected for human motion, it is important to evaluate the characteristic of the signals by measuring the statistical aspects in the temporal domain to investigate the peaks, cross-correlation, standard deviation (SD), etc., or transforming the signals into frequency domain to evaluate the bandwidth, spectral distribution, energy, power, and distortions.

Common signal processing technique for extracting features from human motion related signals can be divided into two categories, namely, statistical methods and transforming methods. The statistical methods, such as SD and root-mean-square-deviation (RMSD), compute the statistical distribution or fluctuation of the signals which could reflect the temporal characteristic of the signal. Those methods are the most common feature extraction approaches in signal processing based classifications (e.g., [5] [6]). Other statistical method, such as cross-correlation, which convolutes two signals to measure the similarities, is mainly used as a good measure of variability and continuity of signals (e.g., [7] [8]). Transforming methods, such as Fast Fourier transform (FFT) and Discrete wavelet transform (DWT), transform the signal into frequency domain, and analyze the distribution of the signals over each frequency band (e.g. [9] [10]).

2.2.1 Fast Fourier Transform

A periodical function can be decomposed into the Fourier series, which are the frequencies it consists of:

$$s_{\mathrm{N}}(x) = \frac{A_0}{2} + \sum_{n=1}^{N} A_n \cdot \sin\left(\frac{2\pi \mathrm{n} x}{P} + \Phi_n\right), \text{for integer } \mathrm{N} \geq 1 \tag{2.6}$$

And when the period of the function is seen as very large, the non-periodical function can be transformed in the similar way to get their frequency form, which is the Fourier transform:

$$\hat{f}(\varepsilon) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x\, \varepsilon} dx \tag{2.7}$$

The spectrum of a periodic function is a discrete set of frequencies, while for a non-periodic signal a continuous spectrum is produced from the Fourier transform. Fourier transforms are used widely in digital technology, it is an extremely powerful mathematical tool that allows the signals to be viewed in a different domain, inside which difficult problems can be analyzed in a simple way.

The discrete Fourier transform (DFT) is defined for the discrete signals, which converts a discrete function with finite squally spaced samples from its original domain to the frequency domain. The DFT result is a list of coefficients of a finite combination of complex sinusoids, ordered by their frequencies, that has the same sample values as the original function.

$$X_k \overset{\text{def}}{=} \sum_{n=0}^{N-1} x_n \cdot \mathrm{e}^{-\frac{i2\pi \mathrm{kn}}{\mathrm{N}}}, \quad k \in \mathbb{Z} \tag{2.8}$$

FFT is the fast algorithm to compute the DFT and its inverse, and it has been widely implemented for many engineering applications. The FFT varies due to different fast algorithms. Many FFT algorithms depend on the fact that $e^{-\frac{i2\pi}{N}}$ is an N^{th} primitive root of unity, and thus analogous transforms can be applied.

2.2.2 Discrete Wavelet Transform

The Fourier transform converts a signal from time domain to frequency domain. From the Fourier transform we can get the frequency components of a signal and their coefficients, but we cannot know when these frequency components occur. For a stationary signal, we do not need the information of the instant a frequency component arises, because the process is stationary and all frequency components are constant and do not change when time shifts. But for non-stationary signals (the joint probability distribution of the signal changes as time shifting), Fourier transform loses information when they are applied.

Short time Fourier transform (STFT) solves the problem with non-stationary signals to some extent. The signal is divided into sections and each section is analyzed with Fourier transform. It is like to apply a sliding window on the signal, and the signal in each window is analyzed independently for frequency content. Because the window's size is constant for all frequencies in the STFT, the resolution of the analysis in the time-frequency domain is always the same (equally spaced). The selection of the most appropriate window size is an essential issue.

Unlike STFT, wavelets transform (WT) [11] provides a multi-resolution solution. Generally, the wavelet transform can be expressed by the following equation:

$$\mathrm{X(a,b)} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \overline{\Psi\left(\frac{t-b}{a}\right)} x(t) dt \tag{2.9}$$

$\Psi(t)$ is the mother wavelet, and we can see that the wavelet transform is the convolution of the signal and a wavelet basis function, which is obtained by dilations and translations of the mother wavelet. The mother wavelet is a kind of function which fulfills some special conditions, like it is time-limited and its average is 0. For instance, Haar wavelet, Meyer wavelet, Morlet wavelet are popular wavelets.

After the sampling and a series of processing, the discrete form of the wavelet transform, discrete wavelet transform (DWT) can be acquired. The fast DWT algorithm was also conducted, and a one-level DWT is realized as follows:

$$\mathrm{y}_{low}[n] = \sum_{k=-\infty}^{\infty} x[k] h[2n-k] \tag{2.10}$$

$$\mathrm{y}_{high}[n] = \sum_{k=-\infty}^{\infty} x[k] g[2n-k] \tag{2.11}$$

The samples are decomposed through a low pass filter with impulse response $g[n]$, and through a high-pass filter $h[n]$ simultaneously. This decomposition makes the time resolution half, since each filter output characterizes only half of the signal. But the frequency resolution has been doubled. The process of a 3-level DWT is shown in Fig.2.2.

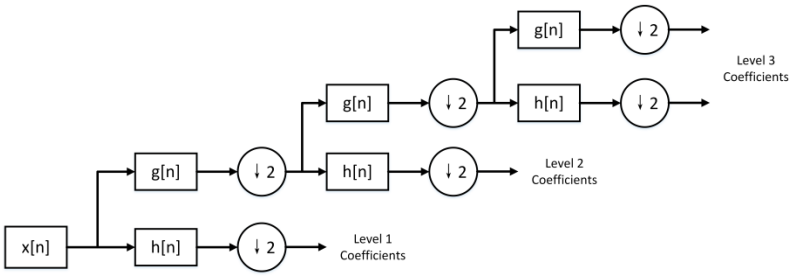


Figure 2.2. A 3-level DWT.

2.3 Feature Selection and Dimension Reduction

Both feature selection and dimensionality reduction deal with features and seek to reduce the number of attributes in the dataset. Different from dimensionality reduction methods, which reduce the number of attributes by creating new combinations of attributes, feature selection approaches include and exclude attributes present in the dataset without changing them.

The feature selection methods are devoted to automatically select the attributes in the dataset that are the most relevant to the predictive modelling problem that we are solving. Feature selection techniques commonly acts as a filter, muting out features that are unneeded, irrelevant and redundant. The feature selection algorithms can be generally divided into three categories: filter methods, wrapper methods and embedded methods. Filter methods apply a statistical measure to rank the features using a scoring to each feature. The filter methods include, for example, the student's t-test, Chi squared test, information gain and correlation coefficient scores. Wrapper methods treat the selection of a set of features as a search problem, where different combinations of features are evaluated and compared to each other. The search process can be a best-first search, or a stochastic algorithm, or a heuristic method, such a recursive features elimination algorithm. Embedded methods, such as Elastic Net and Ridge Regression, try to figure out the features that contribute to the accuracy of the model the best while it is being created.

The dimensionality reduction, or dimension reduction, in machine learning is the process of reducing the number of random variables by eliminating the dimensions that are more likely to be noise. Dimension reduction methods usually transform the dataset in the high-dimensional space to a space of lower dimensions. The main advantages of applying dimensionality reduction techniques in machine learning based classification are as follows: firstly, it reduces the cost of time and storage space; secondly, it usually improves the performance of the machine learning model by removing multi-collinearity; thirdly, it eases the visualization of the data when reducing them into a very low

dimension such as 2D or 3D. The most popular dimension reduction methods are principle component analysis (PCA), linear discriminant analysis (LDA) and generalized discriminant analysis (GDA).

2.3.1 Student's t-test

A t-test is a statistical hypothesis test which the statistics of the test follows a Student's t-distribution null hypothesis. The t-test is widely applied when the test statistically follows a normal distribution and it is usually used to determine if two data sets are significantly different from each other.

Two-sample t-test is one of the most frequently used t-tests, which hypnotized that the means of the two populations are equal. Different from the one-sample t-test, by which the statistical difference between a sample mean and a known or hypothesized value of the mean in the population, the two-sample t-test tries to compare the means of two different samples. Based on our application, the two-sample t-test is more suitable since the data comes from two categories (classes).

The statistic of the two-sample t-test can be expressed as:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \tag{2.12}$$

where $\bar{x}$ and $\bar{y}$ are the means of the two classes, s_x and s_y are their standard deviation, and n and m are their size.

If the case that the two data sets are assumed to come from the population with equal variances, the test statistic under the null hypothesis has Student's t-distribution is replaced by the pooled standard deviation:

$$s = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} \tag{2.13}$$

When the two data sets are not assumed to be from the populations with equal variances, the test statistic under the null hypothesis follows an approximate Student's t-distribution with a number of degrees of freedom given by Satterthwaite's approximation, which is also called Welch's t-test.

Two main output of the t-test are the hypothesis test result and the p-value, denoted as h and p respectively. The h is a logical value: if $h = 1$, this indicates the rejection of the null hypothesis at the Alpha significance level; if $h = 0$, this indicates a failure to reject the null hypothesis at the Alpha significance level. The p-value of the test, returned as a scalar value ranged between 0 and 1, can be found using a table of values from Student's t-distribution. If the yielded p-value is below the threshold chosen for statistical significance, then the null hypothesis is rejected in favor of the alternative hypothesis. The significance level is usually chosen as 0.10, 0.05, 0.01, or 0.001.

2.3.2 Principle Component Analysis

The PCA is a popular and useful linear transformation technique that is commonly used as feature extraction and selection methods in classification. It converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principle components using an orthogonal transformation, the number of which is less or equal to the number of original variables. The transformation is defined in the following way: the first principle component contains the largest possible variance, and each succeeding component in turn has the largest variance with the constraint that it is orthogonal to the preceding principle components. The results obtained after the transformation are uncorrelated orthogonal basis sets (vectors) [12].

PCA can be performed by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix. The PCA approach can be summarized as follows [13]:

1. Standardize the data.
2. Obtain the Eigenvectors and Eigenvalues from the covariance matrix or correlation matrix, or perform singular vector decomposition.
3. Sort eigenvalues in descending order and choose the k eigenvectors that correspond to the k largest eigenvalues, where k is the number of dimensions of the new feature subspace.
4. Construct the projection matrix from the selected k eigenvectors.
5. Transform the original dataset via the projection matrix to obtain a k-dimensional feature subspace.

Suppose that there is a random vector X, with p variables,

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \tag{2.14}$$

with a variance-covariance matrix

$$var(X) = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix} \tag{2.15}$$

We consider the linear combinations:

$$\begin{aligned} Y_1 &= e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p \\ Y_2 &= e_{21}X_1 + e_{22}X_2 + \cdots + e_{2p}X_p \\ &\vdots \\ Y_p &= e_{p1}X_1 + e_{p2}X_2 + \cdots + e_{pp}X_p \end{aligned} \tag{2.16}$$

And each of these combination can be regarded as a linear regression which predict Y_i from $X_1, X_2, \ldots, X_p$. $e_{i1}, e_{i2}, \ldots, e_{ip}$ can be considered as the regression coefficients.

Y_i is random since it is a function of our random data, and its variance can be computed as :

$$var(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = e_i'\Sigma e_i \tag{2.17}$$

Besides, Y_i and Y_j will have a population covariance

$$cov(Y_i, Y_j) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{jl}\sigma_{kl} = e_i'\Sigma e_j \tag{2.18}$$

where e_{ij} are collected in to a vector as:

$$e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{ip} \end{pmatrix} \tag{2.19}$$

The first principle component is the linear combination of x-variables that has maximum variance among all the linear combinations, therefore it accounts for as much variation as possible of the whole data set. Formally speaking, the first principle component Y_1 selects $e_{11}, e_{12}, \dots, e_{1p}$ that maximizes

$$var(Y_1) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{1l}\sigma_{kl} = e_1'\Sigma e_1 \tag{2.20}$$

subject to constraint that

$$e_1'e_1 = \sum_{j=1}^{p} e_{1j}^2 = 1 \tag{2.21}$$

The second principal component Y_2 is the linear combination of x-variables that accounts for as much of the remaining variation as possible, under the constraint that the correlation between the first and the second component is 0. It is therefore decided as selecting the appropriate $e_{12}, e_{22}, \dots, e_{2p}$ that maximizes

$$var(Y_2) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{2k}e_{2l}\sigma_{kl} = e_2'\Sigma e_2 \tag{2.22}$$

subject to the constraint that

$$e_2'e_2 = \sum_{j=1}^{p} e_{2j}^2 = 1 \tag{2.23}$$

along with an extra constraint that it is uncorrelated with the first principle component

$$cov(Y_1, Y_2) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{2l}\sigma_{kl} = e_1'\Sigma e_2 = 0 \tag{2.24}$$

Similarly, all the subsequent principle components have the same property, namely, they are linear combination that account for as much of the remaining variation as possible and they are uncorrelated with each other.

For instance, the i^{th} principle component Y_i is determined by selecting the $e_{11}, e_{12}, \dots, e_{1p}$ that maximizes

$$var(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = e_i^{'}\Sigma e_i \tag{2.25}$$

subject to the constraint that sums of squared coefficients add up to one, along with the additional constraint that this component is uncorrelated with all the previously determined components.

$$\begin{aligned} e_i^{'}e_i &= \sum_{j=1}^{p} e_{ij}^2 = 1 \\ cov(Y_1, Y_i) &= \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{il}\sigma_{kl} = e_1^{'}\Sigma e_i = 0 \\ cov(Y_2, Y_i) &= \sum_{k=1}^{p}\sum_{l=1}^{p} e_{2k}e_{il}\sigma_{kl} = e_2^{'}\Sigma e_i = 0 \\ &\vdots \\ cov(Y_{i-1}, Y_i) &= \sum_{k=1}^{p}\sum_{l=1}^{p} e_{i-1,k}e_{il}\sigma_{kl} = e_{i-1}^{'}\Sigma e_i = 0 \end{aligned} \tag{2.26}$$

2.4 Supervised Learning Models

The most popular machine learning methods can be grouped using their similarities as follows: regression algorithms, such as simple logistic regression (SLR); instance-based algorithms, such as k-nearest neighbor (kNN); regularization algorithms, such as Elastic net; decision tree algorithms, such as classification and regression tree (CART); Bayesian algorithms, such as Naïve Bayes, Bayesian network; clustering algorithms, such as k-means; artificial neural network algorithms, such as perceptron, back-propagation; deep learning algorithms; ensemble algorithms, such as AdaBoost, random forest; and other algorithms, such as support vector machine (SVM).

2.4.1 Artificial Neural Network

Artificial neural network (ANN) is a statistical machine learning model used for data mining and classification purposes, like decision making and pattern recognition. In the recent years, ANNs were very prevalent as methodologies for gait analysis. There are

different kinds of networks, but they all have the basic components: a set of nodes, and connections between nodes. The nodes work as computing units. They process the inputs and produce the outputs. And the connections are used to transfer the information flow between the nodes.

ANN is a kind of network inspired by biological neural network. It sees the nodes as artificial neurons and each neuron is a computational model highly extracted from the natural neurons. The inputs are firstly multiplied by weights before computed by the mathematical function, and the function determines the activation of the neuron.

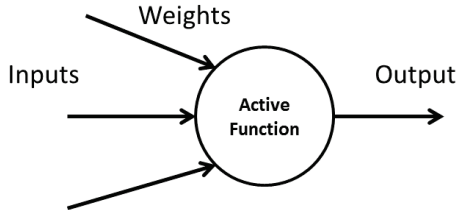


Figure 2.3. An artificial neuron.

An ANN model consists of inputs, some hidden layers and an output layer. The ANN architectures can be grouped into two categories, they are feed-forward networks without loops and feedback networks with feedback loops.

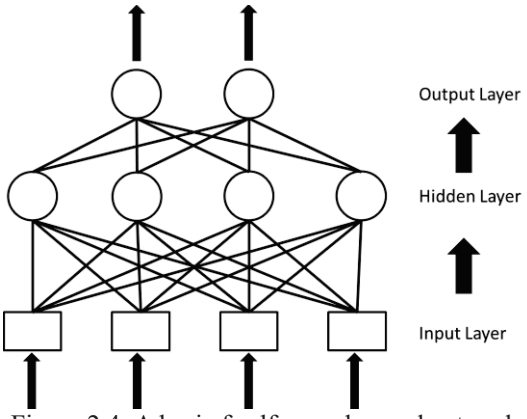


Figure 2.4. A basic feedforward neural network.

In order to get the expected outputs, the model should be constructed with appropriate weights. The process of adjusting the weights is called learning. One of the most commonly used ANN models is the backpropagation (BP) algorithm, on which many other algorithms are based. The BP algorithm is used in feed-forward ANNs, in which the neurons send their signals forward and propagate the errors backwards. It uses supervised learning with examples to reduce the error until it is minimal.

The activation function of the artificial neurons implemented BP algorithm is a weighted sum, the sum of the inputs x_i multiplied by their weights w_{ji}:

$$A_j(\overline{x}, \overline{w}) = \sum_{i=0}^{n} x_i w_{ji} \tag{2.27}$$

And the most common output function is a sigmoidal function:

$$O_j(\overline{x}, \overline{w}) = \frac{1}{1 + e^{A(\overline{x}, \overline{w})}} \tag{2.28}$$

The goal of the training process is to obtain a desired output when certain inputs are given. The error is the difference between the actual output and the desired outputs. Weights need to be adjusted in order to minimize the error. From that we can define the error function:

$$E_j(\overline{x}, \overline{w}, d) = (O_j(\overline{x}, \overline{w}) - d_j)^2 \tag{2.29}$$

The BP algorithm calculates how the error depends on the output, inputs and weights. After this, the weights will be adjusted using the method of gradient descendent.

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \tag{2.30}$$

In the next steps, we compute the derivative of E in respect to O_j, then the derivative of O_j on the weights. The final weights are adjusted according to:

$$\triangle w_{ji} = -2\eta(O_j - d_j)O_j(1 - O_j)x_i \tag{2.31}$$

The above derivations are for the ANN with two layers (one hidden layer); however, if we want to add another layer, we can follow the same procedure, with calculating the error depends on the inputs and weights of the previous layer. The indexes should be adjusted carefully since each layer can have a different number of neurons. For practical reasons, ANNs implementing the BP algorithm usually do not have many layers, since the training time of the network grows exponentially.

2.4.2 Support Vector Machine

Support Vector Machines (SVM) is also a well-known powerful technique of machine learning for classification and regression problems. The SVM algorithm is based on the statistical learning theory and the Vapnik-Chervonenkis dimension introduced by Vladimir Vapnik and Alexey Chervonenkis [14] [15]. Its main idea is to map the data to a usually high dimensional space (by means of kernel functions) and to make the classification in this space through the construction of a linear separating hyperplane. The data vectors near to the hyperplane are called support vectors and the method to determine the optimal hyperplane is to maximize the soft margin, which is the distance to the nearest cleanly split examples in order to split the examples as precisely as possible.

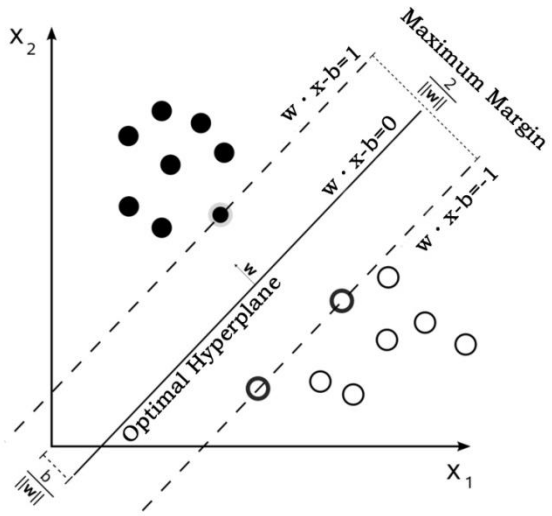


Figure 2.5. Illustration of a separating hyperplane with a maximum margin by SVM.

Assume that we have n training data $x_1, x_2 \ldots x_n \in R$, so $\mathbf{x_i}$ is a vector of features, and their outputs are $y_1, y_2 \ldots y_n$, with $y \in \{1, -1\}$. Define the hyperplane H such that:

$$\begin{aligned} x_i \cdot w &\geq 1, when\ y_i = +1 \\ x_i \cdot w &\leq 1, when\ y_i = -1 \end{aligned} \tag{2.32}$$

And make $x_i \cdot w = 0$ indicate the points on hyperplane. The distance between the two critical hyperplanes is $\frac{2}{||w||}$, so we want to minimize $||w||$. We use a function to do the mapping from input space to some higher dimensional space, which is denoted by

$$\Phi(\mathbf{x}): \mathbf{x} \subset \Re^K \rightarrow \Re^M, K \ll M \tag{2.33}$$

So we get the separating hyperplane of the following universal approximation from

$$f(\mathbf{x}) = \sum_{i=1}^{m} w_i \Phi_i(\mathbf{x}) + b \tag{2.34}$$

As said before, we maximize the soft margin of the hyperplane in order to obtain the best classification performance. The optimal hyperplane can be calculated by solving the dual Lagrangian optimization problem as follows:

$$\mathfrak{T}(w, b, a) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j K(x_i, x_j) \tag{2.35}$$

where a_i are lagrangian multipliers, and the nonlinear function $\Phi(\mathbf{x})$ can be applied by using a kernel function defined as

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = < \Phi(\boldsymbol{x_i}), \Phi(\boldsymbol{x_j}) > \tag{2.36}$$

There are different types of Kernel functions, for example, the Linear Kernel, Polynomial Kernel, Gaussian Radial Basis Function (RBF) Kernel, and so on. However, the calculation in a high dimensional space can be extremely complicated. The Mercer's theorem makes it possible to compute the inner product of vectors in the low dimensional space implicitly, but the results indicate the classification in the high dimensional space.

3. Machine Learning for Gait Analysis and Classification

Machine learning has a broad range of applications in biomedical engineering, whereby the biomedical issues involving large data sets and complex mathematical context are solved promisingly by advanced artificial intelligence techniques, especially the machine learning methods. Gait analysis, the systematic study of human walking behavior, has been rapidly advancing in the sense of both clinical findings and engineering breakthroughs. The findings on gait analysis have been beneficial to millions of patients who are suffering from diseases that cause gait impairment. In this chapter, gait analysis is introduced as an application of machine learning. To understand the work presented in this thesis, a brief overview on the basic science underlying gait and clinical gait analysis is given in Section 3.1, consisting of the definition of gait, gait cycle, and common gait parameters, and the characteristics of normal and pathological gait. Afterwards, gait measurement systems are shortly reviewed in Section 3.2, including vision-based gait measurement systems for the capturing of joint angle trajectories, and the wearable sensor-based gait measurement system for measuring gait joint angles and trunk accelerations. In Section 3.3, the current directions and state-of-the-art machine learning techniques employed in previous gait analysis and classification are explicitly reviewed and summarized, emphasizing their contributions and limitations. At the end of this chapter, the proposed novel machine learning framework for gait classification is explained in depth in Section 3.4. This chapter is necessarily laid ahead of Chapter 4, 5 and 6 for the better explanation of the validation and application of the proposed framework to solving different gait classification problems.

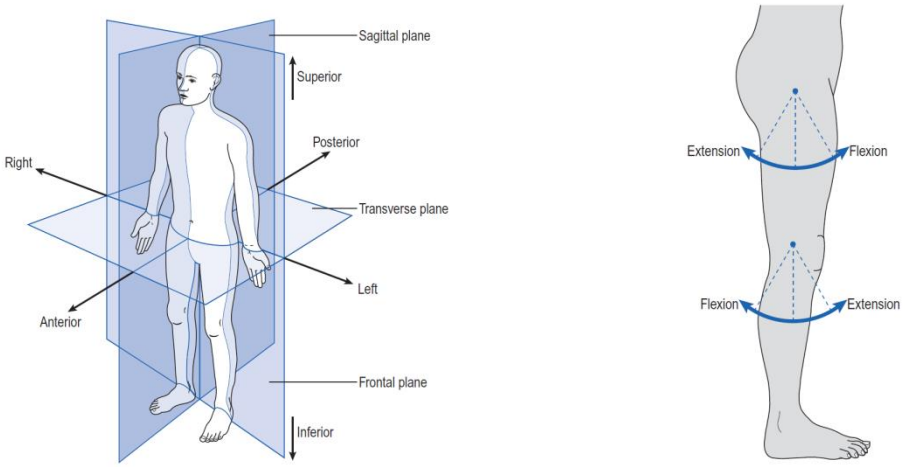
3.1 Basic Science of Human Gait

3.1.1 Gait and Gait Analysis

Human gait is a locomotion achieved through the voluntary movement of the lower limbs. It is defined as bipedal, biphasic forward propulsion of the center of gravity of the human body, and results from a complicated process involving the brain, spinal cord, peripheral nerves, muscles, bones and joints. The behavior of gait involves three scientific disciplines: anatomy, physiology and biomechanics. The systematic analysis of this walking behavior is called gait analysis [16]. One of the most important applications of gait analysis is in the assessment of gait quality for supporting the diagnosis and rehabilitation of related gait disorder, such as NDDs. Therefore it is attracting more and more attentions from

researchers in the fields of physical therapy, biomedical engineering, neurology and rehabilitation engineering.

In order to understand the walking behavior from anatomical perspective, it is important to know the terms that describe the directions and relationships between parts of the body. When a person is standing upright with the feet together, arms by the sides and palms forward, six terms are utilized to represent directions with respect to the center of body: 1) the umbilicus is *anterior*; 2) the buttocks are *posterior*; 3) the head is *superior*; 4) the feet are *inferior*; 5) *Left* is self-evident; and 6) *Right* is self-evident. Furthermore, six terms are used to describe the relationships within a single body part: 1) *Medial* means towards the middle of the body; 2) *Lateral* means away from the midline of the body; 3) *Proximal* means towards the rest of the body; 4) *Distal* means away from the rest of the body; 5) *Superficial* structures are close to the surface; and 6) *Deep* structures are far from the surface. Additionally, the motion of the limbs is described in three planes: 1) A *sagittal* plane is any plane that divides the body into right and left position; 2) A *frontal* plane divides a body part into front and back portions; 3) A *transverse* plane divides a body part into upper and lower portions. The illustration of all the used terms with respect to the anatomical position can be seen in Fig. 3.1 a).



a) b)

Figure 3.1. a) The human anatomical position. b) The movement of hip and knee joints in sagittal plane. (Image source: Chapter 1, Fig.1.1, Fig. 1.2, Whittle, 2007. [16])

The bones of the pelvis and legs are the most involved bones during walking. The pelvis is commonly regarded as a single rigid structure during gait analysis. The femur is the longest bone in the body, which connects the hip and knee joints. The patella or kneecap is a sesamoid bone with an important mechanical function of displacing the quadriceps tendon forwards, so to improve its leverage. The tibia is the bone connecting the knee joint and the ankle joint, and fibula is the bone next to tibia on its lateral side.

The point where one bone is in contact with another is called a joint. Most of the joints that walking involves can only move in one or two of the planes. The movements of joints taking place in the sagittal plane are named flexion and extension; the movements taking place in the frontal plane are abduction and adduction; the movements taking place in transverse plane are internal and external rotation. The hip joint is the only ball-and-socket

joint in the human body and is able to perform all the six movements in the three planes. The knee joint is capable of performing flexion and extension, with a small amount of internal and external rotation. The ankle joint has significant motion only in dorsiflexion and plantarflexion direction, which correspond to flexion and extension in other joints. Since the biggest movements of the hip, knee and ankle joints take place in sagittal plane, the presented thesis takes only the joint movement in sagittal plane into consideration. The definitions of the movements of the hip and knee joints in the sagittal plane are depicted in Fig. 3.1 b).

In addition to bones and joints, there are a large number of muscles and tendons that play a role in moving the joints and maintaining the posture during walking under the support and coordination of the spinal cord, spinal nerves and peripheral nerves. Further anatomical and physiology science on human gait is beyond the scope of this thesis and is therefore not discussed.

3.1.2 Gait Cycle and Parameters

Prior to elaborating on the mathematical description and analysis of gait, it is necessary to name the most important terminologies and parameters commonly used in gait analysis.

Modern gait analysis dates back to the early 20th century, when new techniques were first adapted to capturing human motion pictures. Then in the 1960s the introduction of hip replacement arthroplasty by John Charnley in the UK motived the measurement of hip joint force by John Paul in Glasgow. By the early 1980s, the traditional cine film began to be replaced with many modern commercial systems used today, such as *Vicon* and *Coda* in the UK, and *Motion Analysis Corporation* in the US. Nowadays, the computer games industry is refining the motion capture techniques and providing better visualization and animation techniques integrated with complex mathematical models and simulations. It must be stressed here that, since the 1960s more serious attempts have been taken out of laboratory and applied in the clinic, and from the 1980s gait analysis has been more and more widely adapted as an effective tool for the management of walking disorders in patients. The main limitations on how to obtain high-quality gait data have been overcome to a great extent, and more attention has been paid on understanding the gait data in order to benefit the patients in a more meaningful way.

The gait cycle is the most basic concept of clinical gait analysis. One gait cycle, or one stride, is defined as the time interval between two successive incidences of initial contact, which is defined as the event when one foot just starts making contact with the ground. Initial contact is also often called heel contact in the literatures owing to the fact that the initial contact is performed by striking the ground using the heel in normal gait. One gait cycle can be divided into stance phase and swing phase depending on the status of one interested leg, where stance phase is the period when the foot is in contact with the ground, and swing phase is the period when the foot is in the air moving forward. When considering the status of the other leg, the stance phase of one cycle performed with one leg is divided into four sub-phases: loading response, mid-stance, terminal-stance and pre-swing; the swing phase is divided into three sub-phases: initial swing, mid-swing and terminal swing. A summary of all the phases in one gait cycles is shown in Table 3.1.

	Sub-phase	Start	End	Interval (% of gait cycle)
Stance phase	Loading response	Initial contact (IC)	Opposite toe off (OT)	0-10%
	Mid-stance	Opposite toe off	Heel rise (HR)	10-30%
	Terminal stance	Heel rise	Opposite initial contact (OI)	30-50%
	Pre-swing	Opposite initial contact	Toe off (TO)	50-60%
Swing phase	Initial swing	Toe off	Feet adjacent (FA)	60-73%
	Mid-swing	Feet adjacent	Tibia vertical (TV)	73-87%
	Terminal swing	Tibia vertical	Initial contact (IC)	87-100%

Table 3.1. Definition of sub-phases in one gait cycle and their intervals.

Gait performance is commonly evaluated with four types of measures in clinical and biomechanical studies: temporal-spatial parameters, kinematic parameters, kinetic parameters and electrical parameters. Those measures provide an intuitive insight into the walking subject's level of gait function and have been widely used for gait assessment. In this thesis, the focus lies on the processing of temporal-spatial and kinematic parameters.

Temporal-spatial parameters

Temporal-spatial measures examine the global aspects of gait measured in time and length. Owing to the fact that gait is considered to be a semi-periodic behavior, and that gait cycles are very similar to each other for steady walking, those parameters are often calculated as means by averaging the measures collected from multiple gait cycles. Those measures are:

- *Stride time*: the time duration of a complete gait cycle
- *Stance/swing time*: the time duration of the stance/swing phase
- *Single support time*: the time duration during which one foot has contact with the ground in one gait cycle
- *Double support time*: the time duration during which two feet have contact with the ground in one gait cycle
- *Stride length*: the distance between two successive placements of the same foot. It consists of two *step lengths*, left and right, each of which is the distance the corresponding foot moves forward in front of the other one.
- *Cadence:* the number of steps taken in a given time; cadence is usually measured in steps per minute.
- *Speed*: distance covered by the whole body in a given time; speed is measured in meters per second (m/s).

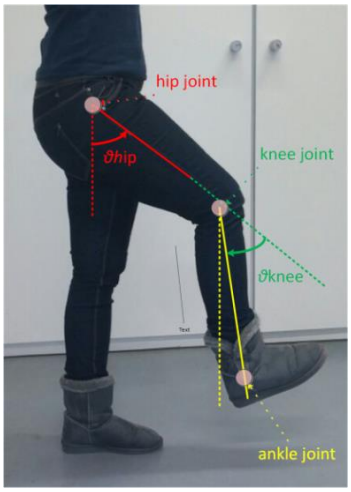
Kinematic parameters

The kinematic measures mainly refer to the joint kinematics of the hip, knee, and ankle. It is important to know how the hip, knee and ankle joint angles are defined before examining diagrams of those angles. The hip angle is defined as the angle between the thigh and the vertical axis, assuming that when the walking subject is standing straight at zero-position, the thigh is in parallel with the vertical axis. The hip angle is positive when the hip joint is in a flexion state and negative when the hip joint is in an extension state.

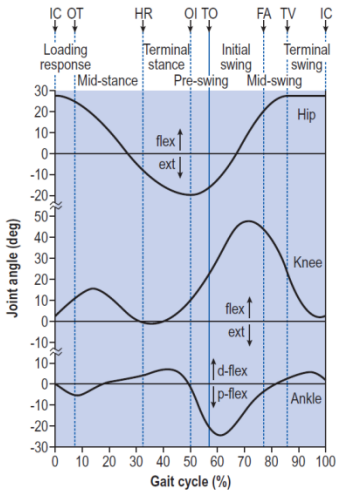
The knee angle is defined as the angle between the femur and the tibia. The ankle angle is defined as the angle between the tibia and an arbitrary line in the foot. The ankle angle is normally around 90° with small fluctuation, while the hip and knee angles undergo more significant changes during walking; therefore, in this thesis, only the hip and knee joints are considered. The definitions of hip and knee angles are illustrated in Fig. 3.2 a).

The hip angle for a normal gait cycle starts at about 30° of flexion at initial contact; then , it steadily extends towards a neutral position in the mid-stance phase, and reaches an approximately 15° of extension in the terminal stance. The direction reverses in the pre-swing phase and rapidly reaches a flexion angle of about 30° in the initial and mid-swing phases. The hip holds this angle for a short period of time, and then, it extends slightly to the same position as in initial contact. The knee is in about 5° of flexion at initial contact, and then continues to flex, reaching a peak position of about 20° of flexion. Then, it begins extending in the loading response and mid-stance phases. In the terminal stance phase the knee angle reaches a minimum peak of about 3° of flexion, then the motion is reversed, and the knee begins flexing in the terminal stance and pre-swing phases. For the swing phase, the flexion reaches its peak of about 50°; then, the motion reverses again, and the knee starts to extend, reaching the neutral position at the end of the cycle. The hip and knee joint angle trajectories for one normal gait cycle are depicted in Fig. 3.2 b).

Besides the joint angles, the accelerations of certain joints or the trunk are also important parameters that may reflect walking variability. Wearable sensors, such as the inertial measurement unit (IMU) or the attitude and heading reference system (AHRS) are attached on specific body parts and the movements of these parts in three dimensions are investigated in order to determine the dynamics of the gait patterns.



(a)



(b)

Figure 3.2. a) Definition of hip and knee angles. b) Normal hip, knee and ankle angle trajectories in one complete gait cycle (image source: Chapter 2. Fig.2.5, Whittle, 2007, [16]).

Kinetic parameters

Kinetic measures refer to the joint torques, also called joint moments, which are the efforts devoted to changing the velocity of the segments. The net torque is the sum of four factors measured at the joint, and they are as follows: the external torque results from the reaction of the ground with the foot, known as the ground reaction force (GRF); the gravitational torque at the center of mass (COM) caused by gravitational field; the intersegmental torque results from the angular movement of the other segments involved; the muscle torques results from the combined action of all musculoskeletal elements acting at the joint [16].

Electrical parameters

The electrical parameters, such as the electromyogram (EMG), is a measure of the muscle activity during gait, and is usually collected using electrodes. It can reveal nerve dysfunction, muscle dysfunction or problems with nerve-to-muscle signal transmission.

A summary of the four basic types of gait parameters can be found in Table 3.2.

Data Type	Major Gait Parameters
Temporal-spatial	stride(stance/swing) intervals
Kinematic	hip/knee joint angles; trunk/feet acceleration
Kinetic	joint moment, GRF
Electrical	EMG

Table 3.2. Summary of four types of gait parameters.

3.1.3 Normal and Pathological Gait

A person with a normal gait should be able to maintain the balance and rhythms of walking, while performing stable and continuous forward motions of the body. To achieve a normal gait, the cooperation of brain, spinal cord, and the whole sensory system is required. As some of the neurological diseases may have direct effect on those systems, changes in gait may occur. Gait abnormalities usually occur in 8–19% of elderly population, in 14% of individuals over 65 years, and in over 50% of individuals aged over 85 [17].

The gait changes normally with aging. Those changes are reflected mainly in four aspects, i.e., the strength, the walking posture, the limb motion, and the joint motion. Major declines in strength can often be observed after the fifth decade, and they become much faster afterwards. Healthy elderly people tend to walk with a larger toe-out angle, and a slower walking speed, especially after the age of 70. The changes in limb motion are mainly seen from the slight increase in the double support time. Joint motion changes can be mainly observed from the reduction in ankle plantar flexion during the late stance phase.

Unlike the normal changes caused by aging, neurodegenerative diseases change gait differently. Pathological gaits are usually divided into neuromuscular or musculoskeletal

etiologies, caused by different musculoskeletal, neurological, or weakness related problems. Depending on the causes and the characteristics of gait, abnormal gait can be divided into several types. A short summary of the pathological gait patterns that are concerned and highlighted in this thesis, along with their causes and characteristics, is listed in Table 3.3.

Gait Pattern	Common Causes	Characteristics
Parkinsonian gait	Parkinson's disease (PD)	• Bradykinesia, tremor, and rigidity • Shuffling steps • Difficulty in initiating steps
Scissor gait	Cerebral palsy (CP)	• Bilateral involvement and spasticity in all extremities • Narrow base, dragging both legs and scraping the toes • Adductor tightness may cause legs to scissor
Steppage gait	Polyneuropathy (PNP)	• Foot drop • Unsteady gait • High-lifted legs
Hemiplegic Gait	Stroke	• Unilateral weakness • Leg dragging • Loss of normal arm swing and slight circumduction
Waddling gait	Multiple Sclerosis (MS)	• Swaying, symmetric, wide-based gait • Toe walking
Choreic gait	Huntington's disease (HD)	• Irregular, jerky, and involuntary movements in all extremities • Baseline movement disorder • Wide-based gait with slow leg raising and simultaneous knee flexion

Table 3.3. Common pathological gait patterns [18] [19] [16] [20] [21].

3.2 Gait Measurement Systems

Gait parameters, are commonly measured by four types of measurement systems: the vision-based system, such as RGBD cameras; the wearable sensor-based system, such as IMUs; the goniometers, which mainly measures the joint angles; and the force and pressure sensors, such as underfoot pressure sensors and pressure mattress.

The electro-goniometer is a useful tool for measuring angles precisely and has been recently widely considered as the gold standard system for joint angle measurements [22]. The pressure and force based measurement systems are mainly used for the detection of stride-related temporal parameters, such as the moments of the heel strike and toe off. In

this section, the two most important gait measurement systems, vision-based and wearable sensor-based systems are discussed.

3.2.1 Vision-based Systems

Vision-based systems for measuring the kinematic parameter, such as the joint angles, can be divided into marker-based and markerless systems. Many of the marker-based systems, such as Vicon [23], utilize four to eight cameras in a restricted indoor environment and detect the location of active markers attached on pre-defined positions on the body. Those systems are usually quite precise, but are comparably more expensive, and they require larger space, and longer setup time. Some other marker-based systems, such as Kinovea [24], utilize passive markers and detect the position of the markers with color-based or shape-based segmentation algorithms. Those systems provide less costly solution; however, the precision of the measurement depends highly on the correct placement of the markers, and the view angle of the camera.

Owing to the drawbacks of the marker-based systems, the markerless vision-based systems have become popular in recent years. Reha@home [25] is a robust markerless gait analysis system developed by the Institute of Automation, University of Bremen, Germany. This system includes a RGBD camera and associated software for image processing. During the recording of data, the subjects are requested to walk in front of the camera from left to right or the other way around. The RGB images along with the depth information are saved and passed to the algorithms for extracting the hip and knee joint angles. The detection algorithm starts by extracting the contour and silhouette of the walking subject in the defined region of interest (ROI) from each frame, and then detects the head, hip, knee and ankle joints according to the proportion ratios of the human body. The hip and knee joint angles are calculated using the detected joint locations in the image, and smoothed with a cascade of filters in the next step. The hip and knee joint angles are computed using the detected hip, knee, and ankle joint based on the kinematic relationship of those joints. This system is considered to be one of the state-of-the-art, low-cost, markerless gait analysis systems.

3.2.2 Wearable Sensor-based Systems

Besides the Reha@home and Kinovea, another low-cost gait measurement solution is the wearable sensor-based systems, which utilizes multiple IMUs to measure the gait parameters from predefined body locations. The IMUs are able to measure the joint angles, as well the accelerations of the trunk using their electronic components. A 6-axis IMU can measure the acceleration using accelerometers and the angular velocity using gyroscopes. Some IMUs have magnetometers in three axes. The magnetometer measures the vector of the earth's magnetic field and uses the measurement as a reference to calibrate against drifts in orientation. Some IMUs have additional outputs, such as roll, pitch, and heading (yaw) angles and they are called vertical reference units (VRUs) if the

heading angle is low-drift unreferenced, or attitude and heading reference systems (AHRSs) if the heading angle is accurate. Table 3.4 shows a comparison of a few IMUs with different functionalities, different number of axis and different accuracies of orientation.

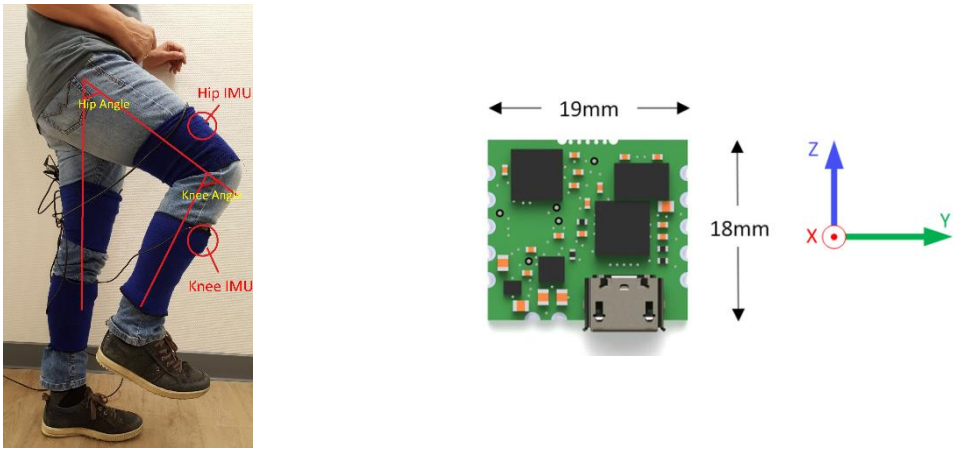
Model	**Manufacturer**	**Accelerometer**	**Gyroscope**	**Magnetometer**	**onboard processing**	**Functionality**	**Typical attitude and heading accuracy**	
		#	#	#			Pitch and Roll	Heading
Navigation IMU	AIMS	3	3	-	no	IMU		
Navigation VRU	AIMS	3	3	-	yes	VRU	Static accuracy: ±0.1° / Dynamic accuracy: ±0.2°	
MTi-100 IMU	Xsens	3	3	3	no	IMU		
MTi-200 VRU	Xsens	3	3	3	yes	VRU	Static accuracy: ±0.2° / Dynamic accuracy: ±0.3°	
MTi-300 AHRS	Xsens	3	3	3	yes	AHRS	Static accuracy: ±0.2° / Dynamic accuracy: ±0.3°	Accuracy: ±1.0°
FSM-9	Hillcrest labs	3	3	3	yes	AHRS	Static accuracy: ±1.5° / Dynamic accuracy: ±1.5°	Static accuracy: ±1.0° / Dynamic accuracy: ±1.0°

Table 3.4. Comparison between different IMU models [26] [27] [28].

Different studies have proposed different numbers of IMU for measuring the gait joint angles [29] [30] [31]. The method can be low-cost depending on the number of sensors and on the model of the IMU. Additionally the accuracy of the measurements also depends on the model of the IMU. The motion sensor that has been chosen for measurement in this thesis is the FSM-9 AHRS module, with has on-board data fusion and is capable of providing inclinations in three axes.

For the purpose of recording the gait joint angle signals, which are one of the three types of signals this thesis concerns itself with, four FMS-9 units (each with dimensions $1 \times 1 \times 0.25$ inch and a weight of 2.8gm), are attached on the lower limbs of the subjects,

two on the frontal upper legs and two on the frontal lower legs. Using the inclinations provided by the sensors, the hip and knee joint angles are computed based on the kinematic correlations between the placement of sensors and the definition of the joint angles. The standards setup of the AHRSs, as well the dimensions and orientations of the units, is illustrated in Fig 3.3.



(a) (b)

Figure 3.3. (a) Setup of AHRSs on subject for the measurement of the joint angles; (b) Dimensions and orientations of the FSM-9 unit.

In addition to the joint angle signals, the trunk acceleration signals are also one important data type studied in this thesis since they are optimal representation of gait balance and stability. The quantification of gait quality via AHRS, particularly the accelerometers, has become popular in the recent years owing to their high accuracy, easy setup, and low cost [32] [33]. As the weight of the trunk is over 50% of the weight of the whole body, and the movement of the trunk is highly prioritized by the central nervous system [34], measurements of trunk sway movement have been shown to be very sensitive to disease-related gait changes [35]. The acceleration of trunk is widely considered to be a presentation of the motion of center of mass (CoM), and has been proposed as a comprehensive indicator of walking motor control [36].

Based on the information provided above, the FSM-9 unit is utilized in this thesis to measure trunk acceleration signals during overground walking as well. During the experiments, the sensor unit is attached on the back of the waist of the subject, at the L3 segment of the lumbar spine (e.g., [37] [38]). The three axes of the unit correspond to the three most important axes of the anatomy dimensions: the x-axis corresponds to the Medio-Lateral (ML) axis, the y-axis corresponds to the Anterior-posterior (AP) axis, and the z-axis corresponds to the vertical (V) axis. The definition of the three anatomy axes can be seen in Fig. 3.1. The setup of the FSM-9 unit is illustrated in Fig.3.4. In order to stabilize the unit so that the influences of cloth or skin can be minimized as much as possible, a belt is bound around the waist of the subject.

(a) (b) (c)

Figure 3.4. Experimental setup to measure trunk acceleration. a) Placement of FSM-9 AHRS; b) Subject wears a belt for stabilization; c) Overground walking for gait measurement.

3.2.3 Comparison of Gait Measurement Systems

In order to compare the performances of different types of gait measurement systems on capturing joint angle signals, the author has carried out an experimental study by collecting and comparing the joint angle data simultaneously from four measurement systems: goniometer as the gold standard; the AHRSs (FMS-9) as the motion sensor-based system; the Reha@hom system as the markerless vision-based system; and, the Kinovea as the marker-based system [39].

The four systems were equipped on one leg of a healthy subject at the same time, and the test subject was requested to walk 2–3 complete cycles several times within the working rage of the camera, and the data are recorded simultaneously from AHRSs and camera.

In principle, the AHRS-based system achieved a better performance than the vision-based systems compared with the goniometer system, which is also mainly composed of MEMS, while the vision-based measurements had a larger lag and larger fluctuations. Besides, the vision-based systems usually work within a limited space, while motion sensor-based systems do not have such restrictions. On the basis of the experimental results, we chose to acquire the joint angle and trunk acceleration signals using the wearable sensor-based systems. The experiments conducted in Chapter 4 and 5 utilize the setups introduced in this chapter. More details about the comparison experiment can be found in [39].

3.3 Related Work

3.3.1 Current Directions

Modern computer based gait analysis started in the 1970s in several hospital based research labs, before undergoing commercial development in the 1980s [40]. Most of the famous techniques developed or employed for modern gait analysis before the 21st century have been very comprehensively summarized by Tom Chau in his reviews [41] [42]. Chau has pointed out that the main technical challenges for gait analysis and classification are as follows: high-dimensionality, meaning that the gait data set may consist of variables from kinematic, kinetic, temporal-spatial, and electrical dimensions; temporal dependence, meaning that the gait data always tend to have semi-periodic characteristic; high variability, meaning that gait data contains inter-subject, intra-subject, within-trial, and between trial variability; and correlation between dimensions, meaning that the gait signals are not always independent from each other, but tend to have certain correlations, and such correlations are usually nonlinear. In addition to identifying these challenges, Chau reviewed carefully the existing studies and stressed five most important techniques, i.e., fuzzy logic, statistical analysis, fractal method, ANN, and wavelet analysis. The advantage of the fuzzy method is that it can group data from multiple dimensions and measure gait improvement objectively; however, it is not suitable for dealing with gait data that has not been parameterized into a proper set of features. Statistical approaches, such as PCA, factor analysis, and multiple correspondence analyses, are believed to be very beneficial for gait analysis, according to Chau, because they can reveal clinically relevant information that would be difficult to interpret from the original variables and are able to identify significant gait parameters. However, those methods cannot uncover the localized structure of data. The fractal methods, such as de-trended fluctuation analysis, and relative dispersion analysis, are able to reveal new pattern. However, those methods require a very large amount of data for realization and therefore are not suitable for all types of gait signals. In the second part of his review, Chau stressed the importance of using machine learning techniques, particularly, ANN, which can model the relationship between gait features regardless of the high complexity of the features.

A more modern review on gait analysis and classification using computational intelligence was done by Lai et al [6] in 2009, when machine learning has been in rapid development. In his work, he summarized most of the gait signal processing techniques, such as FFT and DWT, the gait feature selection approaches, such as PCA, and most importantly, he stressed the importance of gait classification and machine learning. He was the first to point out potential usage of hybrid machine learning systems for gait classification, which are combination of several learning machines for achieving better performance. Owing to Chau's predictions and proposal, the hybrid system has been utilized in this thesis to improve the classification performances, and this will be discussed later in Section 3.4. Eventually, Lai highlighted a few additional challenges and future directions of gait classification, such as predicting diseases progressions, the development of more advanced sensor technologies, and the monitoring of gait rehabilitation of patients with NDDs.

After 2010, as successfully predicted by Lai, both machine learning and sensor technology became more and more essential in gait classification and analysis. More recent reviews by Muro-de-la-Herran in 2014 [43], Ao in 2015 [44], Chen in 2016 [45], and Caldas in 2017 [46] have suggested the advanced wearable sensors as main gait capturing devices, and machine learning techniques as dominate gait analysis approaches. Muro-de-la-Herran investigated all the gait related studies published between 2012 and 2013, and found that over 60% of studies were using wearable systems for gait analysis and classification [43]. Ao has further confirmed the broad usage of wearable sensors and machine learning in gait analysis, fall detection, sleep monitoring and diseases diagnosis [44]. Chen reviewed all the related studies using wearable sensors in gait analysis and concluded his review with an overall summary of gait parameters, and machine learning applications with a focus on gait speed estimation [45]. Caldas made an all-round search of publications on machine learning based gait analysis between 1968 and 2016, and concluded that the machine learning methods significantly support gait analysis [46]. According to the reviews by Chen [45] and Caldas [46], the current direction of gait analysis and classification can be generalized as follows:

1. Gait data types: the most studied gait data types are kinematic data, kinetic data, and temporal-spatial data. Kinematic data, such as hip/knee joint angle signals and trunk accelerations, correspond to pathological gait patterns such as the crouch gait, drop foot, stiff knee, and excessive hip sway. Kinetic data, such as GRF and moment of joints, correspond to gait characteristics of sensory gait, and the weakness during the toe-off moment. The temporal-spatial gait data, such as stance/swing time, correspond to difficulty in certain gait phases. (Table 4, Chen 2016)
2. Gait measurement systems: wearable sensors have shown a greater potential than the conventional systems, such as vision-based and force plates system. (Table 1, Chen 2016)
3. Pathological gait patterns: the most studied gait pathologies are NDDs such as multiple sclerosis (MS), Parkinson's disease (PD), Huntington's disease (HD), polyneuropathy (PNP), amyotrophic lateral sclerosis (ALS), and Alzheimer's disease (AD). (Table 2, Chen 2016).
4. Machine learning in gait classification: after studying over 20 machine learning based gait classification studies, Caldas confirmed the significant contribution of machine learning to the current, as well as the future direction of gait analysis (Table 3. Caldas, 2017). Constructively, Caldas has pointed out that gait classification has so far no standardized method or framework, making the large amount of previous studies difficult to compare (Section 3.2, Caldas, 2017).

3.3.2 State-of-the-art Approaches and Their Limitations

In order to understand the most advanced approaches for machine learning-based gait classification in depth, a comprehensive investigation of the state-of-the-art approaches from the most authoritative journals articles and conferences proceedings was conducted.

Most of the cited publications were published in high-ranking journals within the past five years. A summary of those publications can be found in Table 3.5, where the publications are listed with the studied gait patterns, data types, and extracted features, as well as the number of total subjects, employed machine learning classifiers, and the best results achieved.

First, most (20 out of 26) of the listed publications dealt with the classification of pathological gait patterns caused by NDD and the healthy gait, and only a few studied other patterns, such as the faller and non-faller [47]. Some studies focused on the classification of one NDD versus the healthy pattern, for instance, [48] developed a SVM-based strategy to distinguish the PD pattern from the "healthy control (HC)" pattern, while some studies tried to classify different pathological patterns, such as discussed in [49]. Bilgin conducted classification on ALS vs. PD as well, and all the problems were interpreted as binary classification problems. In some studies like [50], different diseases were grouped together as the "NDD group", and classification was done in order to distinguish the "NDD" from the "HC" class, while in other studies like [51], a multiclass classification was carried out for five different gait patterns using various classifiers.

Second, almost all the major types of data have been investigated, with nine studies focusing on kinematic data ([52] [53] [54] [55] [8] [56] [57] [58] [59]), 13 studies focusing on temporal-spatial parameters ([48] [51] [5] [50] [60] [61] [62] [63] [64] [65] [66] [9] [59]), and 8 studies focusing on kinetic data ([6] [61] [10] [57] [67] [49] [59] [68]). In some of these studies, several types of data were jointly utilized for processing and classification (e.g. [56] [61] [59]). The most investigated kinematic parameters are the joint angle signals and acceleration/angular velocity signals collected from certain body parts like the trunk, shank and feet. The most evaluated kinetic parameter is the GRF. The kinematic and temporal-spatial parameters are the most studied data types because they are more intuitive than the kinetic parameters, easier to collect, and can be interpreted more directly.

Regarding the features extracted for gait classification, different types of features were considered depending on the type of gait data. For example, the most utilized features for kinematic data are statistical features, time and frequency domain features, and variability-based and symmetry-based features. Statistical features consist of maximums, minimums, means, standard deviations (SD), and root-mean-square-deviations (RMSD), as utilized in [54], [55], and [59]. Time and frequency domain features are usually parameters generated using signal processing techniques, such as FFT coefficients (e.g. [54]), power spectral density (PSD) (e.g. [53]), correlation coefficients (e.g. [53] [54] [8]), and spectrum (e.g. [8]). Most of the variability and symmetry related features are in general also based on correlation (e.g. [53] [54]) or basic statistical methods (e.g. [55]). The most utilized features for temporal-spatial gait data are statistical features and wavelet features. Statistical features, such as maximums, minimums, means, standard deviations, skewness, Lempel-Ziv complexity, fuzzy entropy and Teager-Kaiser energy (e.g. [5] [64]), are widely applied. Wavelet features such as the approximation and detailed coefficients were utilized by [48] and [9]. For the kinetic data, similar to the temporal-spatial parameters, the most popular features were extracted using statistical methods or wavelet methods. Statistical parameters, such as peak values at certain

important gait incidences were utilized in [6], while wavelet parameters including the approximation and detailed coefficients were computed as dominate features in [10] [49]. Some other studies use gait phase related features for classification [58].

The number of subjects who participate in the experiments is usually between 10 and 166, with a minimum number of 10 reported in [67] and a maximum 166 reported in [10]. The average number of subjects included in the mentioned studies was 53. PhysioBank [69] is a well-known database that is widely used in gait analysis, as it has a large archive of physiological data and contains gait and balance database. The gait and balance database is composed of nine different sub-databases, such as gait dynamics database, which is a collection of stride interval signals from 64 healthy and NDD patients and was utilized in [5], [50], [65], [64], and [9], and, a PD gait database, which is a collection of multichannel recordings from force sensors placed beneath the feet of 93 PD patients, and 73 HC subjects, and was utilized in [10]. A majority of the other studies used data collected in their own experiments.

Of all the classifiers, the SVM was the most frequently used classifier, with 10 studies employing it as one major machine learning strategy. ANN was the second frequently used classifier, with nine out of 26 publications employing it, followed by Naïve Bayes (NB) and kNN, with five publications employing each. Other famous techniques are Random Forest (RF), Decision Tree (DT) and Simple Logistic Regression (SLR) in order of popularity.

In the majority of the highlighted studies, accuracy, namely, the percentage of subjects correctly predicted as their labelled classes, was reported as the final results. In some other studies results were reported as the sensitivity and Area Under the Curve (AUC) of the classifiers. Except for [8], which mainly deals with the classification of different freeze of gait (FOG) levels of PD patients, as well as [58], which classifies different walking patterns such as walking upstairs, walking downstairs, of the remaining studies that reported accuracy as the main result, the best accuracy was achieved in [67], in which an accuracy of 98% was achieved for classifying NDD and HC patterns from foot switch signals. The worst results were seen in [54], in which an accuracy of 70% was achieved for distinguishing AD and HC patterns using feet acceleration signals. The average accuracy computed from all the studies that reported the accuracy as the final result is 89.6%. However, to the best of the author's knowledge, no classification studies have been conducted for exactly the same groups of gait patterns using the same types of data and the same machine learning methods. Therefore, it is unreasonable to compare the results simply by looking at the reported accuracy. The results, however, should always be evaluated from different perspectives, such as the novelty of the work, the quality of the gait features, and the utilization and optimization of machine learning algorithms.

Even though the state-of-the-art studies have provided a very comprehensive and bright picture for gait classification solutions, there are still many limitations and unanswered questions:

1. Lack of a general framework. As can be seen from the previous discussions, each study worked with one or two specific types of data, and the methods for gait classifications differ from each other to a large extent, from both feature

extraction, and machine learning perspectives. This situation might be difficult for evaluating the solutions and comparing the performances. This issue has been pointed out by Caldas in [46], and so far no researcher proposed a general machine learning framework that could overcome this problem and is applicable to different data types. This framework would take into consideration of the most important characteristics of gait signals of all types.

2. Binary results are no longer sufficient. Most of the previous studies tried to answer one question using machine learning and signal processing methods, which is, whether the subject was healthy or pathological, or whether the subject had one disease as opposed to another. Owing to the rapid advances in technology and clinician skills, a single binary (Healthy or Patient) answer is no longer sufficient. This question, however, can be answered in many cases intuitively from professional medical diagnostic results. Nowadays, instead of only a yes-or-no classification of gait, more and more research efforts are being put towards assessing the level of illness, monitoring the progress of rehabilitation of gait disorders, and evaluating the effect of certain medications or physical therapy. All in all, it is not enough to provide only a binary label for a subject's gait, and a more precise, detailed, numerical metric is needed to further evaluate the degree of illness or healthiness of gait.
3. Limited number of subjects. The number of subject recruited for gait classification studies has not been very large compared with the number of features extracted. As gait data collection is usually time consuming, the number of subjects is commonly quite limited. Performing classification directly at the subject level cannot be ideal since the number of training and testing samples will be very limited, while the number of samples required for conducting an ideal classification is usually ten times larger than the number of features. Therefore, it would be beneficial to generate as many samples as possible for classification, especially when the number of subjects and walking trials are limited.
4. A combination of different machine learning techniques or learning models is in demand. Various machine learning techniques have been validated for use in gait analysis and have achieved promising results. Currently, it is not enough to simply compare each machine learning approach and determine the best one. It is more important to design novel and advanced frameworks to unblock the bottleneck and promote more advanced and efficient strategies. The potential usage of combined models, which aim to combine different machine learning techniques or learning models in order to achieve a better performance, for gait classification problems was first mentioned in [6] by Lai. This strategy might be an optimal solution, as it brings the advantages of different classifiers together and may realize a higher performance.

Publication	Gait Patterns	Data Type	Features	# Subjects	Classifiers	Best Results
Loroche [52], 2014	HO; HC	joint angle	spatial angles; trunk & pelvis motion	40	SVM	Acc= 90%
Yang [53], 2012	CRPS; HC	trunk acceleration	temporal; frequency; symmetry	20	ANN	Acc= 85.7%
Wang [54], 2014	AD; HC	feet acceleration	statistical; frequency; symmetry	30	ANN	Acc= 70.00%
Joshi [48], 2017	PD; HC	stride intervals	wavelet	31	SVM	Acc= 90.32%
Greene [55], 2015	MS; HC	shank angular velocity	statistical; temporal; variability	38	SLR	Acc= 96.90%
Pradhan [51], 2015	NDD; HC	temporal-spatial	statistical	150	kNN; ANN; SVM;NB	Sen=92.0%
Djurić-Jovicić [8], 2014	PD FOG levels	shank angular velocity	spectrum; temporal-spatial; correlation	12	Rule-based	Acc= 78-100%
Lai [56], 2009	PFPS; HC	GRF; kinematic	statistical	27	SVM	Acc= 85.15%
Xia [5], 2015	NDD; HC	stride intervals	statistical	64	SVM; RF; ANN; kNN	Acc≈ 100%
Zeng [50], 2015	NDD; HC	stride intervals	gait dynamics	64	ANN	Acc= 93.75%
Wahid [60], 2015	PD; HC	temporal-spatial	statistical	49	KFD;NB;RF; kNN;SVM	Acc= 92.6%
Pauk [61], 2016	Stroke types	temporal-spatial; joint moment	statistical	41	clustering	-
Ren [62], 2017	NDD; HC	stride intervals	statistical; empirical mode decomposition	64	NB;SVM;RF; MLP;SLR	AUC= 0.949
Ma [63], 2016	Glaucoma; HC	steady-state; temporal-spatial	statistical; temporal-spatial	19	DT; kNN; regression	Acc= 94%
Banaie [64], 2011	NDD; HC	stride intervals	statistical	64	17 classifiers	Acc= 86.957%
Gandomkar [47],2014	fallers; non-fallers	gait energy image	histograms	24	SVM	Sen= 88%
Ren [65], 2016	NDD; HC	stride intervals	conditional entropy	64	MLP;RF; NB	AUC= 0.958
Lee [10], 2012	PD; HC	GRF	wavelet	166	FNN	Acc= 81.63%
Gestel [57], 2011	CP patterns	kinematics; kinetics	states of gait	139	BN	Acc= 88.4%
Agostini [67], 2014	NDD; HC	foot switch signals	gait phases; temporal	10	DT-like	Acc= 98%
Wu [66], 2010	PD; HC	stride intervals	probability density; signal turns count	31	SVM	Acc= 90.32%
Bilgin [49], 2017	ALS; HC; NDD	GRF	wavelet	64	LDA;NB	Acc= 90.93%
Wang [58], 2012	walking patterns	shank acceleration	gait phases	20	DT	Acc= 98.87%
Baratin [9], 2015	NDD; HC	stride intervals	wavelet	64	LDA	Acc= 85%
Ilias [59], 2016	Autism; HC	spatial; kinetic kinematic	statistical	44	ANN; SVM	Acc= 95.00%
Mezghani [68],2008	AS; OA	GRF	kinematic; kinetic; wavelet	42	kNN	Acc= 91%

Table 3.5. Summary of the state-of-the-art researches for machine learning-based gait classification.

CRPS=Complex Regional Pain Syndrome; HC=Healthy Control; HO=Hip Osteoarthritis; FOG=Freezing of Gait; PFPS=Patellofemoral Pain Syndrome; GRF=Ground Reaction Force; NDD=Neurodegenerative Diseases; RF=Random Forest; KFD=Kernel Fisher Discriminant; NB=Naive Bayesian; MLP=Multilayer Perceptron; SLR=Simple Logistic Regression; AUC=Area Under Curve; DT=Decision Tree; BN=Bayesian Networks; LDA=Linear Discriminant Analysis; AS=Asymptomatic; OA= Osteoarthritis; Acc=Accuracy; Sen=Sensitivity; kNN=k-Nearest Neighbors; FNN=Fuzzy Neural Network; AD=Alzheimer's Disease; PD=Parkinson's Diseases; MS=Multiple Sclerosis; CP=Cerebral Palsy; ALS=Amyotrophic Lateral Sclerosis.

3.4 Novel Machine Learning Framework for Gait Classification

The main contribution of this thesis is a proposal for a novel machine learning framework for gait classification. The proposed framework is modified from the basic machine learning framework with additional blocks and functions in order to enhance the ability of the framework to solve gait classification problems. The block diagram of the proposed novel framework is depicted in Fig.3.5.

The definitions and functionalities of the blocks are given below. The blocks with orange background are the newly proposed blocks, and the blocks without any background colors are the unmodified blocks based on the traditional machine learning diagram as depicted in Fig.1.1. The large dashed rectangles are the boundaries highlighting the main novel functions. Pre-processing and feature selection blocks are bound with dashed lines, as they are not indispensable.

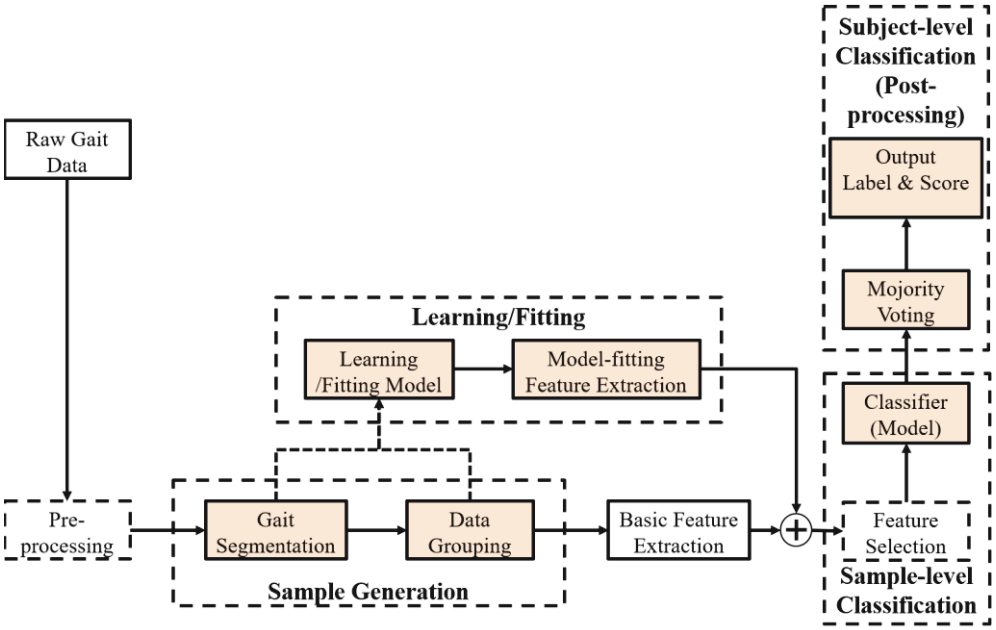


Figure 3.5. Proposed machine learning framework for gait classification.

3.4.1 Term Explanation

Raw gait data: The original input gait signals. The gait signals studied in this thesis are the joint angle signals, the trunk acceleration signals, and the stride interval signals. The

joint angle signals are collected from the hip and knee joint while walking over ground, while the trunk acceleration signals are the acceleration signals collected from the back of the waist. The stride interval signals consists of five signals of related parameters, which are stride time, swing time, swing percentage, double support time, and double support percentage.

Pre-processing: A common block with the original machine learning framework, which aims to make the input signals easier to work with in the following steps. Common techniques are filtering, normalization, data cleaning, outlier removal, interpolation, and so on.

Gait segmentation: This step segments the complete signals and time series into smaller units based on the semi-periodic characteristic. The main function of this block is to prepare the data for sample generation together with the next step, i.e., data grouping. Depending on the data type being processed, the gait signals are segmented into either gait cycles (strides) or steps. For instance, the gait signal of the joint angles collected from one leg will be segmented into cycles, since this is the minimum unit to define a meaningful gait motion. The trunk acceleration will be segmented into steps, since the minimum unit, which corresponds to one simple gait period, is one step performed by either the left or right leg in turn. One gait cycle contains two steps, one made by each leg.

Data grouping: Grouping the segmented data into meaningful groups, with each group considered as a sample for extracting features and performing the classification. The main reason for grouping the segmented cycles or steps into groups is to form a fragment of data for effective feature extraction, since it is usually not enough to extract features from only one single gait cycle, but important to compare between cycles in order to understand the variability of gait. The group mentioned here is the minimum unit used to define a sample for extracting all features that are needed. It has to be mentioned here that data grouping does not only involve putting adjacent cycles or steps together, but also means putting nonadjacent cycles or steps together, and the only criteria used to define the grouping method are the meaning of desired features.

Basic feature extraction: This block extracts the necessary features for classification. Those features are named as basic features to be distinguished from the model-fitting (MF) features. These features are usually extracted using the techniques suggested in authoritative studies, and have already been proved to be effective for gait classification under certain scenarios.

Learning/Fitting model: This block is a novel function that contributes to the overall classification system. It is included to boost the performance of the framework by taking advantage of an additional model. The requirements for selecting a proper second model is that, the two models should be independent from each other. In other words, the system can potentially be a good classifier if the two component classifiers/models represent the data structure from different dimensions.

Model-fitting features extraction: This block extracts the model-fitting features (machine learning features) using an additional model. The extracted MF features will be fed together with the basic features to the high-level classifier that is concatenated with the low-level model. The extraction of MF features depends upon the characteristics and structure of the data type and the features can be extracted from data either before or after

grouping. For example, the MF features are extracted from segmented data when we want to describe the shape of signals of one single gait cycle, while they are extracted from the grouped fragment when we would like to investigate the variability between gait cycles. The low-level machine learning models utilized in this thesis are GMM, least square regression, and kNN.

Feature selection: A common block from the traditional machine learning scheme. The feature selection techniques mainly used in this thesis are the t-test and PCA.

Classifier (Model): The high-level classifier/model is the core classifier that performs classification on the generated samples, based on the two-level classification scheme of the proposed framework. The classification results are the predicted labels for each test sample. In this thesis, a classifier is referred to as a high-level classifier if it performs the final classification, and is different from a low-level model, which aims to extracts features using machine learning methods. The main high-level classifiers employed in this thesis are ANN and SVM.

Majority voting: This is a function that performs the subject-level classification, or namely, the post-processing. This stage is referred to as classification, since it is a procedure that performs the final decision-making at the subject level. However, despite its name, no machine learning technique is utilized at this stage; instead, the majority voting (MV) plays the role of decision making. Majority voting computes the percentages of the samples predicted as being in either class for a given test set of one subject, and predicts the final label and confidence score of this subject based on the majority of sample level results obtained from the sample-level classification. More detailed explanations will be given in Chapter 4, along with practical applications.

Output label and score: They are final output of the framework. Different from traditional classification framework, from which usually only a label is given as the output, the output of the proposed framework could also be a confidence score, or "score" for short, as an additional metric. The label is the class to which the majority of samples are predicted to belong, and the score is the percentage of the samples predicted as belonging to the major class. The main advantage of the proposed score is that it can be potentially used as an indicator of the level of "healthiness" or "pathology level" of the gait patterns, since it is a scalar number. This will be discussed in depth in later chapters.

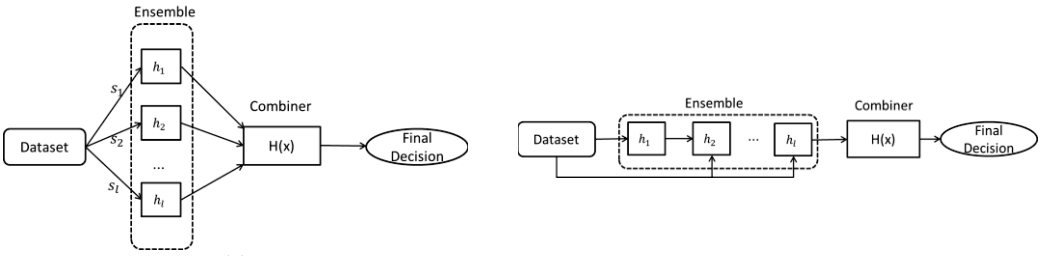
3.4.2 Highlights

The main innovations of the framework have been highlighted with red dashed rectangles in Fig.3.5. Recalling the four limitations of the current gait classification approaches mentioned in Section 3.3.2, the novelties of the proposed framework, as well as how those novelties may overcome the limitations are carefully explained below.

1. **The hybrid model.** Different from the methods in most of the previous studies, where various classifiers were utilized separately for classification, the proposed framework tries to combine models to achieve a better performance. Therefore, the focus of the framework is not to compare different machine learning

techniques and find the best one, but to integrate one classifier into another to enhance the overall performance and efficiency.

According to the "No Free Lunch" theorem, there is not a single classifier that can be considered optimal for all problems [70], and there is no clear guideline for choosing the proper learning methods; moreover, it is rare for one single classifier to have complete knowledge about the data distribution and how the classification algorithm should work. Therefore, it is usually difficult to find a good single classifier in practical pattern classification tasks, especially for gait classification problems, where the size and quality of the dataset are mostly limited. The two most important motivations for combining classifiers are as follows: the statistical motivation, meaning that it is possible to avoid the worst classifier by averaging several classifiers; and the representational motivation, meaning that, under particular situations, fusion of multiple classifiers can improve the performance of the individual classifier [71]. The two typical architectures for combining classifiers are parallel architecture and serial architecture. The parallel architecture is a structure where a set of classifiers are trained in parallel, and their outputs are combined afterwards to give the final decision; the serial architecture is a structure in which a primary classifier is used, and its output errors or parameters are given to the second classifier for training. The final results are the output of the last classifier, which receives and combines all inputs from primary classifiers. The two types of architecture are illustrated in Fig. 3.6.



(a) (b)

Figure 3.6. Two architectures for combined classifiers.

The architecture applied in this thesis follows a serial structure, with the output of one model serving as the input of another. Features will be extracted from the low-level model from independent structures and dimensions of data.

2. **The sample generation**. As has been pointed out, one of the main general limitations of gait classification problems is the number of subjects. In order to generate a large enough number of samples for training and testing the classifier, the sample generation block is proposed according to the semi-periodic characteristic of gait signals. This block consists of two steps: gait segmentation and data grouping. The complete gait signal, which usually contains the walking data collected from a certain distance or time, is cut into smaller units defined by gait cycles or steps. Afterwards, those steps or cycles of data are grouped into fragments that will be used for feature extraction. By using these two steps, the number of samples is greatly increased for performing classification, owing to the

fact that the number of cycles or steps performed by the subjects is considerably more than the number of subjects participated in the experiment.

3. **The two-level classification scheme**. In association with the sample generation step, the two-level classification scheme is generated, where in the first level, classification is performed on the large number of samples generated based on segmented gait cycles or steps, and in the second level, the classification result of the first level is summarized to arrive at the final decision for one subject, who has performed multiple gait cycles of walking in the experiments. The second level of classification is more of a statistical computing step than a real "classification" step since no machine learning techniques are utilized. Compared with traditional approaches, with which the classification of gait patterns is done directly at the subject level, this proposed approach focuses more on the cycle and sample-level quality of gait, as well as the cycle/step level gait variability. By applying this two-level classification scheme, the intra-subject variability and the structure of data contained in the personal walking characteristic can be fully analyzed and sufficiently abstracted for classification purposes.
4. **The output score**. Along with the output label, which indicates the predicted class of the test subject, the score is proposed as an additional metric for measuring the likelihood of the prediction. Computed as the percentage of the samples predicted as belonging to the predicted class, the score can be used as an indicator of the level of "healthiness" or "unhealthiness" after some calculations, or namely, the quality of gait. As mentioned before, one main limitation of previous studies was that they used binary classification results of one or zero, which are not sufficient anymore for higher requirements of the modern applications fields of gait analysis, such as rehabilitation monitoring, medication effect assessment, and illness severity evaluation. Contrastively, the proposed score can potentially provide an additional parameter for more detailed evaluation of gait quality, which can be achieved by comparing the changes in the score before and after treatment or rehabilitation operations.
5. **The wide applicability**. Unlike the methods proposed in most of the previous studies, which were only verified with one certain gait data type or scenario, the proposed framework can be more effective for various types of gait data types and different classification scenarios. This is because the semi-periodic characteristic of gait signals is explicitly integrated into the architecture of the framework. It is evident that all types of gait signals are able to be segmented into cycles or steps. Hence, it is believed that the sample generation, as well as the two-level classification scheme, highlighted in this framework, is applicable to all gait data types. Besides, the combined classifier enables the exploration of data structures within different dimensions, and this is feasible and beneficial for all types of gait data.

In summary, this thesis takes the four main limitations of previous studies into consideration, and proposes a novel framework, which could overcome the limitations and solve gait classification in different scenarios for various data types. This is achieved by integrating four innovative functions and schemes, including a combined classifier to enhance the classification accuracy by exploring data from more dimensions, a sample generation scheme to produce a large enough number of samples for more robust learning,

a two-level classification architecture that enables deeper evaluation of cycle-level gait quality, and a confidence score that can be transferred to an indicator of the degree of gait quality. This framework is designed based on the semi-periodic characteristic that can be seen for all types of gait signals; thus, it is believed that this framework will be able to provide an applicable solution for all types of gait signals and different scenarios. The framework integrates the advantage of previous achievements, such as advanced feature extraction and signal processing techniques, as well as the ideas of state-of-the-art approaches and future directions of development, such as hybrid classifiers and multilevel classification. In order to validate the effectiveness of the framework, comprehensive classification studies are carried out using the framework on three typical types of gait data, i.e., the joint angle signals, the trunk acceleration signals, and the stride interval signals, which will be presented in details in Chapters 4, 5, and 6, respectively.

3.5 Conclusions

This chapter explicitly explains the main contributions of this thesis. In describing the basic science of human gait, measurement systems for gait analysis, related studies, including the current research directions and the state-of-the-art approaches, and the proposed solution framework, this chapter has fully stressed the overall needs, current status, and challenges of gait classification problems. Building on these current advancements and considering the present limitations, the proposed framework tries to link the traditional approaches with new concepts and ideas. The theoretical explanation of how this framework will overcome the limitations and solve various types of gait classifications problems has been discussed briefly in this chapter. Detailed validation work, as well as applications will be presented in depth with real gait signals in the following chapters, with theoretical approvals, experimental results, and practical clinical (case) studies.

4. Gait Classification for Joint Angle Signals

This chapter describes validation of the proposed machine learning framework and its application to the classification of healthy and pathological gait patterns based on joint angle signals. The processed signals are the hip and knee joint angle trajectories recorded using the AHRSs with the setup introduced in Chapter 3; however, the machine learning framework is not limited to the data collected from AHRSs, but is a general one that can be applied to data collected by different types of devices, such as vision-based equipment or goniometers.

The related work, including the characteristics of the hip and knee joint angle signals, the statement of the problem, the challenges that may be encountered, and the previous classification approaches are discussed in Section 4.1. Section 4.2 gives the details of the proposed gait classification solutions with breakdown explanations for each procedure. In Section 4.3, the approach is evaluated on a dataset built from the data of 28 healthy subjects and 30 patients with different types of neurodegenerative diseases collected by the author; detailed results and discussion are also provided. In Section 4.4, the proposed approach is applied to four possible application scenarios, including the management of patients with NDDs, for the assessment of the medication's effect, as well as for long-term monitoring purposes. This chapter is summarized, and conclusions are dawn in Section 4.5.

4.1 Related Work

The data that this chapter focuses on are the hip and knee joint angle signals. Owing to the fact that gait ideally tends to be a semi-periodic behavior, the joint angle signals collected from a walking trial containing multiple gait cycles are commonly analyzed by investigating the individual gait cycles or the averaged gait cycles in clinical gait analysis [16]. Furthermore, the rhythm of the gait is also very important and can be regarded as a factor that determines walking quality [41] [42]. Additionally, the inadequate or excessive extension/flexion behavior of the hip and knee joints always corresponds to specific gait disorder symptoms and therefore has been associated with certain pathologies [16] [20]. For instance, increase in hip and knee flexion can be commonly seen in a step-page gait, also known as a foot drop gait, whose symptoms are highly associated with polyneuropathy, while insufficient flexion of the knee can often be seen in a patient with Parkinson's disease owing to the shuffling gait [20]. In brief, it is

essential to analyze the joint angle behaviors for assisting with gait assessment and rehabilitation, which can considerably contribute to the development of engineering solutions that take advantage of machine learning techniques.

4.1.1 Joint Angle Signals

The signals of the hip and knee joint angles for a walking trial are continuous waveforms containing peaks and valleys, whose incidences are usually the starting or ending points of the gait cycles. An example of the hip and knee joint angle signals of one healthy subject performing straightforward walking along a corridor for 13 gait cycles recorded using IMUs is illustrated in Fig. 4.1. The red curve represents the hip angle, and the blue curve is the knee angle. It should be noted that, the signals are pre-processed raw data, and the data has been compensated with offset values so that both hip and knee signals start approximately at the zero-position, i.e., the initial position where the subject is required to stand straight with both hip and knee angles equaling zero.

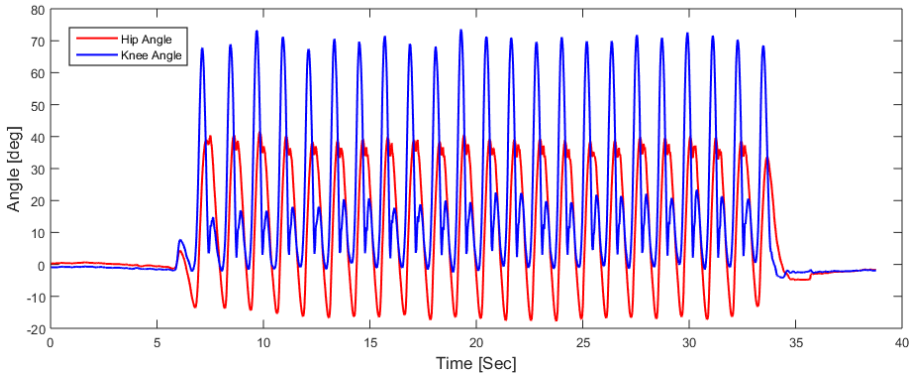


Figure 4.1. Illustration of hip and knee joint angle signals for 13 gait cycles.

Recalling the ideal hip and knee joint angle trajectories for one complete gait cycle, as presented in Fig.3.2 (b), it can be easily observed from Fig. 4.1, that the signals contain multiple gait cycles, and the shapes of trajectories for both joints have un-ignorable differences between gait cycles. Therefore, to understand the walking behavior, it is necessary to investigate the shape of the trajectories of individual gait cycles and look into the differences between them numerically.

4.1.2 State-of-the-art and Limitations

Few classifications have been done with kinematic parameters, i.e., the hip and knee joint angles, in previous research on classifying the gait patterns between healthy control (HC)

groups and pathological (PT) groups. A summary can be seen in Table. 4.1. The best results listed are the classifications with the best accuracy obtained in those studies with only kinematic input.

Several conclusions can be drawn from Table 4.1:

- The hip and knee joint angles have been utilized in all the four studies.
- The peak values and the range of motion (ROM) of the joint angle trajectories are the most considered features.
- The SVM is the most preferred classifier.
- The best classification accuracies obtained were 90% for classifying PT (including PD) and HC, and 100% for classification of healthy young and healthy elderly subjects.

Publication	Kinematic Data	Features	Classification Output	Classifier	Best Results
D. Laroche et al [52]	Joint angles of hip, shank, and foot	Time series of joint angles for gait cycle	PT vs. HC	SVM	Acc=90%
R. Begg et al [72]	Joint angles of ankle and knee	Peak values and ROM	Healthy Young vs. Healthy Elderly	SVM	Acc=100%
H.H. Manap et al [73]	Joint angles of hip, knee and ankle	Peak values	PD vs. HC	SVM	Acc=71.36%
N. M. Tahir et al [74]	Joint angles of hip, knee and ankle	Peak values	PD vs. HC	SVM & ANN	Acc=78.1%

Table 4.1. Summary of state-of-the-art researches for gait classification using kinematic data

Even though very promising classification results have been reported from previous studies using kinematic features, several limitations still exist, and a novel approach that can overcome those limitations is needed. To be specific, for those studies relying on the peak values, such as maximum and minimum flexion/extension angles, and ROM values, in which the peak values are used as the most contributive features for distinguishing HC and PT gait patterns, the shape of the joint angle trajectories may not be represented comprehensively in many cases by solely using those values. In Fig. 4.2 there are two adjacent gait cycles (resampled to same length in percentage), with the same peak values and ROM, but inequivalent shapes. Considering solely the peak values in those cases will lead to a loss of information.

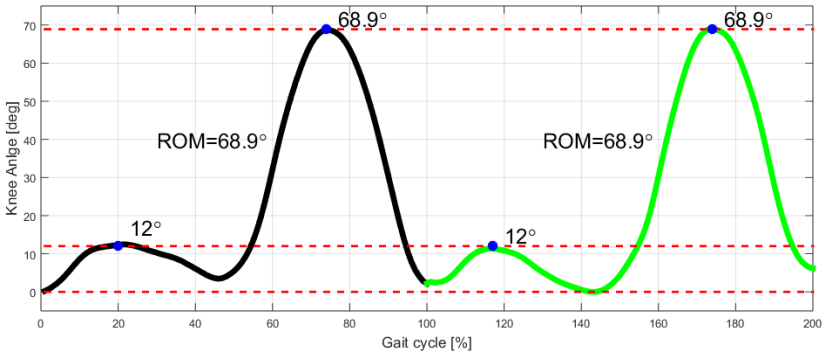


Figure 4.2. Two gait cycles with same peak values and ROM but inequivalent shapes.

Additionally, it has been discovered that the walking speed is highly associated with gait kinematic performances [20]. In [75], B. Koopman proved that the hip and knee joint angles depend highly on the walking speed and the square of walking speed. Moreover, a higher walking speed ideally tends to yield larger ROMs for both hip and knee joints. On the other hand, the natural walking speed is associated with age [76] [77]: human's natural walking speed remains relatively stable until about age 70 and then declines about 15% per decade. As most of the patients suffering from neurodegenerative disease are elderly people whose ages are in excess of 70 years and naturally have much slower walking speeds, the peak values of the joint angles may have significant differences from the healthy subjects, especially the young healthy subjects. The limitation of using the peak values as the dominant features for classification is that, a healthy subject who intentionally or accidently walks with a speed slower than his or her natural speed may be misclassified as a pathological subject. Therefore, it makes sense to investigate, instead of solely utilizing peak values and ROM, the whole shape of the joint angle signals. The normalized shape of hip and knee angles is also worth being investigated, since the influences of walking speed need to be eliminated.

Various studies recently have proposed different variabilities or stabilities based gait features for solving the gait classification problem. For example, P. Ren et al. investigated conditional entropy as the main variability feature for classification of three different types of PT gait (Lateral Sclerosis (ALS), Huntington's disease (HD) and PD) and HC gait [65]; Y. Wu et al. proved that the rhythm, i.e., the variability of PD gait is significantly different from that of HC gait, and a 90.32% classification accuracy has been achieved using those rhythm features and the SVM classifier [66]. However, those state-of-the-art features were only extracted from temporal-spatial parameters, and the variability of kinematic signals has so far not been evaluated. Inspired by recent research achievements, novel features are proposed in this chapter and integrated into the machine learning framework introduced in Chapter 3 for gait classification. In contrast to most of the previous studies that mainly focus on specific values, in this chapter, the features are extracted from two aspects: the shape of the joint angle signals of the gait cycles, and the shape variability between gait cycles.

4.2 Gait Classification Using Variability Features and Shape Features

The machine learning framework introduced in Chapter 3 is deployed and adapted to the data and features this chapter concerns it with. The flowchart of the classification scheme is depicted in Fig.4.3. The most important procedures are emphasized with bold text.

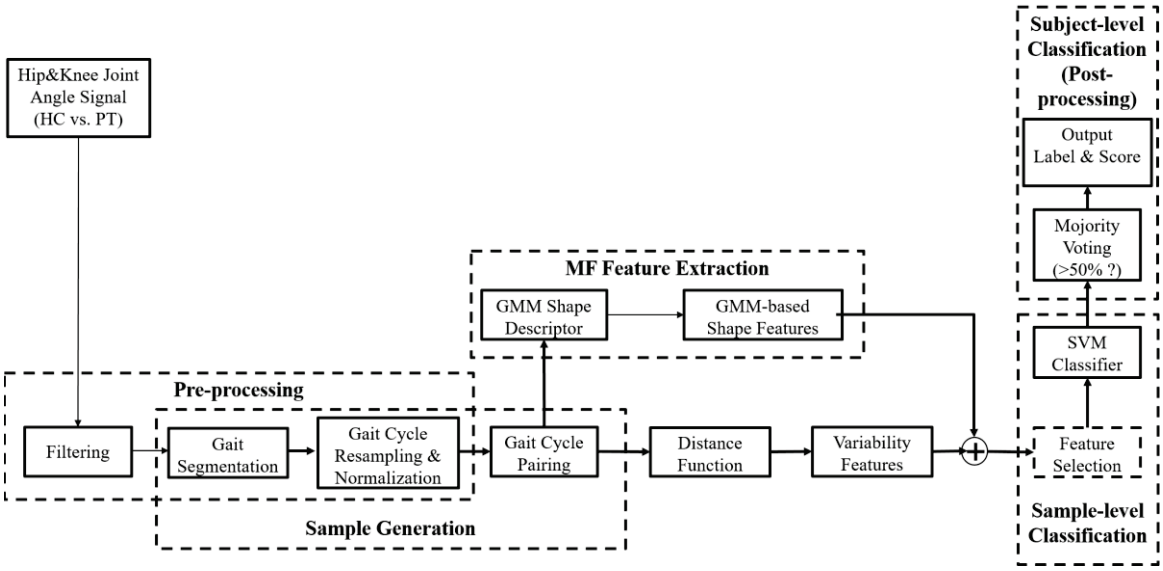


Figure 4.3. Flowchart of the classification scheme for joint angle signals.

Given the hip and knee joint angle signals as raw input, the signals are filtered first to remove noise, and then segmented into individual gait cycles using an enhanced peak detection approach. The segmented gait cycles are resampled to the same number of data points and normalized to have to the same ROM from 1 to 100 afterwards. Then, the gait cycles are paired to generate the samples for purposes of classification, which will be introduced in Section 4.2.2. GMM is employed as a shape descriptor for each pair of gait cycles to generate the model-fitting features, i.e., the shape features; in parallel, four distance functions are defined and utilized to calculate the distances between the shapes of the two cycles in each pair using both normalized and un-normalized cycles. All features are served as input to a high-level classifier trained using SVM. Classification is done for each of the generated samples, and a majority voting scheme is defined to make decisions on the pattern of the subject depending on the majority of the output labels of all samples generated from the corresponding subject. The machine learning framework was validated on the data collected from 28 healthy control (HC) subjects and 30 patients with three types of NDDs. A leave-one-subject-out (LOSO) validation mechanism was employed in the evaluation phase, and the results are presented and discussed in Section 4.3.

4.2.1 Data Pre-processing

The moving average filter is the filter most commonly applied to gait signal data for noise reduction, especially data recorded with inertial sensors [78]. Therefore, a moving average filter is applied to the hip and knee joint angle signals firstly to remove the noise introduced during data recording. The filter smooths the data by replacing each data point with its neighboring data points defined within the span. The values after filtering are calculated as:

$$y_s(i) = \frac{1}{2K+1}(y(i+K) + y(i+K-1) + \cdots + y(i-K)) \tag{4.1}$$

where $y_s(i)$ is the smoothed value for the i^{th} data point, K is the number of neighboring data points in either side, and $2K+1$ is the span. The number of spans used in this step is five. It should be noted here that, since the data miss rate is extremely low and can be ignored in this thesis, the signals this chapter focuses on are considered to be time series with the data sequence taken at successive equally spaced points in time. In case there are missing data, the missing data points are interpolated with the mean of their neighbor values. In this chapter, the term "signal" will always refer to 1-D discrete sequence without considering the temporal dimension.

4.2.1.1 Gait Cycle Segmentation

For the purpose of generating a large number of samples for performing machine learning-based classification, where the variability features and the shape features are extracted from each sample, the filtered signals are segmented into separate gait cycles in the next step using an enhanced peak detection algorithm, *PeakDet* [79]. The algorithm is based on the semi-periodic characteristics of gait cycles and the fact that the knee angle value at the heel contact moment of one cycle tends ideally to be a local minimum in the trajectory. The peak detection method used is a recursive algorithm that assumes that the peak is surrounded by valleys, and the peak value should be δ times greater than the valley value, where δ is a scaling factor representing the ratio between the values of the peak and valley. The method has been enhanced to improve its robustness on segmenting gait cycles since some detected peaks are caused by noises and do not correspond to the starting or ending incidences of gait cycles.

The *PeakDet* method is firstly deployed to detect all possible peaks in the knee joint angle signal, and the peaks are verified successively afterwards; only the peaks that fulfill the following criteria are stored as valid peaks for segmentation.

$$\begin{gathered} Hip(peak_i) > Th_{hip} \\ and \quad Knee(peak_i) < Th_{knee} \\ and \ peak_i - peak_{i-1} > 0.5 \times L_{est} \end{gathered} \tag{4.2}$$

The $Hip(peak_i)$ and $Knee(peak_i)$ are the hip and knee angle values at the location of i^{th}peak, respectively; Th_{hip} and Th_{knee} are the threshold values for hip and knee joints, respectively, and in my thesis they are assigned as 13° and 30° respectively. The L_{est} is the estimated number of data points in one cycle, calculated by dividing the number of data points in the whole signal by the estimated number of gait cycles. In principle, a cycle's starting value for the hip is supposed to be larger than 13°, and smaller than 30° for the knee; the length of the cycle should be larger than half of the estimated average lengths of all cycles. With those criteria, the incorrect detections are eliminated.

The gait cycle detection result for one knee angle trajectory recorded during the straight walking of one healthy subject along a 70 m long hallway is illustrated in Fig. 4.4. Red crosses indicate the detected heel contact incidences, which correspond to the starting and ending points of a gait cycle. In this thesis, the first two and last two cycles that correspond to the acceleration and deceleration phases of one walking trial are neglected. After the segmentation of knee joint angle signals, the hip signals are segmented into gait cycles automatically, since the data are recorded simultaneously for hip and knee joints.

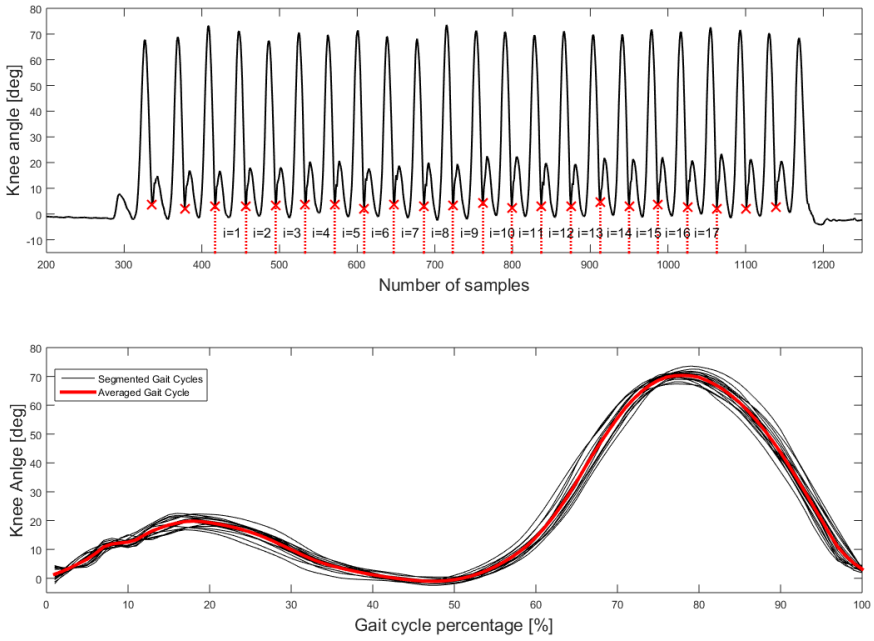


Figure 4.4. Gait cycle segmentation for knee joint angle signal.

4.2.1.2 Gait Cycle Resampling

The segmented gait cycles are resampled to have the same number of data points, that is, $n = 1000$ data points per cycles, using the piecewise cubic Hermite interpolating polynomial (***PCHIP***) method [80]. The purpose of resampling the segmented gait cycle

signals is two-folded: 1) To use the introduced distance functions, which perform measurement of shape differences between two signals both with standardized lengths; and 2) to eliminate the influences of walking speed, since the actual number of data points in the original signal is highly associated with the walking speed factor, and in this thesis, the influence of walking speed should be carefully eliminated as claimed in Section 4.1.2.

PCHIP

The PCHIP method interpolates the signal using a piecewise cubic polynomial function $F(x)$ with the following essential properties:

- The polynomial $F(x)$ is a cubic Hermite interpolating polynomial for the given input data points with derivatives at the interpolation points.
- $F(x)$ interpolates y, that is, $F(x_i) = y_i$, and the first derivative $\frac{dF}{dx}$ is continuous. The second derivative $\frac{d^2F}{dx^2}$ is probably not continuous.
- The cubic interpolant $F(x)$ is shape preserving. The shape of the original input data is preserved and the monotonicity is also respected by choosing appropriate slopes at the x_i .

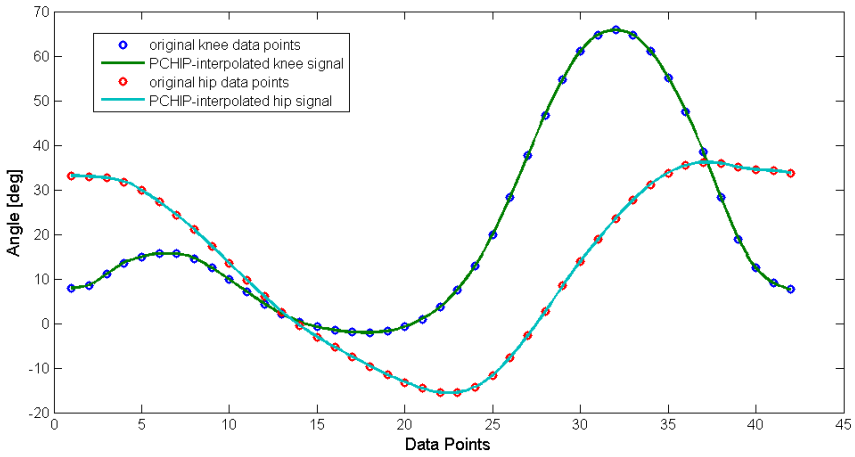


Figure 4.5. A gait cycle of hip and knee joint angle trajectories before and after resampling using PCHIP.

An example of one segmented cycle of hip and knee joint angle trajectories is illustrated in Fig. 4.5 with red and blue circles, respectively, each consisting of 43 original data points. The original data points are resampled using the PCHIP method described above with 10 times the sampling frequency, and the resulting signals are plotted in the same figure. It can be observed, from the figure, that the shape of the resampled signal could follow precisely the shape formed by the original data points, indicating a promising efficiency of the PCHIP method on resampling the gait signals. The number of data points after resampling, n, is a critical factor and the determination of the optimal n is discussed in Section 4.2.3.

4.2.1.3 Gait Cycle Normalization

In order to analyze the shape variability of the hip and knee joint angle signals without considering the influence of the actual walking speed as an additional advantage, the resampled gait cycles are normalized to have values between 0 to 100, representing the percentage of the angle values with respect to the maximum range of motion the joint can reach. The normalized gait cycle is calculated using the following equation

$$X_j' = \frac{100(X_j - \min(X))}{\max(X) - \min(X)}, j = 1,2, \dots, n \tag{4.3}$$

where X_j is the j^{th} data points of the un-normalized gait cycle, and X_j' is the j^{th} data point of the normalized gait cycle. $\min(X)$ and $\max(X)$ are the minimum and maximum value of the un-normalized gait cycle, respectively, and n is the total number of data points of the gait cycle after resampling. In order to analyze the signals comprehensively, both normalized and un-normalized gait cycles are utilized in the next procedures for the extraction of variability features.

4.2.2 Gait Cycle Pairing

Before extracting the necessary features for purposes of classification, the pre-processed (including segmentation, resampling, and normalization) gait cycles (GC) are grouped into pairs using the following method: starting from the first cycle, group the current cycle with each of the rest of the cycles on its right-hand side from its subsequent cycle until the last one. An illustration of the gait cycle pairing procedure for four gait cycles is shown in Fig. 4.6.

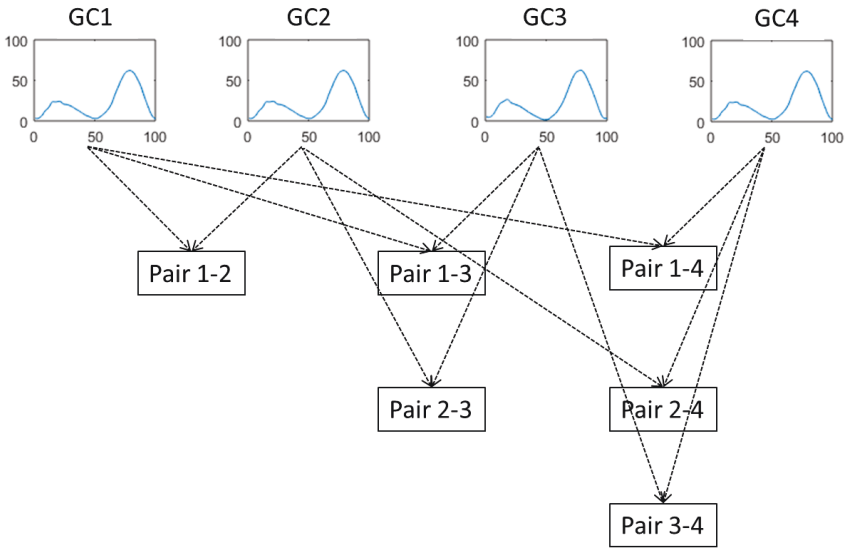


Figure 4.6 Gait cycle pairing for sample generation.

The total pairs of gait cycles as illustrated in Fig. 4.6 are calculated as 3 + 2 + 1 = 6. For a walking trial in which one subject completes m gait cycles, the total number of pairs that can be generated is:

$$N_{pair} = (m-1) + (m-2) + \cdots + 1 = \frac{m(m-1)}{2} \tag{4.4}$$

Two advantage of performing the pairing of gait cycles are: 1) by pairing the gait cycles, the gait cycles are grouped by two, and can be directly used for measuring the differences and generating the variability features afterwards; and 2) the number of pairs N_{pair} is $\frac{m-1}{2}$ times greater than the number of gait cycles m, which is one good solution for the lack of samples during the training and testing phase of the classification. More available samples can be utilized when considering the pairs of cycles as samples for classification, rather than the individual cycles as samples, prior to the assumption that the optimal number of the samples should be much larger than the number of features to perform the classification.

4.2.3 Variability Features Extraction

Owing to the fact that variability is one of the most essential factors used to investigate the performance of gait, it is important to compare the differences between gait cycles. In this thesis, the shape variability of the joint angle signals of gait cycles is considered and 16 features are extracted from the pre-processed gait cycles to measure the variability. Four distance functions are defined as the metrics for the measurement of differences between two cycles.

4.2.3.1 Distance Functions

Given two gait cycles, X and Y, segmented and resampled from one trajectory of hip or knee recorded during one walking trial of one subject, each with n data points, four distance functions are defined to measure the differences between them with regard to the shape of hip and knee joint angle trajectories.

1. ***Mean-Subtraction (MSub):***

$$MSub(X,Y) = \frac{1}{n}\sum_{j=1}^{n}\left|X_j - Y_j\right| \tag{4.5}$$

X_j and Y_j represent the j^{th} data point of the resampled gait cycle X and Y, respectively. The meaning of the output of this distance function is the measure of the sample-standard difference of two signals calculated using the sum of the subtraction of each data point, divided by the total number of data point.

2. ***Root-Mean-Square-Deviation (RMSD):***

$$RMSD(X,Y) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}|X_j - Y_j|^2} \tag{4.6}$$

X_j and Y_j represent the j^{th} data point of the resampled gait cycle X and Y, respectively. RMSD is able to represent the sample-standard deviation of two signals by calculating the square root of the variance.

3. ***Mean-Dynamic-Time-Warping (MDTW):***

$$MDTW(X,Y) = \frac{1}{n}C_{p^*}(X,Y) = \frac{1}{n}min\{C_p(X,Y)|\ p\ is\ an\ warping\ path\} \tag{4.7}$$

The MDTW distance function is defined based on Dynamic Time Warping (DTW), which enables the measurement of the difference between the shapes of two signals by finding an optimal alignment between them [81]. The optimal alignment is a non-linear mapping between the two input signals, achieved by warping the signals in time domain along a warping path. The $MDTW$ is computed by dividing the DTW value, which is the Euclidean distance of the two signals after warping, with the number of data points n.

4. ***Maximum-Cross-Correlation (MCorr):***

$$MCorr(X,Y) = 1 - \max(|R|) = 1 - \max(|xcorr(X,Y)|) \tag{4.8}$$

The Mcorr is computed from the maximum value of the cross-correlation sequence calculated from the two input signals. The cross-correlation sequences are normalized so as to have values that lie between 0 and 1. In order to be consistent with the other distance functions, the maximum correlation is subtracted from 1 so that, for all the four distance functions, smaller values indicate smaller variability of gait.

Dynamic Time Warping

The fundamental functionality of DTW is to construct a non-linear mapping between the two input series to minimize the distance between them. This optimal alignment is accomplished by warping the input sequences non-linearly in the time domain. Given two sequences $Q = q_i, i = 1,2,\dots,n_q$ and $C = c_j, j = 1,2,\dots,n_c$, the alignment cost between the two points q_i and c_j is calculated using a cost function

$$w(q_i, c_j) = (q_i - c_j)^2 \tag{4.9}$$

Subsequently, a n_q by n_c warping matrix is constructed, where the (i^{th}, j^{th}) element of the matrix is the value of the cost function $w(q_i, c_j)$. Having the warping matrix, the global cost matrix G can be obtained, where each element $g(i,j)$ represents the accumulative distance, as follows:

$$g(1,j) = \sum_{r=1}^{j} w(q_1, c_r); \quad g(i,1) = \sum_{r=1}^{i} w(q_r, c_1) \tag{4.10}$$

$$g(i,j) = w(q_i, c_j) + \min\big(g(i-1,j-1), g(i-1,j), g(i,j-1)\big), i \neq 1, j \neq 1$$

A warping path P, is a contiguous set of cost matrix elements that represent the mapping between Q and C, where m_k is the k^{th} element in the path, where $k = 1,2,\dots,K$ and $\max(n_q, n_c) < K < n_q + n_c - 1$.

$$P =< m_1, m_2, \dots m_k, \dots, m_K > \tag{4.11}$$

Then, the optimal path is calculated using dynamic programming to be one of all possible paths that has the minimal cost, which is the accumulative cost between the staring path element $g_P(1,1)$ and the ending path element $g_P(n_q, n_c)$. Consequently, the final DTW-based distance measure D is the sum of all the costs on the resulting optimal path.

$$D = DTW(Q, C) = min\left\{\sum_{i=1}^{K} g_P(q_{il}, c_{jl})\right\} \tag{4.12}$$

The main reason of utilizing $MDTW$ in addition to $MSub$ and $RMSD$ is that DTW is able to find the best alignment of two signals and minimize errors introduced by phase shifting. For instance, as show in the upper plot of Fig. 4.7, gait cycle 2 is the knee angle trajectory of one segmented gait cycle, and gait cycle 1 has exactly the same signal as gait cycle 2 with 5% of shifting to the right-hand side, which can be potentially caused by inaccurate cycle segmentation. The shapes of the two signals are explicitly the same. However, the MSub and RMSD functions give 7.28 and 9.73 as the sample-standard differences of the two signals. The DTW is able to find an optimal alignment to compensate for the errors caused by phase shifting, and give the actual measure of the shape difference. As can be seen from the lower plot, the two signals are very well aligned after performing DTW, and the MDTW gives 0.28 as the sample-standard difference.

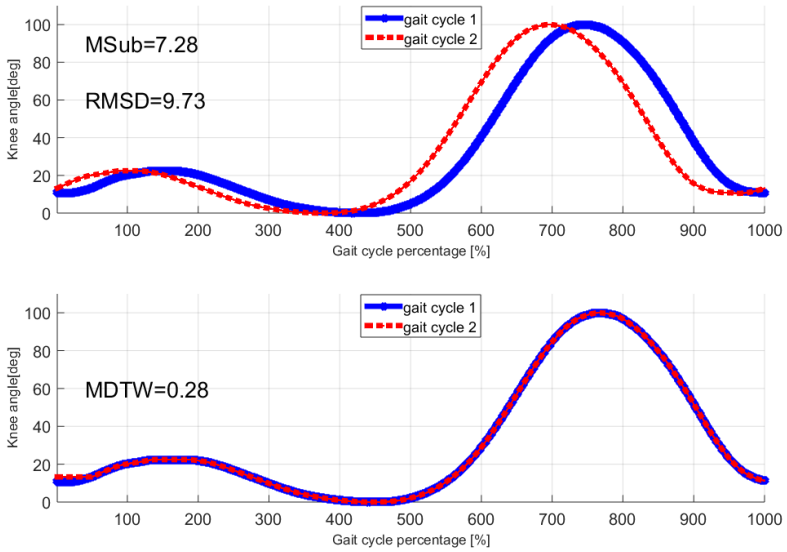


Figure 4.7. Two gait cycles with phase shifting before and after applying DTW.

Cross-correlation

Cross-correlation is a technique that is commonly used for matching signals. In signal processing, it is also used as a measure of the similarity of two series as a function of the displacement of one relative to the other. The function is also known as a sliding dot product. The cross-correlation sequences for two discrete time sequences, x_n and y_n, is calculated using

$$R_{xy}(m) = E\{x_{n+m}y_n^*\} = E\{x_n y_{n-m}^*\} \tag{4.13}$$

where $-\infty < n < \infty$, the asterisk denotes complex conjugation, and E is the expected value operator. Cross-correlation measures the similarity between x_n and lagged copies of y_n as a function of the lag [82]. In our case both signals are resampled to have the same length, therefore no zero-padding is performed at the end.

Practically, as only a finite segment of one realization of the infinite-length random process is available, the sequences of cross-correlation are estimated using

$$\hat{R}_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m}y_n^*, & m \geq 0, \\ \hat{R}_{yx}^*(-m), & m < 0. \end{cases} \tag{4.14}$$

The output vector, which is the cross-correlation sequences, has elements given by

$$c(m) = \hat{R}_{xy}(m - N), \;\; m = 1,2,\dots,2N-1 \tag{4.15}$$

Similar to MDTW, the advantage of using MCorr is that, in case either of the signals is phase shifted by inaccurate segmentation of the gait cycle, the MCorr is still able to compute the similarity of two signals accurately by sliding one of the signals to find the optimal match.

All of the four distance functions are defined based on the existing popular similarity measures and modified according to the gait application for extracting variability features. Even though they are all descriptors representing the shape differences of two signals, they reveal differences from different mathematical aspects and contain information from different dimensions. The significance of these distance functions, as well as the features extracted with them, is analyzed in the feature selection step.

4.2.3.2 Optimal Resampling Rate

It is worth noting that, to define MSub, MDTW, and MRMSD, the values obtained from the original methods are divided by the number of data points n in order to reveal the sample level difference of the two signals. Therefore, the selection of an optimal n is important, and the influence of n on the values obtained with those functions has to be investigated.

Figure 4.8 depicts the relationship between the three measures, along with the computation time, with respective to n, computed from the knee trajectories of two cycles

using PCHIP as the resampling method. The values obtained from the three distance functions are all normalized to 0–1 in order to bring them to the same scale for comparison. The computation time is the time it takes to compute the values from all of the three functions, and is normalized as well. It can be observed from the plot that the values obtained from the three functions have a significant drop, as n increases to 500, indicating that a n smaller than 500 would not suitable, as it has a large impact on the results. On the other hand, the computation time for a n larger than 1500 increases rapidly. Consequently, the optimal range for n lies between 500 and 1500. Therefore $n = 1000$ is selected in this chapter for all gait cycles of hip and knee trajectories as the optimal n.

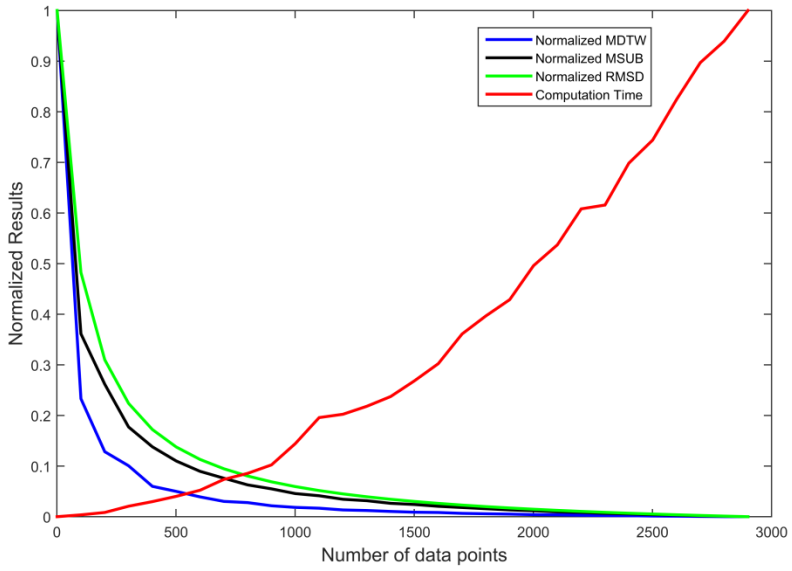


Figure 4.8. Influence of number of data points on the results of distance measures

4.2.3.3 Variability Features

The cycles are grouped to form new samples in order to perform classification, as introduced in Section 4.2.2, and the two resampled gait cycles in each pair are used as the two inputs sequences in this step to calculate the shape variability features. Considering that each cycle contains a hip signal and a knee signal, and both the original signal and the normalized signals are employed, in total 16 features are extracted from each pair of gait cycles using all the four introduced distance functions. It has to mention here that, hip cycles are only compared to hip cycles, and the same rule applies to knee cycles, since it does not provide any meaningful information to compare a hip cycle to a knee cycle. A table of all the variability features and their denotations are listed in Table. 4.2. The whole package of the 16 features is regarded as one observation, namely a sample, during the classification phase.

Feature Index	Feature Denotation	Feature Index	Feature Denotation
V1	V_{HSO}	**V9**	V_{HSN}
V2	V_{HDO}	**V10**	V_{HDN}
V3	V_{HRO}	**V11**	V_{HRN}
V4	V_{HCO}	**V12**	V_{HCN}
V5	V_{KSO}	**V13**	V_{KSN}
V6	V_{KDO}	**V14**	V_{KDN}
V7	V_{KRO}	**V15**	V_{KRN}
V8	V_{KCO}	**V16**	V_{KCN}

Table 4.2. Indexes and denotations for variability features.

(The three capital letters represent the joint (H-Hip, K-Knee), the distance function (S-MSub, D-MDTW, R-RMSD, C-MCorr) and the data type (O-Original, N-Normalized), respectively.)

4.2.4 Shape Features Extraction

In addition to the variability features, the shapes of the hip and knee angles are worth investigating as well, since the kinematic behavior of joints are highly associated with the nervous system, which can be degenerated from neurological disorders. The shape of a signal is commonly described with features extracted from the time domain and frequency domain, such as the peak values and the Fourier transform coefficients. However, those commonly used features may not be suitable for hip and knee joint angles, particularly not for one gait cycle. Specially, the time domain features, like peak values and ROM, as explained in Section 4.1.2, are strongly associated with walking speed, while the frequency domain features are usually more suitable for signals with a larger length, i.e., a large number of data points, and with a broader bandwidth. The hip and knee signals, however, have almost fixed frequency for each person based his or her walking pace.

In this thesis, an unsupervised machine learning technique, GMM, is employed as a signal shape descriptor, as well as a tool for feature extraction. GMMs are a parametric representation of the probability density function, based on a weighted sum of Gaussian distributions. The theory and related learning algorithms of GMM have been largely studied in the past [83] [84]. The algorithm has been mainly used for the generating and learning of motion sequences [85], and has been used also as a classifier for classifying gait activities [86], as well as a modelling method for EEG signals during patients' rehabilitation [87]. To the best of our knowledge, no study has applied GMM to the characterization of the shape of gait signals, and in this thesis, we would like to investigate the potential use of GMM beyond the trajectory learning scope, and deploy it as a powerful feature generator.

4.2.4.1 Gaussian Mixture Model

In order to train a GMM consisting of K Gaussian distributions for a given gait signal, three parameters, i.e., mean (μ_k), covariance ($\sum_k$) and prior (π_k) are required. Those parameters are retrieved using the expectation-maximization (EM) algorithm from the mixture of all Gaussians. The probability function $p(x_i)$ is estimated using the following formula for the GMM with K components (Gaussians):

$$p(x_i) = \sum_{k=1}^{K} \pi_k p(x_i|k) \tag{4.16}$$

where x_i represents one of the trajectories used for training the GMM. As one pair of gait cycles is defined as one unit for feature selection, in total, four trajectories, consisting of two hip angle trajectories and two knee angle trajectories, are used. The number of dimensions D equals five, including one dimension for each trajectory, plus one temporal dimension. π_k is the prior, $p(x_i|k)$ is the conditional probability defined as the D-dimensional Gaussian distribution depending on the mean vector μ_k and the covariance matrix $\sum_k$ of the k^{th} Gaussian.

$$p(x_i|k) = \frac{1}{\sqrt{(2\pi)^D \cdot |\sum_k|}} \cdot e^{-\frac{1}{2}((x_i-\mu_k)^T \sum_k{}^{-1}(x_i-\mu_k))} \tag{4.17}$$

The probability density function modeled by GMM is described by the three parameters mentioned above, and those three parameters are the function of cumulated posterior probability E_k, given as

$$E_k = \sum_{i=1}^{N} P(k|x_i) \tag{4.18}$$

Those parameters are learned using the expectation–maximization (EM) algorithm [88].

As a shape-descriptor, the output parameters of the GMM models are served as features to the high-level classifier for performing the classification. For each pair of gait cycles that contain two hip angle trajectories and two knee angle trajectories, a multiple-dimensional GMM algorithm is used to train a model with multiple components. The number of components is five, meaning that we would like five Gaussian distributions to be generated in the model of one pair of cycles. The number of dimension is five for each pair of gait cycles, including one temporal dimension, two dimensions for the hip angle for both cycles, and two dimensions for the knee angle of both cycles.

Figure 4.9 shows the results of modeling the shape of one pair of hip and knee joint angle trajectories using GMM. The green dots in the upper plots are the centers of the Gaussian distributions after modeling, and the red and blue curves are the original trajectories. The lower plots are the illustrations of the Gaussians, with the ellipses representing the locations and the characteristics of the Gaussians: thicker ellipses represent Gaussians with higher dispersion. The features extracted from this GMM model of the hip and knee angles are the normalized locations of the centers for each component in all dimensions. The equation for computing one feature using its corresponding component's center location is as follows:

$$\mu = \frac{\hat{\mu} - a}{b - a} \tag{4.19}$$

where μ is the normalized center location for one component, $\hat{\mu}$ is the center location of the corresponding component, and a and b are the minimum and maximum values of the dimension, respectively. For example, the values for a and b are 1 and n, respectively, for the temporal dimension, n is the number of data points for one cycle after resampling, which is 1000 in this thesis; $b - a$ is the ROM for the hip and knee dimensions.

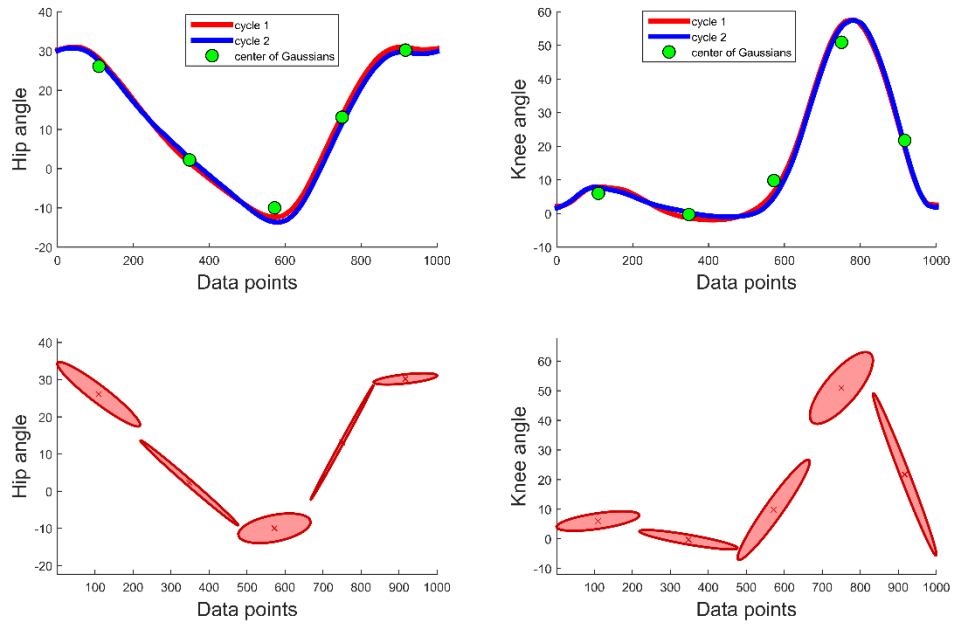


Figure 4.9. Illustration of GMM as shape descriptor for hip and knee joint angle signals

Features are extracted for all the five dimensions for each component. In total, 15 shape features are extracted for one pair of gait cycles. The list of all shape features is summarized in Table. 4.3. In contrast to the traditional shape descriptors, the novel GMM-based shape descriptor is based on an advanced unsupervised machine learning algorithm and can be generated when multiple trajectories are taken into account simultaneously.

Feature Index	Feature Denotation	Feature Index	Feature Denotation	Feature Index	Feature Denotation
S1	μ_{X1}	**S6**	μ_{H1}	**S11**	μ_{K1}
S2	μ_{X2}	**S7**	μ_{H2}	**S12**	μ_{K2}
S3	μ_{X3}	**S8**	μ_{H3}	**S13**	μ_{K3}
S4	μ_{X4}	**S9**	μ_{H4}	**S14**	μ_{K4}
S5	μ_{X5}	**S10**	μ_{H5}	**S15**	μ_{K5}

Table 4.3. Indexes and denotations for all shape features.
(The capital letters represent the dimensions (X-Temporal, H-Hip, K-Knee), and the numbers represent the components (from 1 to 5), respectively.)

4.2.5 Feature Analysis and Classification

After the extraction of all variability features and shape features, the features were first analyzed using a two-sample t-test, whereby the p-values were an indicator of the significance of the features. The features were subsequently ranked according to the p-value in ascending order. Afterwards, classifications were performed based on different feature sets: 1) only the variability features; 2) only the shape features; 3) all features; and 4) only the top ranking features. The high-level classifier used for distinguishing the healthy control patterns from the pathological patterns was the SVM. The two classes were denoted as HC and PT for the healthy control group and pathological group, respectively. Additionally, the PCA was deployed as the second feature selection method by projecting the feature matrix from a high-dimensional space to a low-dimensional space, and selecting the most contributive principle components as the input features.

The SVM is regarded as one of the most promising methodologies for solving binary classification problems, and is able to achieve a good performance even with a limited number of training samples. During the training and testing stage, all the features extracted from one pair of gait cycles were regarded as one sample. All the samples, generated from the healthy subjects were labeled as "HC," and all samples generated from the pathological subjects were labeled as "PT." The SVM models were trained with different kernel functions, and each feature vector was centered and scaled by the weighted mean and SD. The kernel scale factor was selected using a heuristic procedure automatically. Five percent of the training samples were discarded during the learning process using robust learning, which removed 5% of the observations that corresponded to gradients that were large in magnitude.

An example of a feature matrix for one subject is illustrated in Fig. 4.10. The input feature matrix is a $N \times n_f$ matrix, where N is the number of samples and is equivalent to the number of pairs N_{pair}, and n_f is the number of features used for the classification depending on the feature set. The output matrix is a $N \times 1$ matrix indicating the actual labels of all the samples. All samples are labeled as 1 if the subject comes from the HC group, and they are labeled as -1 when subject comes from the PT group.

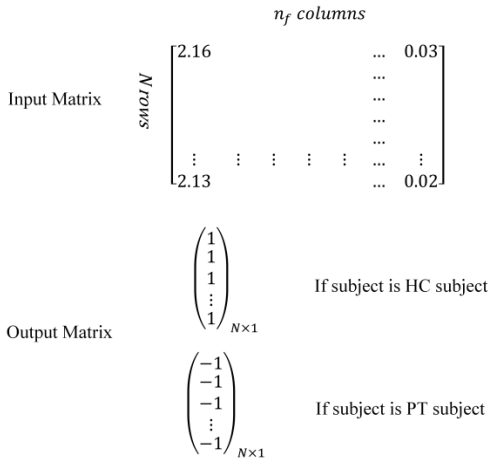


Figure 4.10. Illustration of feature matrix.

4.2.5.1 Majority Voting

The classification performances of the SVM classifiers on samples are evaluated with the accuracy, sensitivity, specificity, and the AUC values separately for each feature set. Afterwards, the final classification results of the subjects are evaluated by a majority voting (MV) approach. The final label of one subject is decided by the majority classification results of all samples, and the confidence score of the final classification of this subject, α, is computed as the percentage of samples of this subject classified as belonging to the major class as follows:

$$\alpha = \frac{\widehat{N}}{N} \times 100\% \tag{4.20}$$

where N is the number of samples, i.e., the number of gait pairs, generated from this subject, and $\widehat{N}$ is the number of samples classified as belonging to the major class, which is the class that has the most predictions. For instance, if 100 samples are generated from a healthy subject, and 60 out of the 100 samples are classified as "PT," then the final label of the classification is "PT," and the confidence score is 60%, and this is indeed a false negative detection.

It has to be stressed that, instead of just providing a class label for each subject, the proposed MV approach is also capable of providing the confidence score. The confidence score is an indicator of the level of confidence, and this indicator is able to provide a new measure for the assessment of walking quality and fro measuring the changes in walking performance; therefore, it is potentially useful for the monitoring of the gait of patients with NDDs. The applications of the score will be discussed in Section 4.4.

4.2.5.2 Leave-one-subject-out Cross Validation

In order to evaluate the classifier's performance, a leave-one-subject-out (LOSO) cross-validation is carried out on the whole data set, so as to assess the algorithm's behavior when evaluating a new subject. For each iteration, the feature set for one subject is reserved as the test set, and all the rest of the subjects' feature sets are used for training the SVM classifier. The test set is passed to the trained classifier, and the final result for the tested subject is reported by the MV approach with an assigned label and a confidence score, both generated from the results on individual samples of the test set. Then, the tested subject's feature set is stored as part of the training set, and the same procedure is repeated on the next subject, until the last subject. The block scheme of the SVM classifier LOSO validation is depicted in Fig. 4.11.

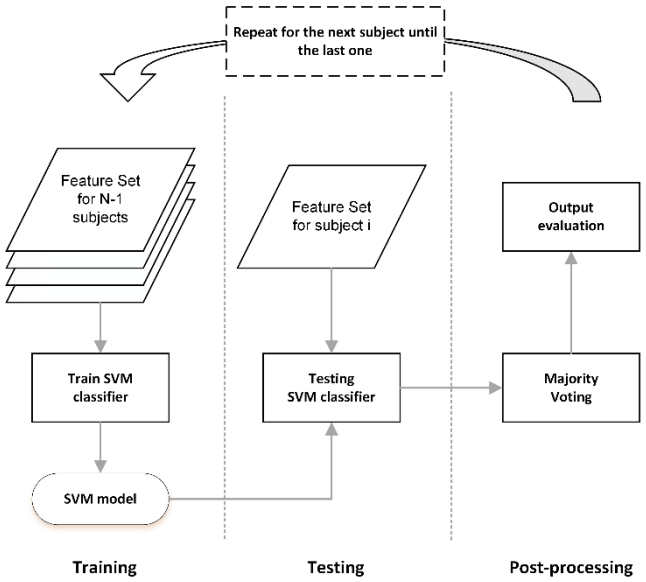


Figure 4.11. Block diagram of the LOSO validation.

The overall accuracy of the LOSO test is the percentage of the subjects successfully classified as their ground truth labels. Sensitivity, i.e., true positive rate, is the percentage of healthy subjects successfully classified as "HC," and specificity, i.e., the true negative rate, is the percentage of pathological subject correctly labeled as "PT," assuming that HC represents the positive class and PT represents the negative class. Apart from those parameters, the average confidence scores for the correctly labeled subjects are calculated for both classes separately, as well as for both classes jointly.

4.3 Experimental Results

In order to evaluate the proposed methods on solving gait classification problems for joint angle signals, an experimental study was carried out using the data collected using the AHRSs from a number of subjects.

4.3.1 Experiment

The gait joint angle signals were collected from 28 HC subjects (age: 34 ± 8, one men and 10 women) and 30 PT subjects with three different types of neurodegenerative diseases. The PT group consisted of six patients with muscular sclerosis (MS) (age: 44 ± 6, two men and four women), 12 patients with Parkinson's disease (PD) (age: 76 ± 4, two men and four women), and 12 patients with polyneuropathy (PNP) (age: 79 ± 6, nine men and seven women). All the patients were mildly affected by the diseases and had provided informed consent as approved by Neuropsychiatricum, Bremen, Germany.

There was an age mismatch between the two groups. However, many previous studies ignored such mismatches and justified their study design by mentioning that the effects of a neurodegenerative disease are much more pronounced than those caused by physiological aging [89] [9] [90].

According to the experimental protocol, all subjects were requested to walk four times at their natural paces along a 70-m straight hallway back and forth on level ground without stopping. No freezing of gait was observed during the experiments with PD patients, and no abnormal gait behavior caused by tiredness or uncomfortableness was identified for all HC subjects and PT subjects. Four AHRSs (FSM-9) were employed for the recording of the hip and knee joint angle signals, using the setup introduced in Chapter 3. The number of gait cycles that each walking trial contained lay in the range of 20 to 50 depending on the walking speed of the subject. After the data collection, all the hip and knee joint signals were processed according to the procedures explained in Section 4.2.

4.3.2 Feature Analysis

4.3.2.1 Statistics on Features

The statistics on the extracted 31 features are reported as the mean and SD in Table 4.4. The significances of the features are indicated by the p-value of the two-sample t-test between the two classes, where a feature with $p < 0.01$ is regarded as a significant feature, a feature with $p < 0.001$ is considered as a very significant feature, and the rest are non-significant features. The two-sample t-test, which is used in statistics to determine whether the means of two groups are equal, has been widely used in gait classification studies (e.g. [59] [66]). The features are ranked according to the p-value in ascending order.

It can be seen from the feature statistics that, only two out of the 31 features (V_{KRO} and V_{KSO}) do not have significant differences between the two groups (HC and PT); all the rest of the 29 features are justified as contributive features, including one significant and 28 very significant features. Those statistical outcomes have shown a successful extraction of features. Besides, more findings can be drawn if we analyze the features in more depth:

- Eight out of the top 10 features are shape features, indicating that the shape features are more significant than the variability features.
- For all of the variability features, the HC group yields smaller mean values, indicating that the HC subjects intend to have on average a smaller variability than the PT subjects with regard to joint movement during walking, which is consistent with the previous research outcomes [91].
- The overall ranking of variability features computed using original data is not as high as that of the features computed using normalized data, with two and five features ranked in top 20, respectively. In particular, regarding variability features associated with the knee joint, only one feature computed from the original data,

V_{KCO}, is scoped as a very significant feature, while all the four features computed from the normalized data are very significant features. This finding further justifies the importance of normalizing data to eliminate the influence introduced by walking speed.

- Among the four distance functions, the features generated using the *MCorr* function has the highest ranking, with three ranked in the top 20.
- All the shape features are very significant features, and among them, the center locations for hip and knee dimensions are ranked the highest, especially for the 4th and 3rd components, which are mainly associates with the late stance and early swing phase.
- For the temporal dimension shape features, the most significant features are for the 4th and 5th components, which mainly correspond to the late swing phase.

In order to understand the features in a more intuitive way, the top four variability features are visualized for the two classes using histograms, as shown in Fig. 4.12. It is evident from the histograms that the distribution of feature values differs between the two groups: the variability of the PT group tends to be larger and have a more dispersed distribution, while the HC group has the opposite distribution.

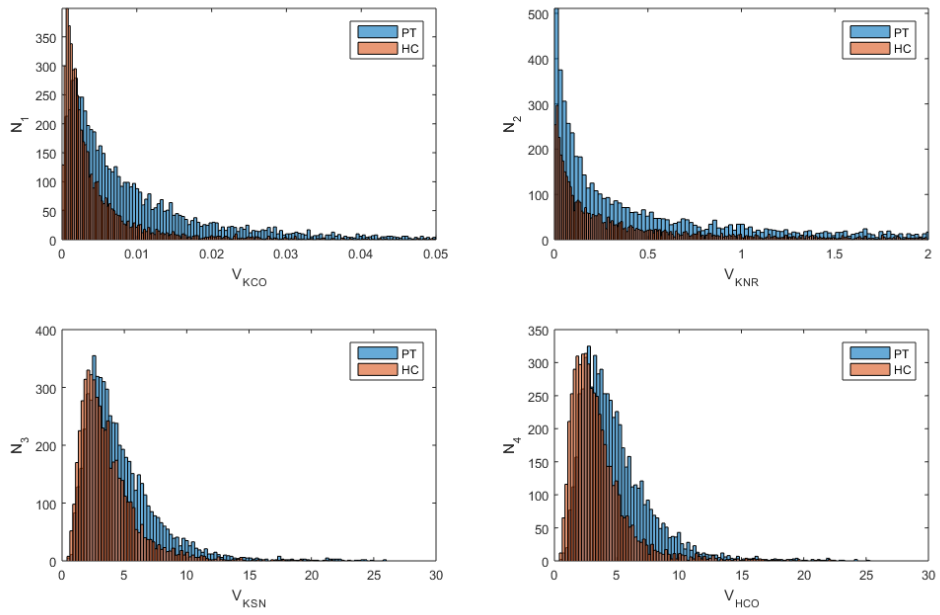


Figure 4.12. Histograms of the most significant variability features for HC and PT classes.

The averaged trajectories are computed by averaging all trajectories for each data point, and are calculated for both classes. They are modeled with GMM, and the centers of all components are plotted in Fig. 4.13. It can be seen from the plots that, the locations of the Gaussian centers have observable differences between the two classes for both the hip and knee joint, and the locations in the temporal dimension of the HC group is always on the right side of the PT group.

Rank	Feature	HC		PT		p-value	significance
		Mean	STD	Mean	STD		
1	μ_{H4}	0.5898	0.1084	0.4647	0.1407	≈ 0	**
2	μ_{K3}	0.1836	0.0736	0.1274	0.0726	≈ 0	**
3	μ_{K4}	0.8539	0.0559	0.7927	0.1163	6.12e-249	**
4	V_{KCO}	0.0023	0.0032	0.0048	0.0043	1.53e-254	**
5	μ_{H2}	0.4266	0.0999	0.4929	0.1057	3.57e-240	**
6	μ_{H5}	0.9488	0.0295	0.9115	0.0828	1.84e-200	**
7	μ_{K1}	0.1396	0.0682	0.1826	0.0790	5.08e-197	**
8	μ_{X5}	0.9136	0.0120	0.9066	0.0123	1.20e-196	**
9	μ_{X4}	0.7275	0.0178	0.7173	0.0185	6.12e-183	**
10	V_{KRN}	4.4924	3.1034	6.3307	3.8962	8.10e-159	**
11	V_{KSN}	3.4915	2.3758	4.8594	2.8415	1.61e-159	**
12	μ_{K5}	0.3876	0.0935	0.4382	0.1044	9.30e-154	**
13	V_{HCO}	0.0050	0.0087	0.0114	0.0162	6.10e-140	**
14	V_{KCN}	0.0025	0.0042	0.0045	0.0040	2.71e-150	**
15	μ_{X1}	0.1026	0.0126	0.0966	0.0131	8.06e-129	**
16	μ_{H1}	0.8683	0.0786	0.8997	0.0593	1.64e-124	**
17	μ_{K2}	0.1222	0.0763	0.1599	0.0928	2.38e-116	**
18	V_{HSN}	2.9211	1.9387	3.8611	2.5932	8.07e-100	**
19	V_{HRN}	3.6352	2.3157	4.7351	3.0683	4.08e-97	**
20	μ_{H3}	0.0754	0.0245	0.0888	0.0428	6.62e-87	**
21	V_{KDN}	2.7762	6.9191	4.9363	7.8447	1.78e-52	**
22	V_{HCN}	0.0019	0.0037	0.0033	0.0063	8.12e-48	**
23	V_{HSO}	1.4854	0.8247	1.7196	0.9930	7.98e-41	**
24	V_{HRO}	1.8048	0.9906	2.0809	1.178	5.46e-40	**
25	μ_{X2}	0.3141	0.0210	0.3088	0.0232	1.73e-36	**
26	V_{HDO}	0.3664	0.6697	0.5023	0.8159	1.89e-21	**
27	μ_{X3}	0.5259	0.0214	0.5235	0.0227	4.06e-09	**
28	V_{HDN}	1.0989	2.7762	1.4416	4.6157	3.12e-06	**
29	V_{KDO}	1.0225	1.8234	1.1124	1.2909	2.41 e-03	*
30	V_{KRO}	2.8007	1.6607	2.8351	1.5199	0.25	
31	V_{KSO}	2.2032	1.2629	2.2174	1.1262	0.53	

Table 4.4 Statistics on features for joint angle signals.

(* indicates the feature is a significant feature. ** indicates the feature is a very significant feature)

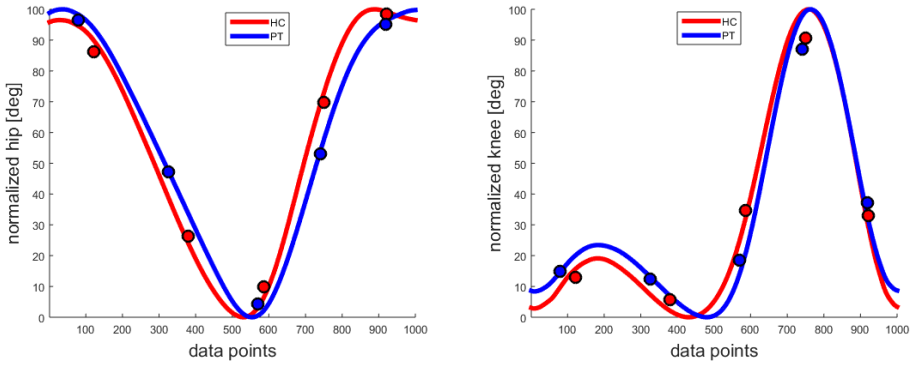


Figure 4.13 Averaged trajectories for HC and PT classes and the centers of Gaussians.

4.3.2.2 Inter-subject and Intra-subject Variability

In traditional medical studies, the intra-subject and inter-subject variabilities are very important factors for the interpretation of the effects of differences in pharmacologic response in different individuals. Intra-subject variability is the difference between walking trials performed by the same individual, while the inter-subject variability is the difference between the walking trials performed by different individuals. In this presented thesis, we would like to investigate whether the extracted features are feasible for representing those variabilities.

The p-values are computed for both intra-subject and inter-subject variabilities for all the subjects in this manner:

- Intra-subject: compute the significance level of each feature for each subject by comparing between different walking trials using the two-sampled t-test, and average the results for each feature.
- Inter-subject: compute the significance level of each feature by comparing each subject to each of the rest of the subjects using the two-sampled t-test, and average the results of all subjects.

The results are summarized in Table. 4.5. Except of two features, S4 and S11, which correspond to μ_{X4} and μ_{K1}, for all the rest of features, the inter-subject variability is more significant than the intra-subject variability, since the p-values are smaller for inter-subject variability. This finding has proved the repeatability of the experiment, i.e., by repeating the experiments for the subjects, the existing differences between walking trials for one subject are always smaller than the differences to other subjects. On the other hand, the features that were extracted have been proved to be feasible for representing the walking performances comprehensively, in the sense of both inter-subject and intra-subject variability.

Feature	p-intra	p-inter	Feature	p-intra	p-inter	Feature	p-intra	p-inter
V1	0.1361	0.0743	**V12**	0.1051	0.0715	**S7**	0.0439	0.0239
V2	0.1007	0.0691	**V13**	0.1139	0.0672	**S8**	0.1678	0.0559
V3	0.1149	0.0741	**V14**	0.1257	0.0467	**S8**	0.1238	0.0193
V4	0.1073	0.0531	**V15**	0.1003	0.0721	**S10**	0.1446	0.0182
V5	0.1898	0.0891	**V16**	0.0538	0.0393	**S11**	0.0038	0.0119
V6	0.1135	0.0663	**S1**	0.1851	0.0949	**S12**	0.1116	0.0156
V7	0.1646	0.0974	**S2**	0.1388	0.0903	**S13**	0.1818	0.0302
V8	0.1326	0.0391	**S3**	0.1636	0.0893	**S14**	0.1776	0.0539
V9	0.1696	0.0774	**S4**	0.1157	0.1519	**S15**	0.0863	0.0362
V10	0.1047	0.0577	**S5**	0.1202	0.0905			
V11	0.1245	0.0793	**S6**	0.1236	0.0144			

Table 4.5. Intra-subject and inter-subject variability.

4.3.3 Classification Results

4.3.3.1 Results on Different Feature Sets

The classification results using three feature sets, i.e., variability features, shape features, and all features, are reported for both classifications of samples and classification of subjects. The performance on classifying samples is measured with accuracy (Acc), true positive rate/Sensitivity (TPR/Sen), true negative rate/specificity (TNR/Spe), positive predictive value/precision (PPV/Pre), negative predictive value (NPV), and the AUC. The performance on classifying subjects is evaluated with Acc, TPR, TNR, PPV, NPV, and the average score for each class separately and jointly. The final results can be found in Table 4.6. The numbers in the brackets, (), are the number of elements for calculating the corresponding parameters, represented by "numerator/denominator". Different kernels are tested as well, consisting of the linear kernel, RBF kernel, and polynomial kernel. The results reported in Table 4.6 are the best ones obtained from all kernels. The classification procedures were repeated for 10 runs, and the average was computed as the final result in order to enhance the robustness.

From the sample-level classification results, we can see that, the accuracy achieved with using only shape features is significantly higher (11.91%) than that achieved by only using variability features. Moreover, the highest accuracy was obtained by using all features, with a very slight difference (0.06%) when using only shape features. The TPR and TNR are the proportions of positives (HC) and negatives (PT) that are correctly assigned as such. For all the three feature sets, the values of TNR are higher than those of TPR, showing that the PT samples are more rarely misclassified compared to the HC samples. This can be reliable in practical applications, since the patients have less probability to be misdiagnosed. The TPR value slight increases when both types of features are combined compared to when only one type is used, while the best results for TNR are obtained with only the shape features. The PPV and NPV are the proportions of true positives and true negatives in all the results that are predicted as positive and negative, respectively. The PPV values are higher than the PNV values for all three

feature sets, which shows that, among all the samples classified as HC, there are very small proportions that actually belong to PT, while for all sample predicted as PT, slightly larger proportion are due to HC actually. This result has proved that the classifier has a strong capability of ensuring a high classification rate of the PT samples. Therefore, in practical scenarios, the probability that the patients will be diagnosed as healthy, delaying treatment is low.

		Variability Features	Shape Features	All Features
Sample-level Classification	Acc	86.93% (2913/3351)	98.84% (3312/3351)	98.90% (3314/3351)
	TPR (Sen)	82.11% (1276/1554)	98.33% (1528/1554)	98.58% (1532/1554)
	TNR (Spe)	91.10% (1637/1797)	99.28% (1784/1797)	99.17% (1782/1797)
	PPV (Pre)	88.86% (1276/1436)	99.16% (1528/1541)	99.03% (1532/1547)
	NPV	85.48% (1637/1915)	98.56% (1784/1810)	98.78% (1782/1804)
	AUC	0.9442	0.9993	0.9994
Subject-level Classification	Acc	89.66% (52/58)	93.10% (54/58)	96.55% (56/58)
	TPR (Sen)	82.14% (23/28)	85.71% (24/28)	92.86% (26/28)
	TNR (Spe)	96.67% (29/30)	100% (30/30)	100% (30/30)
	PPV (Pre)	95.85% (23/24)	100% (24/24)	100% (26/26)
	NPV	85.29% (29/34)	88.24% (30/34)	93.75% (30/32)
	HC score	72.32%	79.55%	85.47%
	PT score	87.34%	95.40%	98.11%
	Overall score	79.59%	87.22%	91.79%

Table 4.6. Classification results on samples-level and subject-level using SVM.

The AUC has shown a very promising behavior of the classifiers, especially the classifiers trained with solely shape features and with all features. The AUC value is an indicator of the diagnostic ability of the classifier as its discrimination threshold is varied, and is an important measure of the classifier. An optimal classifier will yield an AUC value of 1, indicating an always-true prediction for all positive samples. The AUCs achieved from the three classifiers are 0.9442, 0.993, and 0.994 respectively. The classifiers trained with shape features and all features are very close to the ideal classifiers.

The best classification accuracy was achieved using all the features of the subjects, with only two subjects misclassified, while six subjects and four subjects were misclassified when only variability features and only shape features were used, respectively. For the TPR and TNR of the subjects, all PT subjects were correctly classified by using all the features, as well as using only the shape features, while two HC subjects were misclassified out of 28 by using all features. For all the features sets, the proportions of subjects correctly assigned with their desired labels were higher for the PT class than for

the HC class. This is consistent with classification results obtained for the samples, showing that a more reliable detection rate for unhealthy patients compared to healthy patients. The PPV and NPV results of the subjects were in line with the results of the samples as well, indicating that no PT patients were mislabeled as being HC patients, and a very small portion of the HC subjects were mislabeled as PT patients. This outcome ensures that the PT subjects are not wrongly identified as healthy, which is usually important for diagnostic related medical assessments. In addition to those parameters computed from the confusion matrix, the confidence scores were used as important factors for the evaluation of the results. The HC score, PT score, and overall score were the average scores computed from all the true positives, true negatives, and all the trues, respectively. The overall scores significantly increased when all features were used, and the same conclusion can also be drawn for the HC scores and PT scores. By comparing the HC and PT groups, it can be seen that there is noticeable difference, i.e., the scores for the PT class are much higher than those of the HC group. This is correlated with classification results on samples, as the samples of the PT group are more likely to be correctly predicted than those of the HC group.

Overall, very promising classification results were achieved for both samples and subjects. This proved the effectiveness of the feature extraction methods and classification scheme. In particular, the unsupervised machine learning based feature extraction method had a significant effect on the results. The standalone model-fitting features were able to yield a 98.84% and 93.10% accuracy for samples and subjects, respectively. The best results were obtained with the RBF kernel of the three employed kernels for all classification scenarios; therefore it will not be further discussed in this chapter. Moreover, the variability features showed a larger walking fluctuation in joint angle behaviors for the PT group and were consistent with the pre-known medical findings. Instead of directly classifying the subjects, the advantage of the proposed sample generation scheme has been justified by a high classification rate of the subjects. Additionally, there is significant enhancement of the classification performance compared to the state-of-the-art approaches, where only a maximum accuracy of 90% was achieved. Lastly, the confidence score, yielded by the MV approach, is an indicator of the extent to which the subject belongs to its predicted class, and can be potentially used as a measure of the changes in gait quality, and this will be discussed in Section 4.4.

4.3.3.2 Results with Feature Selection

As introduced in Section 4.2, two feature selection methods were deployed to optimize the classifiers, as well as to achieve better classification performances. The first one ranks the original features according to the p-value of the t-test and uses the top n_{f1} features, $n_{f1} = 1, 2, \dots, 31$; the second one transforms the complete original feature matrix using the PCA and uses the top n_{f2} components, $n_{f2} = 1, 2, \dots, 31$, after ranking the components according to their importance. The accuracy of the method in classifying subjects, the scores for the HC group and PT group, the overall performance, and the elapsed time are reported in Table 4.7. The elapsed time is the average time spent completing the validation of one subject using the LOSO approach, i.e., the average time

needed to train a classifier using 57 subjects' data and test the results on one test subject. The algorithm was run with MATLAB 2015 10 times, and the average was regarded as the final result. A computer with an Intel Core i7-4790 CPU at 3.60 Hz, 8 GB of RAM, and 64 bit Win 8 system was used. To visualize the results more intuitively, the results are plotted in Fig. 4.14, with the scores and the accuracy versus the number of elements.

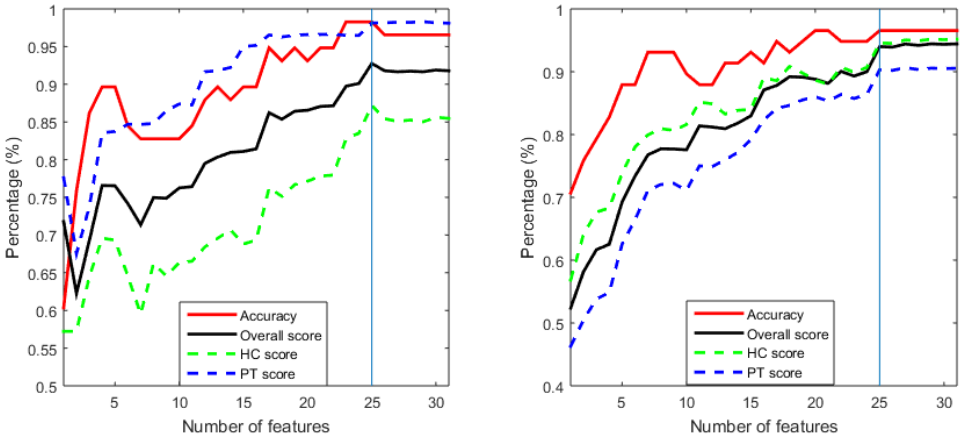


Figure 4.14 Classification performance comparisons for feature selection methods. Left: t-test; Right: PCA.

It is shown in Fig. 4.14 that for the t-test based feature selection method, with an increase in the number of features involved, the scores and the accuracy experienced fluctuations and reached maximums at $n_{f1} = 25$. A slight drop in all four parameters can be observed afterwards in Table. 4.7. The best performance was achieved, therefore, with the top 25 features with the t-test based approach. This approach outperformed the method that used all features, showing the necessity of performing feature selection. For the PCA based feature selection method, the highest accuracy was achieved with the top 25 principle elements, and remained the same as the number of components was increased. The best results for the HC score, PT score, and overall score were obtained when the top 31, 27, and 29 components were used, respectively, but with no significant difference compared to those achieved using the top 25 elements. Hence, it can be concluded that, 25 is the minimum number of components that is able to provide one of the optimal results. Compared to the PCA-based method, the t-test based method provided a higher accuracy with the top 25 features, and the scores for the PT group were higher; hence it is regarded as more optimal, assuming that avoiding the misclassification of PT subjects is one of the main considerations. However, the classifier trained using PCA with top 25 components was more balanced for the two classes, as the scores were closer to each other between the two groups. Regarding the results for the PT group, both feature selection methods correctly predicted all PT subjects using the top 25 elements, indicating the feasibility of avoiding misclassification of patients.

Number of elements	Feature Selection Method									
	T-test					PCA				
	ACC (%)	HC score (%)	PT Score (%)	Overall Score (%)	Elapsed time (sec)	ACC (%)	HC score (%)	PT Score (%)	Overall Score (%)	Elapsed Time (sec)
1	60.34	57.22	77.61	71.78	17.01	70.69	56.84	46.28	52.36	15.10
2	75.86	57.24	67.53	62.24	19.74	75.86	64.10	50.48	58.16	14.79
3	86.21	64.47	73.76	69.25	17.11	79.31	67.64	53.82	61.66	13.50
4	89.66	69.57	83.57	76.57	13.93	82.76	68.33	54.85	62.53	12.57
5	89.66	69.34	83.74	76.54	13.89	87.93	73.84	62.56	69.28	12.09
6	84.48	64.56	84.69	74.18	11.26	87.93	78.04	66.41	73.37	11.77
7	82.76	59.51	84.68	71.35	8.42	93.10	79.94	71.12	76.76	10.60
8	82.76	66.23	84.80	74.97	7.63	93.10	80.94	72.03	77.73	10.17
9	82.76	64.56	86.50	74.88	5.72	93.10	80.66	72.24	77.69	9.38
10	82.76	66.34	87.39	76.25	5.85	89.66	81.58	71.10	77.57	9.52
11	84.48	66.53	87.25	76.43	5.49	87.93	85.16	75.00	81.38	8.81
12	87.93	68.42	91.66	79.50	4.46	87.93	84.89	74.91	81.19	9.31
13	89.66	69.57	91.81	80.34	4.44	91.38	83.26	75.97	80.93	9.58
14	87.93	70.69	92.23	80.96	4.12	91.38	83.86	77.18	81.85	9.48
15	89.66	68.81	95.01	81.09	3.86	93.10	83.98	79.22	82.96	8.11
16	89.66	69.32	95.16	81.44	3.99	91.38	89.05	82.33	87.10	6.31
17	94.83	76.26	96.54	86.23	3.26	94.83	88.56	84.15	87.80	5.45
18	93.10	75.14	96.28	85.37	3.30	93.10	90.84	84.64	89.20	5.16
19	94.83	76.68	96.49	86.43	3.42	94.83	89.81	85.51	89.13	4.96
20	93.10	77.14	96.60	86.55	3.27	96.55	88.65	86.00	88.81	4.55
21	94.83	77.82	96.61	87.06	3.52	96.55	88.01	85.34	88.15	4.51
22	94.83	77.96	96.64	87.15	3.44	94.83	90.74	86.39	90.05	3.69
23	98.28	82.78	96.49	89.75	3.23	94.83	89.93	85.74	89.31	3.98
24	98.28	83.52	96.48	90.11	3.10	94.83	90.66	86.36	90.00	3.50
25	98.28	87.19	98.13	92.75	2.87	96.55	94.56	90.31	93.99	2.69
26	96.55	85.43	98.16	91.79	2.96	96.55	94.52	90.21	93.92	2.70
27	96.55	85.09	98.23	91.66	2.85	96.55	95.06	90.65	94.42	2.74
28	96.55	85.26	98.21	91.73	2.90	96.55	94.93	90.34	94.19	2.65
29	96.55	85.02	98.30	91.66	2.90	96.55	95.17	90.57	94.43	2.65
30	96.55	85.62	98.16	91.89	2.85	96.55	95.10	90.50	94.36	2.71
31	96.55	85.47	98.11	91.79	2.92	96.55	95.18	90.56	94.43	2.71

Table 4.7 Classification results using two feature selection methods.

Regarding the elapsed time of completing one validation process of one subject, a remarkable decrease can be observed from Table 4.7 when the number of elements increases, for the two feature selection methods. When 25 elements were used for the classification, the elapsed time dropped to below 3 s. It has to be stressed here that, the elapsed time computed here is composed of the time for two steps, namely, the training phase and the testing phase, and in principle, the training phase takes a larger portion of the elapsed time. The average elapsed time for testing the data of one subject is 0.026 s and 0.012 s for the t-test and PCA based feature selection methods, respectively. This indicates the high potential of enabling the algorithm in real time, as in real time applications, the model is pre-trained and validated, and only the testing phase is needed.

In summary, better results were obtained with both feature selection approaches than with all features, and the optimal results for both methods were achieved using the top 25 elements. The overall results indicate the feasibility of both feature selection methods.

4.4 Applications to Patient Management and Rehabilitation

Except for the LOSO validation performed in Section 4.3, in this section, four case studies were carried out in in order to further evaluate the effectiveness of the machine learning framework, especially the outcomes of the framework, i.e., the classification label and confidence score, for possible applications to patient management and rehabilitation. Firstly, an experiment was conducted on heathy subjects by simulating the impaired gait, and the differences between the walking conditions with various constraints were compared based on the classification results. In the second study, patients diagnosed with PD were recruited to perform walking tasks before and after taking medications that reduce the symptoms of the patients. The performances were compared based on the classification results for evaluating the possible usage of the resulting scores on measuring medication influences. For the third case study, a patient diagnosed with hereditary spastic paraplegia (HSP) receiving physical therapy was monitored for a long time, and the walking differences between the recording phases were analyzed according to the framework, for the sake of proving the feasibility of the framework for long-term gait monitoring. In the last case study, the walking data of a subject using a state-of-the-art gait rehabilitation system, MOPASS, designed for supporting patients with gait impairments by actively providing torque and force support on the pelvis and legs, was recorded and analyzed using the framework.

4.4.1 Classification of Simulated Impaired Gait

A classification study was done on three healthy subjects while simulating impaired gait. Each subject was equipped with the IMUs on the left leg, with one attached on the thigh and one placed on the shank. In the first part of the protocol, he/she was requested to walk along the 70-meter long hallway with his/her preferred speed. The IMUs recorded the so called “normal” gait signals for each subject. In the second part of the study protocol, subjects were additionally equipped with a functional orthosis on the right leg, as illustrated in Fig. 4.15. Subjects were requested to walk in the same walkway for the same distance so that the IMU-sensors recorded the gait trajectories corresponding to different gait constraints as the functional orthoses that introduced 40° and 60° of knee constraint, respectively.

The hip and knee signals recorded using the IMUs were processed, and classification was done using the optimal SVM model trained using all the data from the 58 subjects mentioned in Section 4.3. The final results are summarized in Table 4.8 with the classification label and confidence scores for all the three tests, i.e., normal, 40° of constraint, 60° of constraint. The trajectories of all segmented hip and knee cycles of subject 1 were plotted in Fig. 4.16 for all the three tests.

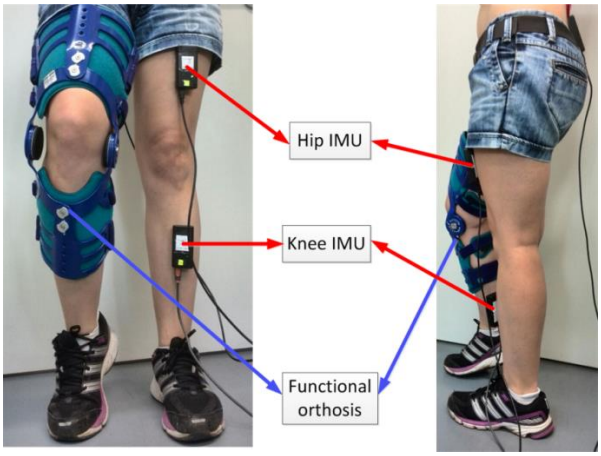


Figure 4.15. Experimental setup with functional orthosis and IMUs.

	Normal		40 degrees		60 degrees	
	Label	**Score (%)**	**Label**	**Score (%)**	**Label**	**Score (%)**
Subject 1	HC	83.53	PT	77.09	PT	96.30
Subject 2	HC	59.35	PT	66.01	PT	75.52
Subject 3	HC	77.19	PT	66.61	PT	82.23

Table 4.8. Classification results on simulated impaired gait.

The results reported in Table 4.8 showed that, for all the three subjects, the SVM model successfully assigned their "normal" walking with the HC label, while for the walking tests with 40° and 60° of constraint introduced by the orthoses, the classification results suggested PT as the final labels. The orthosis was introduced because we wanted to limit the movement range of one leg, which would considerably affect the movement pattern of other leg strongly. The loss of symmetry of the two legs would cause a decrease in the balance level, and in turn increase the variability of the free leg. Therefore, the left leg, which is measured, is supposed to have more pathological behavior theoretically. Even though there are no fixed values indicating the level of impairment, the level of variability in the gait cycles and the shapes of the trajectories, however, could be observed and evaluated from the plots, such as those in Fig. 4.16. The shapes of the knee cycles, as can be observed from the lower three plots, were highly uniformed for free walking, and became more and more chaotically distributed when the subject walked with 40° of constraint and 60° of constraint. For the walking test with 60° of constraint, the knee had very significant negative bending caused by the large difference in height between the two legs. Besides, the shapes of hip cycles' trajectories significant changed after the subject wore the orthosis: instead of having smooth curves, the hip joint's motion was very rigid. The results of the subject in Table 4.8 are therefore believed to be consistent with the observations plotted in Fig. 4.16, indicating that the impaired gait patterns simulated with a functional orthosis were successfully predicted with the trained SVM model using the proposed framework.

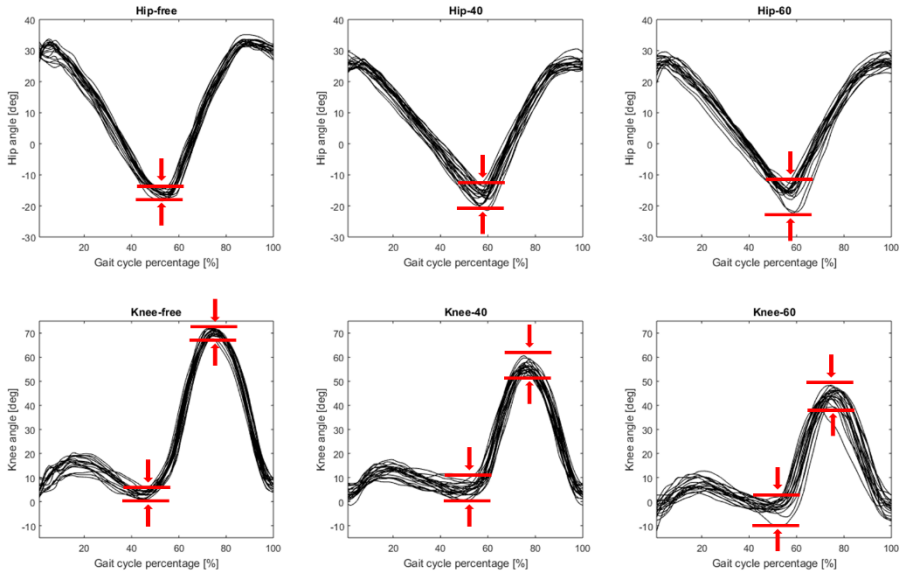


Figure 4.16. All the hip and knee cycles for three walking tests of one subject.

Overall, the results of this study showed the high potential of the machine learning framework to measure the level of impairment of gait, as well as to detect changes in walking quality.

4.4.2 Evaluation of Medication Effect

We would like to investigate the possible usage of the machine learning framework in evaluating the changes in gait quality introduced by medication, which we consider to be one of the most important applications of the proposed framework. This is beneficial to clinicians and therapists, as it may help inform decisions regarding the type of medication to prescribe and strategies for optimizing the medication. The gait quality of a patient is traditionally evaluated through self-evaluation forms completed by the patients at specific times. Patients try to walk and do other movements that are desired, and then they are asked to evaluate their performance according their own feelings. A standard form for evaluating the daily walking gait and mobility designed for a PD patient provided by Paracelsus-Klinik Bremen is shown in Fig. 4.17. This method is very subjective and can sometimes be inaccurate. However, the proposed machine learning approach, along with the resulting label and score, is able to provide objective, numerical measures, which can be used as additional information that can help clinicians better understand their patients' gaits.

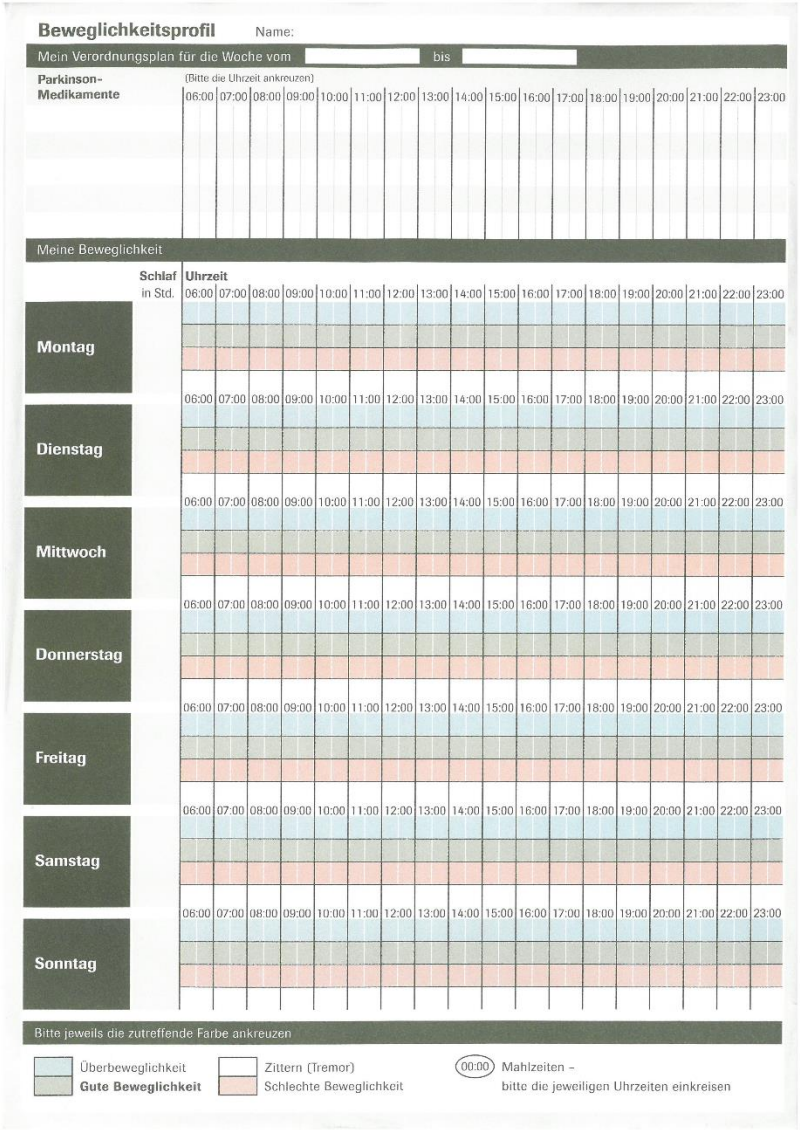


Figure 4.17. Standard form used in clinics for evaluating the daily gait quality of PD patients.

Four patients diagnosed with PD participated in the experiment, and the same protocol as introduced in Section 4.3.1 was employed. The walking data were recorded for different phases, before and after taking the medication was taken. The medication, L-dopa, is able to relieve the symptoms of PD, such as tremor, and increase the mobility of the patient for a short period of time. Along with the data recorded using IMUs, the self-evaluation of the walking performance from the patients themselves, denoted with numbers 1, 2, 3, 4, and 5, which represented very bad, bad, normal, good, and very good, respectively, were documented as well. The full results of the four patients, consisting of the history (number of years after first diagnosis), label and score obtained using the machine learning framework, and self-evaluations, are listed in Table 4.9. For each phase of recording, four walking trials were requested, and only the one with the best performance (smallest number of PT score) obtained with the framework was reported.

Subject: PD1				
Age: 80 **History:** 12 years **Gender**: Male				
	Before	**15min after**	**30min after**	**60min after**
Label	PT	PT	PT	PT
Score (%)	97.8	71.19	87.89	98.39
Self-evaluation	1	3	2	1
Subject: PD2				
Age: 70 **History:** 1 years **Gender**: Male				
	Before	**20min after**	**50min after**	
Label	PT	PT	PT	
Score (%)	81.72	72.99	67.85	
Self-evaluation	1	3	4	
Subject: PD3				
Age: 72 **History:** 2 years **Gender**: Male				
	Before	**10min after**	**30min after**	
Label	PT	PT	PT	
Score (%)	84.23	96.80	76.38	
Self-evaluation	1	1	3	
Subject: PD4				
Age: 78 **History:** 10 years **Gender**: Male				
	Before	**40min after**	**80min after**	
Label	PT	PT	PT	
Score (%)	97.57	90.33	99.26	
Self-evaluation	1	3	1	

Table 4.9. Evaluation results of medication effect on PD patients using the framework.

From Table 4.9 we can see that, for subject PD1, the results obtained using the machine learning framework suggested an improvement in gait performance 15 min after taking the medication, and a decline after 30 min and 60 min, while the self-evaluation provided by the patients suggested the same changing trend. The decline in walking performance may be caused by the fading of the medication effect, or the tiredness of patients, and this requires medical explanations and is beyond the scope of this thesis. For the second subject, the gait evaluation suggested by the proposed framework is consistent with the patient's self-evaluation, namely, significant improvement can be observed for both phases after treatment. Medication starts to have significant impact only from the third phase (30 min after) for PD3, while the patient's self-evaluation suggests the same conclusion. As for the last patient, PD4, a small improvement occurred in the second phase, and faded by the third phase. A similar phenomenon was proved by his self-evaluation. Additionally, if we compare the level of illness among the patients, the two patients first diagnosed 10 years prior to their involvement in the experiment, PD1 and PD4, had generally higher scores than the two patients who had been diagnosed only 1 to 2 years prior, which is logically in line with the medical expectation.

This study showed the capability of the framework to evaluate changes in gait performance introduced by medication for patients with neurological disorders. The label and score can provide clinicians and therapists with additional parameters and can supplement the traditional self-evaluation based approach. Those parameters are numerical values generated from advanced machine learning algorithms, and are

objective and sensitive to small changes in gait. However, those parameters are not meant to give a diagnosis, and should only be used along with other clinical measures for supporting and optimizing the management of patients' disorders.

4.4.3 Long-term Gait Monitoring

To investigate the capability of the framework to monitor the gait performances for long-term usages, a third study was carried out on a patient diagnosed with a rare genetic disease, hereditary spastic paraplegia (HSP), which is an inherited disease whose main feature is a progressive gait disorder. The patient was involved in a 12-month treatment session at home using stochastic resonance therapy (SRT), which allows the patient to exercise with variable vibration stimuli. SRT has been shown to reduce neuronal activity, and by this, the increased muscle tone, which causes gait disturbance in HSP when applied for a longer time, can be reduced by habituation of muscular reflex responses. Fig. 4.18 is an illustration of the HSP patient during SRT-based exercise.

During the SRT treatment period, the patient was not on medication. The HSP patient performed SRT three times per day for approximately 15 min each time. For the purpose of presented case study, the hip and knee joint angles of the patient were recorded while the patient walked on the ground in a straight line at a natural walking pace freely chosen by the patient using the IMUs and processed using the framework for four different therapy phases: just before the start of the exercise-based intervention (Pre), two, four, and 12 months after the start of the SRT (2M, 4M, and 12M respectively). For the purpose of evaluating the proposed framework, as well as for the purpose of clinically investigating the patient's progress besides the sensor-based recording of gait parameters, in each recording phase, the patient was requested to perform two additional standard clinical tests: the 10 m test and the "The timed up and go" (TUG) test. For both tests, the time it takes the patients to complete required walking or reaching tasks is measured, with a shorter time indicating better mobility and hence better gait quality.

Figure 4.18. HSP patient during SRT-based exercise.

The classification results of the HSP patients for all the four phases, the standard gait parameters, and the results obtained from standard clinical tests are listed in Table 4.10. The stride time and ROM of the knee angle are reported as mean (SD), and are computed from all the gait cycles of the four walking trials for each phase; the results for 10 m and TUG are the best results achieved for the four trials.

		HSP Pre	HSP 2M	HSP 4M	HSP 12M
Classification results	**Label**	PT	PT	PT	PT
	Score (%)	92.02	63.11	57.03	61.03
Standard gait parameters	**Stride Time(s)**	1.525 (0.078)	1.090 (0.035)	1.055 (0.034)	1.044 (0.059)
	Knee ROM(°)	46.34 (1.72)	54.38 (1.69)	63.40 (2.14)	67.79 (1.68)
Standard clinical tests	**10m (s)**	5.0	3.6	3.6	3.6
	TUG(s)	5.4	3.6	4.5	4.6

Table 4.10. Classification results and clinical test results on HSP patient during SRT.

The stride time and ROM indicated very significant improvement in gait performance after the first therapy session (2M), which was maintained afterwards (4M and 12M), as the patient was able to achieve much higher walking speeds, and a much larger range of movement for the knee joint at 2M compared to at Pre, while maintain the same level with smaller improvements for 4M and 12M. The standard clinical test results showed similar changes, namely, a larger improvement at 2M compared to at Pre, and a smaller improvement at 4M compared to at 2M, and at 12M compared to at 4M. The classification results obtained from the machine learning framework indicate that the patient had a very high PT score before the SRT intervention; at 2M, a very significant decline in the PT score was observed, and at 4M, a small decline was observed. At 12M, the PT score remained fairly constant.

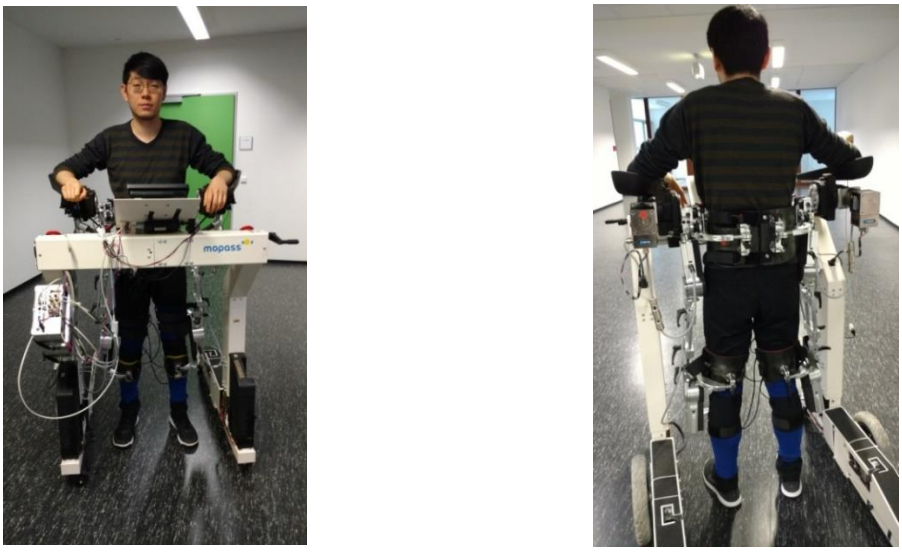
It can be concluded that, each type of measure was consistent with the other on evaluating gait quality, and the measures added up all together to build a more comprehensive long-term profile of the patient, especially for patients like the HSP patient involved in this study, who are not under the observation of clinicians very often. For these patients, the framework is able to provide an efficient, fast, and objective evaluation of gait quality remotely, especially under home conditions.

4.4.4 Application to Gait Rehabilitation System

MOPASS [92] is a state-of-the-art gait rehabilitation system developed by the Institute of Automation, University of Bremen, Germany. It aims to support patients with gait rehabilitation, and is able to cope with the individual needs of the patients during the

walking training. The system contains a mobile platform, which provides support for the subject's pelvis while the subject is walking, and an orthosis that can be attached to both legs to provide force and torque supports to compensate for gait limitations. A study was conducted using this system in order to determine whether the framework could be integrated into rehabilitation systems. The walking data were collected on a healthy subject who was performing slow walking with and without the MOPASS platform. The pelvis module was enabled during the walking platform, while the orthosis was attached to both legs, but no forces and torques were applied. The hip and knee signals were recorded using the IMUs attached on the leg. The whole setup with MOPASS system can be seen in Fig. 4.19.

It is known that, it is usually difficult for people to control their stability very well while walking at a slow speed. The same applies to patients, who naturally have comparably slower walking speeds than healthy subjects. The platform is able to secure the balance of the subject by stabilizing the pelvis while enabling enough range of freedom, and therefore can reduce the walking variability and prevent fallings. The average stride times for slow walking with and without the platform are 1.49 s and 1.64 s, respectively, which are much longer than the subject's natural stride time, 1.09 s. The hip and knee joint angles for all cycles with and without the MOPASS platform are depicted in Fig. 4.20. It can be seen that the variability of the gait cycles is higher for walking without the platform, and after using the platform, the gait cycles are more uniform in shape and better stability can be confirmed. Moreover, the scores generated from the machine learning framework are 64.1% and 76.92% for without and with the platform, respectively, with the HC label for both conditions. The framework showed in this study has the potential capability of evaluating changes in gait quality during the rehabilitation progress for patients using robotic rehabilitation systems. The parameters, as well as the models, can be interpreted in the future in certain ways, so that can they can be used as input to the rehabilitation systems for numerically visualizing and monitoring the whole rehabilitation process.



(a) (b)

Figure 4.19. MOPASS gait rehabilitation system from frontal and back sides.

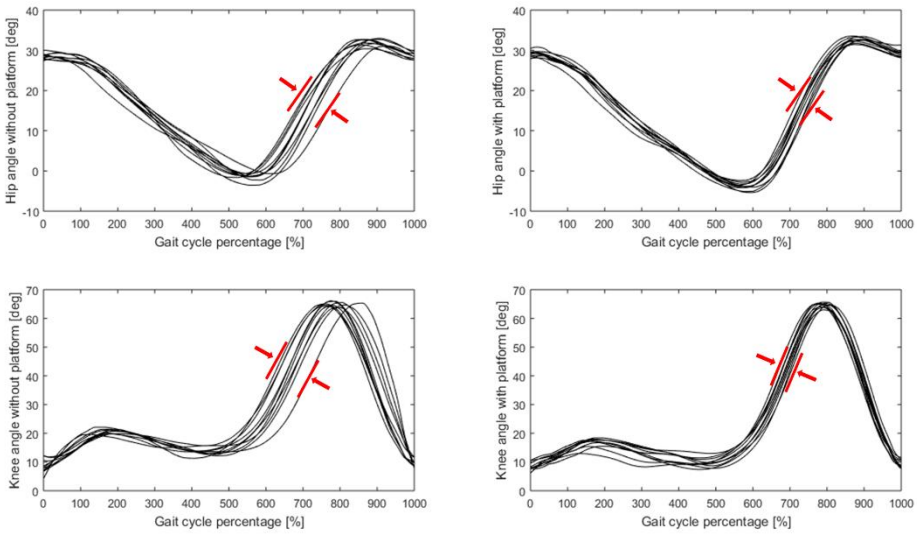


Figure 4.20. Joint angles of subject performing slow walking with and without MOPASS platform.

4.5 Conclusions

The solutions for gait classification using joint angel signals were presented in depth in this chapter, based on the proposed machine learning framework. The novel classification scheme described in this chapter took advantage of combining two machine learning algorithms, and is also capable of providing a numerical score as an indicator of gait quality with a MV approach. The framework has been validated on 58 subjects, and the significance of the extracted features, as well as the efficiency of the classifier has been carefully discussed. The validated framework was introduced to four practical application scenarios, and all showed promising results. The framework is proved to be novel, robust, and efficient, and is able to overcome the previous limitation caused by walking speed and a small number of subjects. The testing phase of the framework is computationally effective so that it can be realized for real time applications. The framework can be used as a standalone platform for assessing and monitoring gait quality, and also can be used as an additional input by clinicians and therapies for supporting decision-making and patient management.

5. Gait Classification for Trunk Acceleration Signals

In this chapter, the machine learning framework proposed in Chapter 3 is deployed to solve gait classification problems on signals of trunk acceleration. The trunk acceleration signals are data corresponding to the movement of the trunk usually collected from the back of the waist, and can reflect the walking stability and balance.

This chapter is organized in the following manner: related work, including the characteristics of the signals, and the state-of-the-art gait classification methodologies are introduced in Section 5.1; the proposed gait classification solution based on the framework is presented in depth in Section 5.2; an experimental classification study is carried out in Section 5.3 to validate the proposed framework and solution using data collected from 54 subjects; and the potential application of this proposed framework and solution is validated with case studies on patients with neurodegenerative diseases under medical treatment in Section 5.4.

5.1 Related Work

Walking balance and stability are one of the most essential factors for analyzing human gait, as the displacement pattern of the trunk, especially the CoM point, may be considered as the end result of all forces and torques affecting the body while waking from one point to another [93]. The movement of CoM is traditionally investigated using the segmental analysis method, the force plate method, and the sacral method. A comprehensive review and comparison of the traditional methods can be found in [94]. In principle, those methods all have the main drawback that they are restricted to laboratory conditions, and are therefore expensive, time consuming, and not optimal for continuous measurement of body movement. Recently, taking advantage of the rapid development of MEMS technology, wearable sensors, such as AHRSs, have been widely adopted in the latest gait and balance studies. R.P. Hubble et al. have reviewed comprehensively the usage of wearable sensors for assessing both standing and walking balance, and concluded, with strong support, the use of the wearable sensors for detecting differences in walking balance between PT and HC subjects [95]. The IMU/AHRSs are commonly attached on the back of the waist using the setup explained in Chapter 3, and the accelerometer data of the trunk are usually collected for analysis.

5.1.1 Trunk Acceleration Signals

The signals analyzed in this chapter are the acceleration signals collected from the AHRSs attached on the back of the waist, which measure the movement of the trunk. The accelerations in three dimensions correspond to the movement of CoM in three directions, i.e., the ML direction, AP direction, and vertical (V) direction, respectively. The orientations of the IMU and its correlation with the human autonomy planes and directions have been introduced in Chapter 3. An example of the acceleration signals in three directions collected from one PD patient walking over ground along a straight corridor for 20 seconds is plotted in Fig. 5.1.

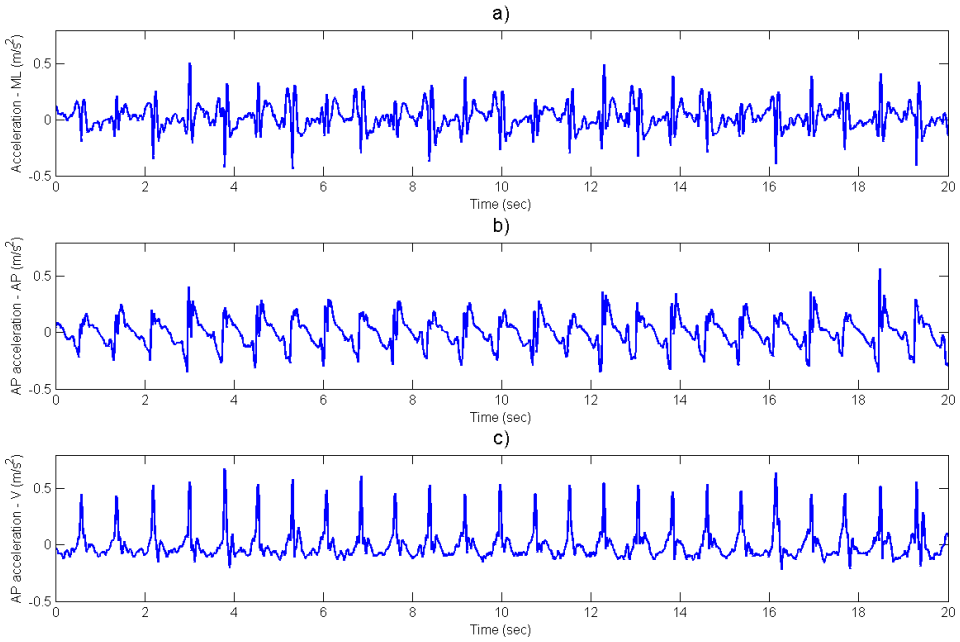


Figure 5.1. Example of trunk acceleration signals collected from a PD patient.

It can be observed from the plots that the trunk sway movement is semi-periodic, and involves movement in all the three directions. There are numbers of peaks and valleys in the signals, which are temporarily related to important events in the gait cycles.

5.1.2 State-of-the-art and Limitations

The acceleration signals have been widely studied recently for understanding the walking behaviors of human, particularly for patients with gait disorders. The standing balance and walking balance were analyzed in [96] using a gait cycle decomposition method, with which the temporal parameters, such as swing time and stance time, were extracted,

and statistically analyzed to distinguish healthy subjects and those with Alzheimer's disease. A similar study can be found in [97], where the trunk acceleration signals were characterized for PD patients by extracting traditional gait parameters, and the PD group and HC group were shown to have significant differences in those parameters. The main limitations of this study were that, firstly, the signals were primarily used for extracting gait parameters for further analysis, yet the results of any analysis relies highly on the accuracy of the extracted parameter; secondly, traditional gait parameters are highly associated with walking speed; therefore, they may not be suitable for classification purposes, which should also eliminate the influence of walking speed, as explained in Section. 4.1.2.

M. Yoneyama et al. have proposed a novel method for quantifying the gait performance of PD patients using the relationship between the gait cycle and the acceleration of the vertical trunk [98]. This approach shows promise in finding the difference between the two concerned groups; however, since the indices were proposed on a subject-level, this method may not be suitable for classification purposes that require a large number of samples for training, based on the number of subjects, which is 27 in this study.

Recently, M.J. Floor-Westerdijk et al. proposed to double integrate the acceleration signals to get the displacement of the CoM for analysis [94]. This approach is too complex for a classification study since the error introduced during the recording of data would be integrated with the acceleration data, which will have a negative impact on the quality of features to be extracted.

A comprehensive assessment of the signals has been done in [38] and [99], where features were extracted from time, frequency, and time-frequency domains in three directions. The study deployed popular statistical, informatics, and signal processing techniques, and compared the features obtained using them in detail for the three groups, i.e., PD, peripheral neuropathy, and HC. Based on these state-of-the-art studies, we would like to further conduct a classification study using some of the popular features proposed in them. Additional machine learning features will be extracted, and served to the classification framework to distinguish the PT and HC walking patterns.

5.2 Gait Classification Using Time-domain, Frequency-domain and Contour Features

The flow chart of the classification procedures for trunk acceleration signals is depicted in Fig. 5.2 based on the general framework proposed in Chapter 3.

Given the acceleration in ML, AP, and V directions as raw input data, the algorithm first applies pre-processing methods on data from all of the three dimensions, consisting of noise removal, artifact compensations, and normalization. These procedures are repeated for the fourth dimension, namely, the magnitude of the acceleration, calculated as the square root of the three signals. A segmentation method is deployed in the next step to detect and segment the signal into separate steps, and a sliding window approach is utilized afterwards to group the steps into windows as samples (observations) for feature

extraction and classification. Features are extracted from both time and frequency domains for all the four dimensions. In addition to those features, model-fitting features are extracted by modelling the ML–AP acceleration diagram with an ellipse using the least square criteria, which is a standard method in regression analysis. All the features are ranked using the t-test, as well as transformed using PCA, before being served as input features to the high-level classifier for final gait classification. SVM and ANN are employed in this chapter as the high-level classifiers for distinguishing HC and PT groups, as well as the PD and PNP groups. The Holdout method and LOSO validation are used to validate the classifiers for samples-level classification and subject-level classification, respectively.

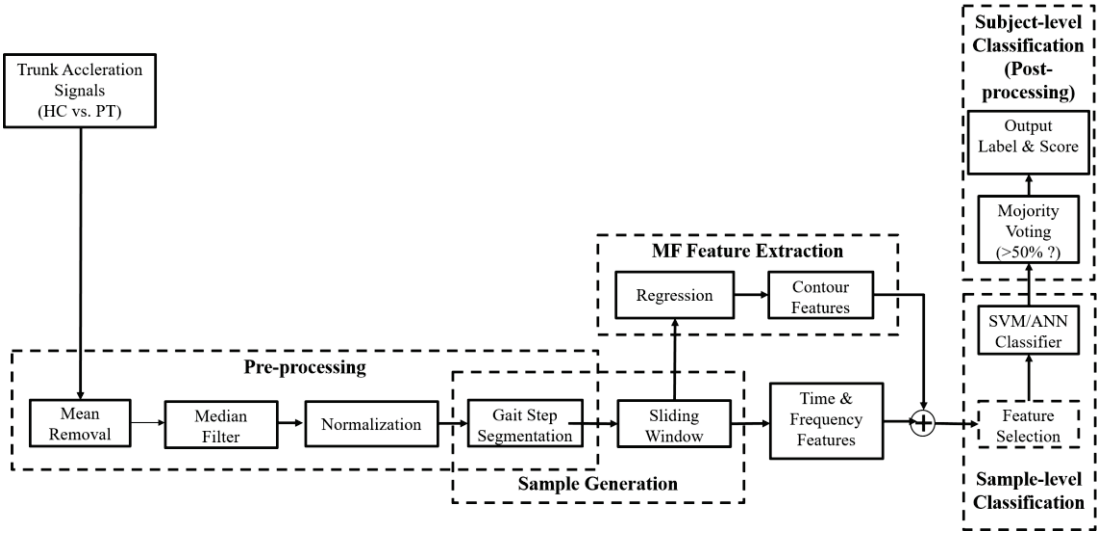


Figure 5.2. Flowchart of the classification scheme for trunk acceleration signals

5.2.1 Data Pre-processing

The raw data collected from the AHRS (FSM-9) can be noisy and easily affected by the artifacts related to gravity and mismatch of orientations. Therefore some procedures are necessary before performing data segmentation.

The artifact related to gravity is removed from the signals by removing the mean from the corresponding signals:

$$a_{ML1}(k) = a_{ML}(k) - \mu_{ML} \tag{5.1}$$

$$a_{AP1}(k) = a_{AP}(k) - \mu_{AP} \tag{5.2}$$

$$a_{V1}(k) = a_{V}(k) - \mu_{V} \tag{5.3}$$

where $a_{ML}(k)$, $a_{AP}(k)$, and $a_V(k)$ are the acceleration signals from the ML, AP, and V directions, respectively, and μ_{ML}, μ_{AP}, and μ_V represent the calculated means of those signals; k is the time step.

In the next step, the impulse-like artifacts unrelated to gait are removed with a 5th order median filter:

$$a_{ML2}(k) = \Lambda\{a_{ML1}(k), m\} \tag{5.4}$$

$$a_{AP2}(k) = \Lambda\{a_{AP1}(k), m\} \tag{5.5}$$

$$a_{V2}(k) = \Lambda\{a_{V1}(k), m\} \tag{5.6}$$

where $\Lambda\{\cdot, m\}$ is the median filter operation with the number of order $m = 5$ in this chapter.

The filtered signals are normalized to unity amplitudes to reduce the inter-subject variability within groups:

$$a_{ML3}(k) = a_{ML2}(k)/max|a_{ML2}(k)| \tag{5.7}$$

$$a_{AP3}(k) = a_{AP2}(k)/max|a_{AP2}(k)| \tag{5.8}$$

$$a_{V3}(k) = a_{V2}(k)/max|a_{V2}(k)| \tag{5.9}$$

The magnitude of acceleration is defined as:

$$M_{acc}(k) = \sqrt{a_{ML}^2(k) + a_{AP}^2(k) + a_V^2(k)} \tag{5.10}$$

where the three component signals are filtered firstly with the median filter before computing the magnitude.

Figure 5.3 shows a sample signal in the V direction before and after each pre-processing step.

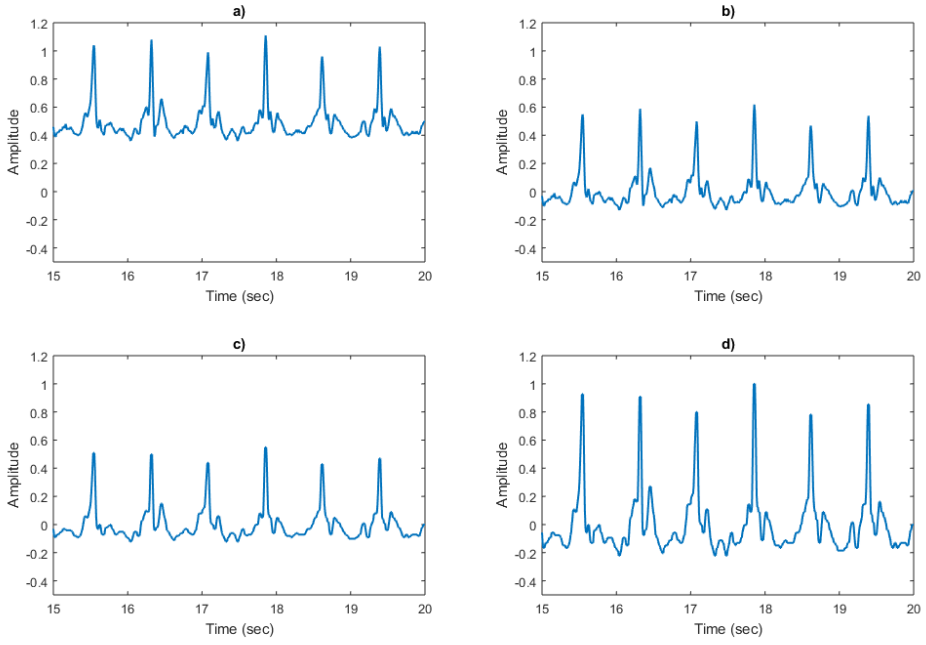


Figure 5.3. Sample acceleration signal in V direction. a) raw signal; b) signal after removing mean component; c) signal after applying median filter; d) signal after normalization. The unit for signals in a) – c) is m/s^2.

5.2.1.1 Gait Step Segmentation

Unlike the joint angle gait signals, from which the data are segmented into gait cycles, for acceleration signals collected from the trunk, data are segmented into steps. This is because each joint angle signal corresponds to the data from only one side, while the acceleration signal collected from the back waist corresponds to movement of the whole trunk, and reflects the movement from both sides. One stride, or temporally speaking, one gait cycle, contains two steps, one performed by each leg, and both steps contribute to the signals.

The step segmentation algorithms have been lately proposed in [100] and [101]. The method utilized in this chapter is based on the method proposed in [100], in which E. Sejdic et al. proposed that heel strike incidences can be detected from the local minima points in the acceleration signals in the AP direction. This criteria is considered in this chapter as the base method owing to its low complexity and high performance, and it is further enhanced in order to achieve better performance for our data: the peak detection method introduced in Section 4.2.1.1 is used to detect all possible peaks, i.e., the local minima, from the pre-processed V direction acceleration signal, and only the peaks that fulfill the flowing criteria are regarded as the correct detections: 1) the distance between the current peak and the subsequent peak is larger than 0.5 times the estimated average distance between two random peaks; and 2) the acceleration value at this peak point is larger than –0.2. These two conditions are generated from experience, and proved to be helpful for the robustness of the segmentation algorithm. An example of the segmentation results on a fragment of acceleration data points in the V direction is illustrated in Fig. 5.4, with the red crosses indicating the detected heel strike incidences.

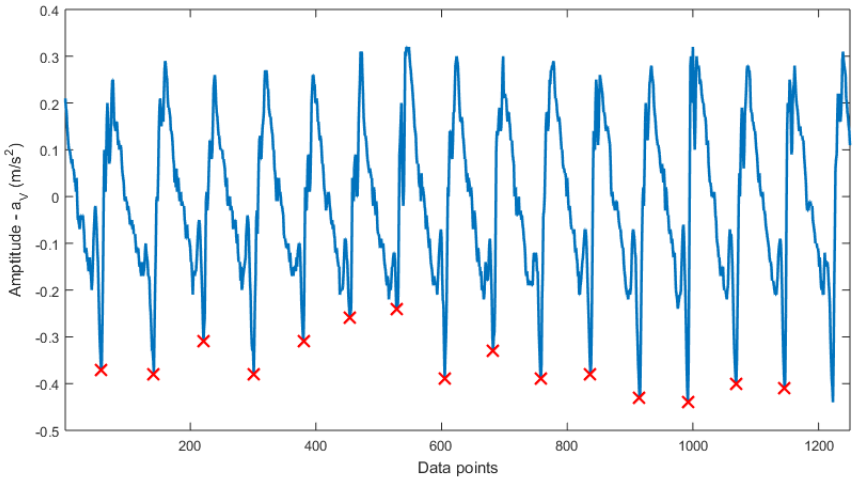


Figure 5.4. Example of step segmentation results on a trunk acceleration signal in vertical direction.

5.2.2 Sliding Window for Sample Generation

After the segmentation of steps, the next step is to group the steps to generate samples for classification. Different from the approach introduced in chapter 4, where the gait cycles are paired, in this chapter, the steps are grouped using a sliding window approach based on the characteristic of the data being processed. The sliding window is famous approach that runs through the signals to produce an underlying selection of data with a fixed window size. Instead of using windows with a fixed size, a fixed number of steps is selected as the main grouping method in this study. More precisely, the window is defined as the data fragment containing four continuous steps, i.e., two steps, or one complete gait cycle from each leg. The actual temporal lengths of the windows may differ from each other while sliding through the signal, depending on the length of the involved steps. There is a 0.25 overlap between two adjacent windows, namely, the window moves ahead for one step in each iteration. Each window is regarded as one sample for performing feature extraction and sample-level classifications. An illustration of the sliding window approach for sample generation can be seen in Fig 5.5, where only the first three windows are visualized.

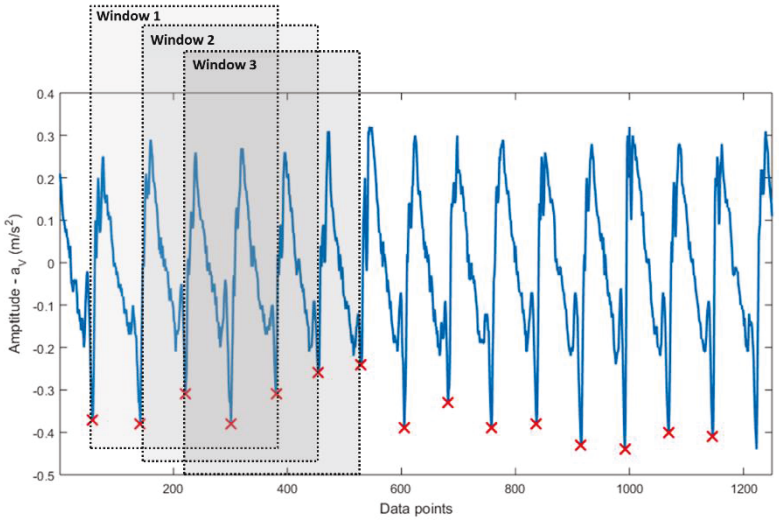


Figure 5.5. Sliding window for sample generation.

The main purpose and advantages of this proposed approach are: 1) the window size is fixed by the number of steps, rather than the length of data, so that each window contains the same amount of information, which corresponds to four actual steps performed by the subject. This is superior to using a window with a fixed length, which may contain a random amount of information, especially in the starting and ending phase; 2) the 0.25 overlap between windows ensures a full utilization of data, and prevents the loss of data or the underestimating of important information. Besides, this approach enables a large number of samples to be generated; and 3) by grouping four steps in one window, a larger scale of data is provided for extracting frequency domain features, which usually requires the length to be statistically representable.

5.2.3 Time and Frequency Domain Feature Extraction

Features are extracted by investigating the most important aspects of the windows from both the time and frequency domains, and the time–frequency domain using common statistical and signal processing algorithms. In total, 15 features are extracted from one window of one dimension.

Statistical Features

Six statistical features in the time domain are extracted from each window. The first four are the absolute maximum, absolute minimum, absolute mean, and the SD. They are defined as following:

$$Max = |\max(a(k))| \tag{5.11}$$

$$Min = |\min(a(k))| \tag{5.12}$$

$$\mu = \left|\frac{1}{n}\sum_{k=1}^{n} a(k)\right| \tag{5.13}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(a(k) - \mu_X)^2} \tag{5.14}$$

where the $a(k)$ is the current window in the acceleration signal. All features introduced in this section are extracted from all the four dimensions, i.e., ML, AP, V, and M.

Except for the standard statistical features, Max_X, Min_X, and μ_X, the standard deviation σ_X is an important feature that measures the spread of the signal aptitude distribution and its squared value is related to the alternating current (AC) power. A larger value of σ_X indicates that the data points are spread over a larger interval of values.

Two additional statistical features are extracted using the cross-correlation method to estimate the similarity between two windows. These two features are defined as the maximum normalized cross correlation between the current window and the subsequent window, and between the current window and the window after the next, respectively. They are defined as:

$$\eta_1 = \max(|xcorr(a(k), a_1(k))|) \tag{5.15}$$

$$\eta_2 = \max(|xcorr(a(k), a_2(k))|) \tag{5.16}$$

where $a_1(k)$ and $a_2(k)$ are the subsequent window and the window after the subsequent window, respectively, and the $xcorr$ is the cross-correlation operator, whose output is the cross-correlation sequence between the two windows. The output sequence is normalized so that the autocorrelations at zero lag equal 1. Therefore, the values of η_{X1} and η_{X2} lie between 0 and 1, larger values represent larger similarity. Two inputs signals are padded with zeros at the end of the same length before computing the correlation features.

Frequency domain features

Four frequency domain features are considered in this chapter for an analysis based on the previous state-of-the-art contributions [38].

The peak frequency, denoting a frequency at which the maximum spectral power occurred, is evaluated using the following equation:

$$f_p = \arg \max_{f \in [0, f_{max}]} |F_X(f)|^2 \tag{5.17}$$

where $F_X(f)$ is the Fourier transform of the signal a_X, and $f_{max} = 125\ Hz$.

The spectral centroid, featuring the "Center of Mass" of the spectrum, is calculated as:

$$\hat{f} = \frac{\int_0^{f_{max}} f|F_X(f)|^2 df}{\int_0^{f_{max}} |F_X(f)| df} \tag{5.18}$$

In the current study, the occupied bandwidth (OBW) of the signal, OBW, is defined as the occupied bandwidth that contains 99% of the integrated power of the signal. It is usually used in signal processing application to measure the spread of the signal energy in frequency domain.

The frequency domain average power of the signal is measured by dividing the sum of power spectral density (PSD) with the length of the PSD, using the following equation:

$$P = \frac{\int_0^{f_{max}} |F_X(f)|^2 df}{f_{max}} \tag{5.19}$$

Besides the above-mentioned time and frequency domain features, the time-frequency domain features are also considered, according to previous research contributions. The discrete wavelet transform (DWT) has been widely used to learn about the relative energies in different time-frequency bands [38]. We used, in this thesis, a 4-level discrete wavelet decomposition to compute the following features, based on the complexity of the signal.

The energy described by the approximation coefficients is computed as follows:

$$E_{a_4} = \|a_4\|^2 \tag{5.20}$$

where $\|\cdot\|$ is the Euclidean norm operator and a_4 is a vector containing wavelet approximation coefficients at the 4th level. The energy described by the k^{th} level detail coefficients can be calculated by

$$E_{d_k} = \|d_k\|^2 \tag{5.21}$$

The relative energy contribution from each decomposition level is computed as the time-frequency domain features using:

$$\Phi_a = \frac{E_{a_4}}{E_{a_4} + \sum_{k=1}^{4} E_{d_k}} \times 100\% \tag{5.22}$$

$$\Phi_{d_k} = \frac{E_{d_k}}{E_{a_4} + \sum_{k=1}^{4} E_{d_k}} \times 100\% \tag{5.23}$$

The wavelet entropy is computed as:

$$\Theta = -\Phi_{a4} log_2 \Phi_{a4} - \sum_{k=1}^{4} \Phi_{d_k} log_2 \Phi_{d_k} \quad (5.24)$$

The wavelet energy features utilized above are considered to be measures of the degree of the time-frequency based order-disorder of the signal [38]. For instance, a periodic mono-frequency signal is considered to be very ordered; thus, its wavelet representation is usually on one unique wavelet resolution level, and Θ will be have a very small value. However, if the subject signal has a very disordered behavior, its wavelet energy distribution will have a significant equivalence with all frequency bands, and Θ for this signal will reach high values. A summary of all time and frequency domain features can be found in Table 5.1.

Features	Definition
$Max_{ML}, Max_{AP}, Max_V, Max_M$	Maximum value of the window.
$Min_{ML}, Min_{AP}, Min_V, Min_M$	Minimum value of the window.
$\mu_{ML}, \mu_{AP}, \mu_V, \mu_M$	Mean value of the window.
$\sigma_{ML}, \sigma_{AP}, \sigma_V, \sigma_M$	Standard deviation of the window.
$\eta_{1ML}, \eta_{1AP}, \eta_{1V}, \eta_{1M}$	Normalized maximum cross correlation with subsequent window.
$\eta_{2ML}, \eta_{2AP}, \eta_{2V}, \eta_{2M}$	Normalized maximum cross correlation with after next window.
$f_{pML}, f_{pAP}, f_{pV}, f_{pM}$	Peak frequency of the window.
$\hat{f}_{ML}, \hat{f}_{AP}, \hat{f}_V, \hat{f}_M$	Spectral centroid of the window
$OBW_{ML}, OBW_{AP}, OBW_V, OBW_M$	Occupied bandwidth.
P_{ML}, P_{AP}, P_V, P_M	Average frequency domain power.
$\Phi_{aML}, \Phi_{aAP}, \Phi_{aV}, \Phi_{aM}$	Relative energy of the wavelet approximation coefficients at the 4th level.
$\Phi_{d_4ML}, \Phi_{d_4AP}, \Phi_{d_4V}, \Phi_{d_4M}$	Relative energy of the wavelet detail coefficients at the 4th level.
$\Phi_{d_3ML}, \Phi_{d_3AP}, \Phi_{d_3V}, \Phi_{d_3M}$	Relative energy of the wavelet detail coefficients at the 3rd level.
$\Phi_{d_2ML}, \Phi_{d_2AP}, \Phi_{d_2V}, \Phi_{d_2M}$	Relative energy of the wavelet detail coefficients at the 2nd level.
$\Phi_{d_1ML}, \Phi_{d_1AP}, \Phi_{d_1V}, \Phi_{d_1M}$	Relative energy of the wavelet detail coefficients at the 1st level.
$\Theta_{ML}, \Theta_{AP}, \Theta_{ML}, \Theta_M$	Wavelet entropy.

Table 5.1 Time and frequency domain features.

5.2.4 Contour Features Extraction

In addition to investigating the acceleration signals separately for each dimension, the displacement of the CoM are commonly investigated using an ellipse-like plot, which represent the excursion of the acceleration motions in two directions with one versus another [102] [103] [104]. As for walking balance, the most significant movement occurs in the ML and AP directions. Therefore, a regression analysis is conducted on the excursion plots to fit the trajectory with an ellipse-shaped contour using the least square regression methods.

An example of the excursion trajectory of the acceleration plotted with ML vs. AP directions can be found in Fig. 5.6 a). As the figure shows, the trajectory is an irregular curve bound in a certain area, and we aim to find an optimal fit to this curve using an ellipse-shaped contour. The fitting method utilized is the approach proposed by Ohad Gal in 2003 [105].

The mathematical representation of an ellipse is:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{5.25}$$

For a single measure, (x, y), the estimator is defined as:

$$g\big((x,y),A\big) := ax^2 + bxy + cy^2 + dx + ey = -f \tag{5.26}$$

where A is the vector of the parameters to be estimated, i.e., (a, b, c, d, e).

The cost function can be defined using the least square as:

$$\begin{aligned} cost(A) &= \big(g_c\big((x_c,y_c),A\big) - f_c\big)^T\big(g_c\big((x_c,y_c),A\big) - f_c\big) = (XA + f_c)^T(XA + f_c) \\ &= A^T X^T XA + 2{f_c}^T XA + Nf^2 \end{aligned} \tag{5.27}$$

where $g_c\big((x_c, y_c), A\big)$ is the vector function of all the measurements, and each element of $g_c\big((x_c, y_c), A\big)$ is $g\big((x, y), A\big)$. X is a matrix of the form $[x_c^2 \;\; x_c y_c \;\; y_c^2 \; x_c \; y_c]$, and f_c is a $N \times 1$ matrix defined as $[f \; f \; f \; \ldots f]^T$.

The fitting problem can be interpreted as a linear regression task, of which the optimal solution can be found when the derivation of the least square cost function is 0. Hence the solution for the optimal A is

$$A = (X^T X)^{-1} X^T (-f_c) \tag{5.28}$$

After solving the linear regression problem, the next step is to extract the parameters from the conic equation. As the ellipse usually has a tilt (orientation), when $b \neq 0$. The tilt is removed so that the ellipse to remain with a conic representation without a tilt, this is done by using the following substitution: replace x with $\cos(\alpha)\, x + \sin(\alpha) y$ and y with $-\sin(\alpha)x + \cos(\alpha)y$ such that the conic presentation is:

$$\begin{aligned} &\mathrm{a}(\cos(\alpha)\, x + \sin(\alpha)y)^2 + b(\cos(\alpha)\, x + \sin(\alpha)y)(-\sin(\alpha)x + \cos(\alpha)y\,) + \\ &c(-\sin(\alpha)x + \cos(\alpha)y)^2 + d(\cos(\alpha)\, x + \sin(\alpha)y) + \mathrm{e}(-\sin(\alpha)x + \cos(\alpha)y) + \mathrm{f} = 0 \end{aligned} \tag{5.29}$$

After some derivation, finally the long-axis, the short axis and orientation of the ellipse are defined as:

$$r_l = 2 \times \max(a, b) \tag{5.30}$$

$$r_s = 2 \times \min(a, b) \tag{5.31}$$

$$\varphi = \alpha \tag{5.32}$$

Four contour features are considered for classification computed from the ellipse-shaped contour: the length of long axis, r_l, the length of short axis, r_s, the orientation, φ, and the distance of the center of ellipse (X_0, Y_0) from the origin $(0,0)$, which is computed as:

$$o = \sqrt{X_0^2 + Y_0^2} \tag{5.33}$$

Detailed derivations can be found in Ohad Gal's documentations. An example of an acceleration excursion plot, and the ellipse fitted on the excursion plot using the least square based regression method are illustrated in Fig. 5.6. The long and short axes are shown in green, and the origin point of the coordinate is shown in red.

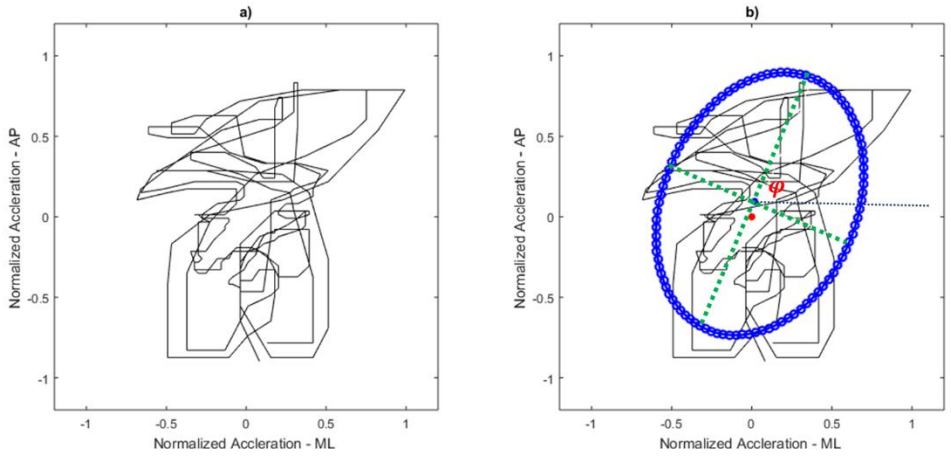


Figure 5.6. Fitting acceleration excursion in ML and AP directions with an ellipse using least square based regression.

5.2.5 Feature Analysis and Classification

The extracted 64 time and frequency domain features, along with the four contour features, are analyzed and selected using the same procedure introduced in Chapter 4. To be specific, the significance of the features is investigated using a two-sampled t-test between the groups, and the features are ranked using the p-value obtained from the t-test. The classification study is conducted according to the steps in Chapter 4. The SVM and ANN classifiers are deployed, and the classification is done at both the sample level and subject level. A holdout validation is carried out to evaluate the classifiers on sample-level classification performances, and a LOSO validation approach is used for the subject-level classification. In this chapter, the samples are defined as the windows generated from the sliding window approach, and the label and score of a subject is computed based on the sample-level classification results, with its percentage of samples

classified as its major class. The t-test- and PCA-based feature selections are employed to optimize the classifiers. The final results are reported with accuracy, TPR, TNR, PPV, NPV, and the scores for the groups.

5.3 Experimental Results

5.3.1 Experiment

The trunk acceleration signals were collected from 54 subjects, consisting of 27 healthy subjects (age: 29 ± 9, 14 men and 13 women), and 27 subjects with four different types of NDDs. The PT group consisted of four patients with muscular sclerosis (MS), (age: 44 ± 6, two men and two women), nine patients with Parkinson's disease (PD) (age: 76 ± 4, six men and three women), three patients with stroke (age: 68 ± 4, two men and one women) and 11 patients with polyneuropathy (PNP) (age: 79 ± 6, six men and five women). All the patients had been mildly affected by the diseases and had provided informed consent as approved by Neuropsychiatricum, Bremen, Germany.

All the subjects were requested to freely walk four times along a 70-m straight hallway on level ground without stopping at their preferred speeds. The AHRSs was attached on the back of the waist according to the setup described in Chapter 3. No abnormal behavior caused by tiredness or discomfort was observed from any subjects during the entire experiment. The number of steps performed by each subject ranged from 30 to 50 depending on the walking speed. The experimental scenario of collecting trunk acceleration signals from one PD patient is shown in Fig. 5.7. In order to reduce the movement of the sensors, which is irrelevant to trunk movement, a belt was fastened around the waists of all subjects to stabilize the sensors.

Figure 5.7. Experiments for collecting trunk acceleration signals.

5.3.2 Feature Analysis

The statistical analysis results are listed in Tables 5.2, 5.3, 5.4, and 5.5 for all the features in all the dimensions, including the means and SDs for the HC and PT groups, the p-value between the two groups, the ranking of features ranging from 1 to 68, and the significance of the features. The significance is indicated as follows: * indicates a significant feature, with $p < 0.01$, while ** indicates very significant features with $p < 0.001$, and the rest of the features are non-significant features.

Features for trunk acceleration -ML						
Feature	**HC**		**PT**		**p-value**	**rank**
	Mean	**SD**	**Mean**	**SD**		
Max_{ML}	0.7461	0.1554	0.7067	0.1330	4.76e-9	49**
Min_{ML}	0.6899	0.1465	0.6768	0.1750	0.09	62
μ_{ML}	0.0089	0.0085	0.0111	0.0103	6.96e-7	51**
σ_{ML}	0.2815	0.0493	0.3032	0.0632	3.33e-15	37**
η_{1ML}	0.7634	0.0519	0.7649	0.0581	0.5798	65
η_{2ML}	0.7778	0.0976	0.7632	0.1073	2.8e-3	56*
f_{pML}	3.9975	2.8083	2.1231	2.1819	8.8e-56	14**
$\hat{f}_{ML}$	5.9480	1.2462	4.4840	1.7289	4.1e-81	2**
OBW_{ML}	32.8778	8.0734	25.8077	9.5500	1.1e-58	10**
P_{ML}	22.7215	7.8783	28.1962	12.7562	3.5e-25	29**
Φ_{aML}	0.4748	0.1772	0.6160	0.2022	1.2e-51	13**
Φ_{d_4ML}	0.3819	0.1317	0.2898	0.1571	1.6e-38	19**
Φ_{d_3ML}	0.1342	0.0802	0.0882	0.0662	3.8e-40	23**
Φ_{d_2ML}	0.0076	0.0056	0.0050	0.0047	8.5e-27	30**
Φ_{d_1ML}	1.4784e-3	8.1419e-4	9.5689e-4	8.3390e-4	2.4e-39	21**
Θ_{ML}	1.3774	0.1922	1.1547	0.3003	2.6e-67	4**

Table 5.2. Statistics on features for ML dimension.

Features for trunk acceleration -AP						
Feature	**HC**		**PT**		**p-value**	**rank**
	Mean	**SD**	**Mean**	**SD**		
Max_{AP}	0.7621	0.1421	0.8038	0.1148	3.1e-12	46**
Min_{AP}	0.7886	0.1590	0.6567	0.1587	1.0e-64	8**
μ_{AP}	0.0143	0.0125	0.0136	0.0132	0.2418	64
σ_{AP}	0.3447	0.0456	0.3291	0.0526	3.8e-11	45**
η_{1AP}	0.8508	0.0527	0.8003	0.0596	7.5e-73	5**
η_{2AP}	0.8860	0.0557	0.8350	0.0909	4.6e-41	15**
f_{pAP}	1.7713	0.1280	1.7222	0.2023	3.9e-9	47**
$\hat{f}_{AP}$	3.9523	0.5895	3.7609	1.1415	2.3e-5	53**
OBW_{AP}	28.2889	6.1621	28.2573	7.5481	0.9238	67
P_{AP}	33.9577	9.4451	32.5528	11.0204	0.0042	58*
Φ_{aAP}	0.7624	0.0951	0.7809	0.1341	0.0011	55*
Φ_{d_4AP}	0.1773	0.0862	0.1544	0.0958	1.3e-7	50**
Φ_{d_3AP}	0.0525	0.0282	0.0573	0.0610	0.0418	61
Φ_{d_2AP}	0.0068	0.0033	0.0063	0.0039	0.0099	59*
Φ_{d_1AP}	0.0010	4.3922e-4	0.0011	6.3408e-4	9.7e-5	54**
Θ_{AP}	0.9745	0.2094	0.8930	0.3460	7.8e-9	48**

Table 5.3. Statistics on features for AP dimension.

Features for trunk acceleration -V						
Feature	**HC**		**PT**		**p-value**	**rank**
	Mean	**SD**	**Mean**	**SD**		
Max_V	0.8752	0.1017	0.8412	0.1063	6.1e-12	44**
Min_V	0.4967	0.1461	0.4484	0.1002	3.3e-17	40**
μ_V	0.0037	0.0036	0.0041	0.0034	0.0083	60*
σ_V	0.3346	0.0713	0.2932	0.0605	7.3e-40	22**
η_{1V}	0.8664	0.0535	0.8131	0.0670	4.5e-69	6**
η_{2V}	0.9085	0.0607	0.8302	0.1200	1.3e-57	7**
f_{pV}	1.8137	0.4113	1.9415	0.6842	4.4e-6	52**
$\hat{f}_V$	3.8658	0.5096	4.1065	0.8294	1.5e-12	41**
OBW_V	18.4939	3.8671	20.1628	4.6095	4.3e-16	36**
P_V	32.1021	12.8314	25.3682	9.0187	4.4e-39	24**
Φ_{aV}	0.7329	0.0970	0.6538	0.1573	1.1e-33	20**
Φ_{d_4V}	0.1998	0.0925	0.2758	0.1450	7.8e-36	18**
Φ_{d_3V}	0.0640	0.0287	0.0661	0.0353	0.1748	63
Φ_{d_2V}	0.0029	0.0017	0.0038	0.0023	4.1e-20	32**
Φ_{d_1V}	4.6811e-4	2.2125e-4	5.8845e-4	3.0256e-4	1.9e-20	31**
Θ_V	1.0260	0.1923	1.1093	0.2639	1.4e-13	39**

Table 5.4. Statistics on features for V dimension.

Features for trunk acceleration -M						
Feature	**HC**		**PT**		**p-value**	**rank**
	Mean	**SD**	**Mean**	**SD**		
Max_M	12.0034	6.6545	9.4756	8.6817	1.6e-11	43**
Min_M	5.3550	3.1140	3.5979	3.4222	8.9e-29	28**
μ_M	7.5864	4.1835	5.5504	4.8538	1.3e-20	33**
σ_M	1.5630	0.9257	1.3968	1.4915	0.0064	57*
η_{1M}	0.9755	0.0217	0.9377	0.0509	1.0e-75	1**
η_{2M}	0.9802	0.0173	0.9497	0.0456	4.7e-64	3**
f_{pM}	0	0	0	0	-	68
$\hat{f}_M$	0.3190	0.3105	0.5821	0.4914	2.1e-37	17**
OBW_M	9.4417	5.1815	12.4368	5.4584	1.4e-31	26**
P_M	4.2498e+4	2.4631e+4	3.1243e+4	2.9755e+4	1.6e-17	35**
Φ_{aM}	0.9817	0.0181	0.9623	0.0328	4.4e-47	11**
Φ_{d_4M}	0.0116	0.0109	0.0250	0.0221	2.7e-50	9**
Φ_{d_3M}	0.0060	0.0072	0.0112	0.0112	1.0e-28	25**
Φ_{d_2M}	5.9545e-4	9.2259e-4	1.3590e-3	1.8543e-3	5.6e-25	27**
Φ_{d_1M}	9.4570e-5	1.5143e-4	1.6129e-4	2.1183e-4	1.2e-13	38**
Θ_M	0.1401	0.1165	0.2507	0.1812	3.1e-47	12**

Table 5.5. Statistics on features for M dimension.

Features for trunk acceleration –Ellipse features						
Feature	**HC**		**PT**		**p-value**	**rank**
	Mean	**SD**	**Mean**	**SD**		
r_l	1.4838	0.1878	1.3656	0.1481	1.7e-49	16**
r_s	1.0583	0.1424	1.0073	0.1481	1.6e-13	42**
o	0.1347	0.0587	0.1067	0.0674	4.3e-20	34**
φ	0.3933	0.1982	0.3897	0.2084	0.7037	66

Table 5.6. Statistics on contour features.

The features are mainly analyzed from two perspectives: the differences between groups, and the difference between dimensions.

- For the maximum, Max_X, there is very significant difference between groups for all the four dimensions. To be specific, the Max_X value is larger for the HC group than for the PT group in the ML, V, and M dimensions, and smaller in the AP dimension. The SD values are larger for the HC group in the ML and AP dimensions, and smaller in the V and M dimensions. When comparing the differences between the dimensions, the values for the V direction are larger than those for the ML and AP directions.
- For the absolute minimum values, Min_X ,very significant differences were observed in all the AP, V, and M dimensions, and no significant differences were observed in the ML dimension. For all the four dimensions, the mean values for the HC group are larger than those of the PT group, and the means in the V dimension are smaller than those in the ML and AP dimensions.
- For the mean value, μ_X, the values for the PT group are significantly larger and very significantly larger in the V and ML dimensions, respectively, while significantly smaller in the M dimension. No significant differences were observed in the AP direction. When making comparisons between dimensions, the values in the V direction were smaller than those in the ML and AP directions.
- The σ_X for the PT group was very significantly larger than that of the HC group in the ML direction, and significantly and very significantly smaller in the M, and AP and V dimensions, respectively.
- For the cross correlation based features, η_{1X} and η_{2X}, but not η_{1ML}, for all the rest of the features in all dimensions, the values for the PT group are significantly or very significantly smaller than those of the HC group, showing that the windows have smaller similarity and hence a larger variability for the PT group. For the differences between dimensions, the similarities between the windows are highest for the M dimension, followed by the V dimension, and then the AP and ML directions.
- For peak frequency features, the values are very significantly larger for the HC group in the ML and AP direction, while very significantly smaller in the V direction. The peak frequency for the M dimension was 0 for both groups.
- The spectral centroids for the HC group are very significantly larger for the ML and AP dimensions, and very significantly smaller for the V and M dimensions. The values for the M dimension are generally much smaller than those of the other three dimensions.
- A very significant difference can be observed in the occupied bandwidth in the ML, V, and M dimensions. The values are smaller for the PT group in the ML dimension and larger in the V and M dimensions.
- For the average power, all the features are significant or very significant. The HC group has in principle a larger average energy in the AP, V, and M dimensions and a smaller energy in the ML dimensions.
- Regarding the time-frequency domain features in the ML direction, all five features are very significant. For the HC group, the approximation coefficient

takes in average 47.48% of energy, while for PT group, it makes up 61.60% of energy. Four out of five features are considered as significant or very significant features for the AP and V directions, and all the five features are regarded as very significant features for the M dimension.

- For the wavelet entropy, all the four features are very significant features. The PT group has smaller entropies in the ML and AP dimensions, and larger entropies in the V and M dimensions. Moreover, it is noticeable that for both groups, the entropies for the ML direction are larger than those of the AP and V dimensions, and much larger than those of the M dimension.
- For the four contour features, three features are significant. On average, the short axis and the long axis are longer for the HC group than for the PT group, and the distance of the ellipse center from the origin is larger for the HC group.

To visualize the differences in the sense of the novel contour features, the ellipses formulated using the mean values listed in Table 5.6 for the two groups are plotted in Fig. 5.8. It can be seen that the tilt between the two groups has no significant difference, while there is an observable difference in the length of the long axis.

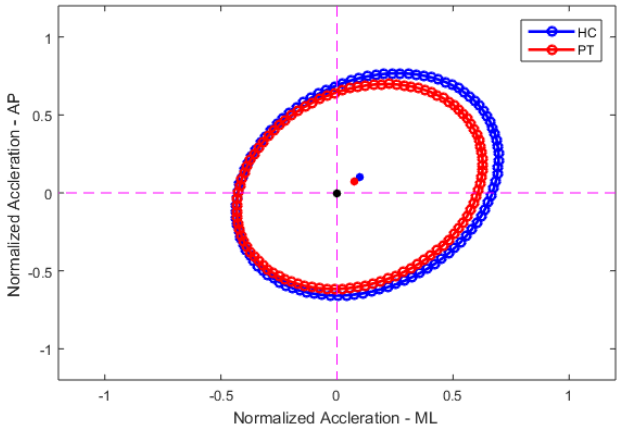


Figure 5.8. Ellipse for HC and PT groups.

Consistent with previous findings [38] [106] [107], it can be concluded from the features extracted that, most of the frequency domain and the time-frequency domain features are able to distinguish the two groups. These measures provide unique but complementary information. For instance, the HC group exhibited greater disorder and randomness in the ML and AP directions than the PT group, as the control of ML motion needs continuous feedback control, allowing online step-to-step adjustments for effective balance control [106]. For the PT group, less ML disorder and complexity, and loss of smoothness in forward progression was found. Moreover, when examining the randomness of the entropy, it can be seen from both the features and the ellipse plots that the movement and randomness in the ML direction was larger than in the other dimensions, which is consistent with previous research [38]. Besides findings that correlated with previous work, new findings in our results can also provide new insights. For instance, the disorder and randomness for the PT group were larger in V direction than for the HC group, which

may be associated with diseases related reasons, such as foot drop, shaking, and shuffling; from the features in the M dimension, we can see that the HC group had on average a larger acceleration magnitude and a smaller variability between the steps, as shown by the η_{1M} and η_{2M}. This finding is consistent with the findings in Chapter 4, namely, HC subjects are able to perform larger trunk movements while controlling their movements with good balance. As the medical related findings are not the goal of this thesis, whose one goal is to extract novel features for classification purposes, the underlying clinical explanations of the features will not be discussed in more depth.

5.3.3 Classification Results

5.3.3.1 Results on Sample-level and Subject-level Classification

The classification results performed on the trunk acceleration signals collected from the 54 subjects are summarized in Table 5.7, including the results for both sample-level and subject-level classifications. The positive class is HC. The numbers in the bracket are the values used for the calculation of parameters.

		Time and frequency domain features		All features	
		SVM	**ANN**	**SVM**	**ANN**
Sample-level Classification	**ACC**	99.26% (5618/5660)	96.57% (5466/5660)	99.86% (5652/5660)	97.86% (5539/5660)
	TPR (Sen)	98.94% (2230/2554)	96.09% (2165/2253)	99.84% (2552/2556)	96.55% (2181/2259)
	TNR (Spe)	99.47% (3388/3406)	96.89% (3301/3407)	99.89% (3400/3404)	98.74% (3358/3401)
	PPV (Pre)	99.20% (2230/2248)	95.33% (2165/2271)	99.84% (2552/2556)	98.07% (2181/2224)
	NPV	99.30% (3388/3412)	97.40% (3301/3389)	99.89% (3400/3404)	97.73% (3358/3436)
	AUC	0.9997	0.9876	0.9999	0.9903
Subject-level Classification	**ACC**	94.44% (51/54)	94.44% (51/54)	96.30% (52/54)	94.44% (51/54)
	TPR (Sen)	92.59% (25/27)	92.59% (25/27)	96.30% (26/27)	92.59% (25/27)
	TNR (Spe)	96.30% (26/27)	96.30% (26/27)	96.30% (26/27)	96.30% (26/27)
	PPV (Pre)	96.15% (25/26)	96.15% (25/26)	96.30% (26/27)	96.15% (25/26)
	NPV	92.86% (26/28)	92.86% (26/28)	96.30% (26/27)	92.86% (26/28)
	HC score	90.09%	78.52%	93.97%	89.06%
	PT score	90.45%	84.42%	96.67%	90.58%
	Overall score	89.73%	81.41%	95.16%	89.84%

Table 5.7 Classification results on sample and subjects

It can be seen from the sample-level classification results that, the overall accuracy achieved using all features are 99.86% and 97.86% for SVM and ANN classifiers,

respectively, which are higher than that achieved by just using the time and frequency domain features. When comparing the TPR and TNR, the results showed that a higher percentage of PT samples than HC samples was correctly predicted, based on the knowledge that TPR is the percentage of HC sample correctly classified and TNR is that for PT samples. For the SVM classifier, the NPV is higher than the PPV for both feature sets, while for the ANN classifier, the NPV is higher than the PPV for only time and frequency domain features, and lower than PPV for using all features. From the AUC values, it can be observed that the performances of the classifiers are very promising, with the highest AUC achieved with the SVM classifier and all features. In general, the SVM outperforms the ANN classifiers for both feature sets. Three kernels for the SVM classifier were tested, including the linear kernel, RBF kernel, and polynomial kernel. The best result was obtained using the RBF kernel and listed in the table, with no significant differences between the linear and polynomial kernels. The best results for the ANN were achieved with 20 neurons. The type of active functions did not have a considerable impact in our study, and the sigmoid function was chosen as the default function. In summary, the utilization of machine learning features is contributive to the classification results; the SVM outperforms the ANN on sample-level classification; better results were achieved for the PT group.

For the subject-level classification, the best accuracy, 96.30%, was achieved with all features using SVM classifier, with only two subjects out of the 54 misclassified. Higher accuracy, TPR, PPV, and NPV were achieved by using all features, rather than by using only the time and frequency domain features, for SVM classifier, while the results for those parameters are the same for the ANN classifier for both features sets. A significant increase in all the three scores for both classifiers can be observed when all the features are used. When comparing the two classifiers, the SVM classifier provided better results for both features sets for all the parameters. Regarding the two groups, the average scores of the PT class were higher than those of the HC class, indicating that the possibility of misclassifying the PT samples as HC samples was low. Consistent with the findings in Chapter 4, this has value in clinical scenarios, as it is usually more dangerous if a subject with NDD is misdiagnosed as healthy, which may cause a delay in treatment.

5.3.3.2 Results with Features Selection

The p-values obtained with the two sampled t-test were utilized in this chapter as the feature selection criteria. Using the same procedures as introduced in Chapter 4, the features were ranked according to their significance level, and the features sets composed of top n features, $n = 1, 2, 3, \ldots, 68$, were tested for subject-level classification using the LOSO validation approach. The accuracy and overall score are compared for SVM and ANN classifiers. The results are depicted in Fig. 5.9.

The results showed that, for the SVM classifier, the best accuracy could be reached with a minimum of 37 top features, which is higher than that when all features are used (96.30%) as listed in Table 5.7. The best score was achieved with the top 56 features, and was slightly higher than that achieved when all features are used. For the ANN classifier, the highest accuracy was obtained with the top 40 features, and was higher than that

obtained when all features were used. The highest score was obtained with the top 65 features. By comparing the two classifiers, it can be concluded that the overall performance of the SVM is better than that of the ANN. When the same number of features are used, the SVM classifier tends to yield better results.

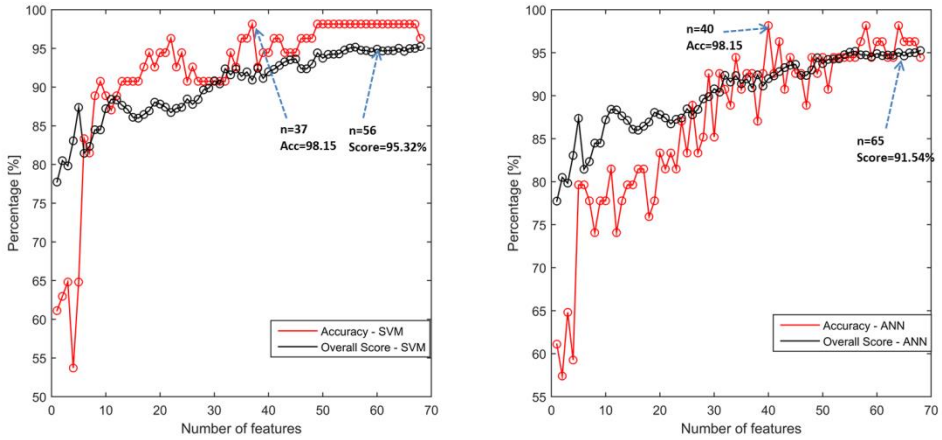


Figure 5.9. Classification results with t-test based features selection.

5.3.3.3 Classification of Pathological Classes

As the PT class is composed of subjects with four types of pathologies, it is worthwhile to investigate if the proposed framework is also capable of distinguishing between the PT classes as well. The classification study was carried out to distinguish the PD and PNP subjects using the SVM classifier, and all 68 features and the results are listed in Table 5.8.

The results revealed that the extracted features were also able to distinguish the PD and PNP classes, with only one subject from the PNP group misclassified as being a PD subject. The scores showed promising classification performance on the sample level, with 86.71% of the samples on average correctly predicted as belonging to the correct group.

Even though patients are rarely misdiagnosed as having other diseases, the main purpose of conducting a classification between two diseases is to provide new insights and interesting features that may reveal the differences in gait pattern brought on by similar underlying causes, and therefore to help clinicians to understand more comprehensively the correlations between the symptoms of an illness and the walking performance of a subject with that illness, and further manage the patients and optimize the rehabilitation treatment strategy. Practically, it is worthwhile to explore how the interventions impact the features, such as entropy and peak frequency.

Accuracy	95% (19/20)
TPR	88.89% (8/9)
TNR	100% (11/11)
PD score	81.25%
PNP score	91.89%
Overall Score	86.71%

Table 5.8 Results for classification of PD and PNP classes.

5.4 Applications to Patient Management and Rehabilitation

As has been stated, one of the most important reasons for developing gait classification algorithms is to provide support for patient management and rehabilitation processes in clinical scenarios. Therefore, a case study was carried out on four patients diagnosed with PD to monitor their gait performances under medication for a short duration. Those patients participated in the monitoring experiment for the joint angle based study introduced in Section 4.4.2 as well, and the trunk acceleration signals were recorded along with the joint angle signals. The classification results, including the label, score, and values for some top features are listed in Tables 5.9 to 5.12. The self-evaluation of walking performance was considered as the standard for comparison, with larger number representing better self-evaluation.

Subject: PD1				
Age: 80 **History:** 12 years **Gender**: Male				
	Before	**15min after**	**30min after**	**60min after**
Label	PT	PT	PT	PT
Score (%)	95.45	79.49	87.23	95.00
Self-evaluation	1	3	2	1
$\hat{f}_{ML}$	1.6747	2.2823	2.3796	2.0375
η_{1M}	0.9105	0.8904	0.8954	0.8926
Θ_M	0.3217	0.3129	0.2928	0.3363
r_l	1.7267	1.4615	1.3894	1.4216

Table 5.9. Classification results on PD patients 1 for monitoring medication effect.

Subject: PD2			
Age: 70 **History:** 1 years **Gender**: Male			
	Before	**20min after**	**50min after**
Label	PT	PT	PT
Score (%)	82.35	70.59	63.64
Self-evaluation	1	3	4
$\hat{f}_{ML}$	6.4851	7.6713	7.1863
η_{1M}	0.9122	0.8953	0.9070
Θ_M	0.5872	0.5810	0.5312
r_l	1.2184	1.5409	1.5364

Table 5.10. Classification results on PD patients 2 for monitoring medication effect.

Subject: PD3			
Age: 72 **History:** 2 years **Gender**: Male			
	Before	**10min after**	**30min after**
Label	PT	PT	PT
Score (%)	83.33	95.65	69.23
Self-evaluation	1	1	3
$\hat{f}_{ML}$	6.7139	6.7929	6.0283
η_{1M}	0.8833	0.8557	0.8762
Θ_M	0.5963	0.6338	0.6156
r_l	1.4410	1.5946	1.3895

Table 5.11. Classification results on PD patients 3 for monitoring medication effect.

Subject: PD4			
Age: 78 **History:** 10 years **Gender**: Male			
	Before	**40min after**	**80min after**
Label	PT	PT	PT
Score (%)	88.89	81.08	90.63
Self-evaluation	1	3	1
$\hat{f}_{ML}$	5.2761	4.5654	5.3387
η_{1M}	0.8328	0.8428	0.8517
Θ_M	0.6588	0.5946	0.6465
r_l	1.1960	1.3294	1.2779

Table 5.12. Classification results on PD patients 4 for monitoring medication effect.

Based on the self-evaluation of subject PD1, the subject felt the medication taking effect after 15 and 30 min, and the effects fading away after 60 min; the same trend was reflected in the scores, with lower scores for 15 and 30 min, and a higher score for 60 min. For the features listed in Table 5.9, the entropy, Θ_M, showed a similar trend as the score, namely, a decrease during the first two phases, and a decrease during the last phase. As concluded in Section 5.3.2, a healthy gait pattern intends to have a smaller value of Θ_M, representing a less random and disordered behavior. However, other features are not able to reveal the changes in gait performances by themselves. The subject PD2 felt a significant improvement after the first 20 min and slight improvement after 50 min. The same trend was observed in the scores and the entropy Θ_M. Subject PD3 did not feel any improvement after the first 10 min, but felt a large improvement after 30 min, and similar changes could be seen with the classification results as well. Subject PD4 felt a gait improvement 40 min after taking the medication, but did not feel an improvement after 80 min. The same conclusion can be drawn from the classification results. It can be noticed that, except of the wavelet entropy, Θ_M, all the other listed features cannot reflect the changes in gait in isolation. In other words, statistically speaking, standalone features can have significant differences between the HC and PT groups; however, a simple increase or decrease in one individual feature cannot be used to evaluate the entire walking performance. More studies are needed to reveal the correlations between each individual feature and the overall gait quality.

As has been proved in many previous research studies, traditional interventions can improve the walking quality of patients by increasing the walking speed or improving the motor control process essential for performing stepping [108] [109]. However, not many studies have focused on the correlations between the extracted parameters, including the time and frequency domain parameters, such as the peak frequency, bandwidth, and the novel machine learning features, and the rehabilitation progress. Therefore, it would be beneficial in future work to further evaluate the features in coordination with the cause, severity, and rehabilitation progress of the specific diseases.

5.5 Conclusions

In this chapter, the proposed general machine learning framework for gait classification is validated with trunk acceleration signals. The proposed solution invokes a sliding window approach for generating samples for classification. The least square based regression method was deployed for extracting additional novel machine learning features, which were shown to contribute to the classification results. The results of validations on 54 subjects have shown the effectiveness of the proposed framework. The extracted features were analyzed in depth, and t-test based feature selection method was used for generating optimized classifiers. The SVM showed a better performance than the ANN for both sample-level and subject-level classifications. At the end of the study, four PD patients were monitored under clinical scenarios with medication, and the outcome indicated the potential use of the framework for supporting patient management and rehabilitation monitoring.

6. Gait Classification for Stride Interval Signals

The machine learning framework proposed in this thesis is validated by solving gait classification problems on stride interval signals in this chapter. The data analyzed in this chapter are the time durations of the stride-related parameters. The chapter is organized in the following way: Section 6.1 describes related work, including illustration of the signals, and the state-of-the-art classification approaches and their limitations; the proposed gait classification approach based on the proposed machine learning framework is discussed in depth in Section 6.2; experimental results are presented in Section 6.3; and the entire chapter is summarized and concluded in Section 6.4.

6.1 Related Work

In Chapters 4 and 5, two of the most important types of gait signals, i.e., the joint angle signals and the trunk acceleration signals have been processed for classification. In addition to those two types of signals, the stride interval signal is also essential in gait analysis, since it could comprehensively reflect the rhythm and dynamic of gait patterns [66]. The stride-to-stride fluctuation can be considered as the walking variability, and analyzing the performance of walking in temporal domain is always a simple and representative solution for understanding the walking behavior.

6.1.1 Stride Interval Signals

The data being concerned in this chapter are the stride interval signals, including five channels, which are the stride time signal, the swing time signal, the double support time signal, the swing percentage signal and the double support percentage signal. Recalling the definition of those parameters introduced in chapter 3 that, the stride time is the duration of the gait cycle, while the stance and swing time are the duration of the phase where the corresponding foot has contact with the ground, and has no contact with the ground, respectively. The double support time is the duration in one cycle that both feet have contact with the ground simultaneously. The stance time and swing time add up to the stride time; therefore, it is enough to consider one of them owing to this linear relationship; in this chapter, only the swing time signal is considered. The swing

percentage and double support percentage are the proportion of swing time and double support time in one gait cycle (stride), calculated by dividing the swing time and double support with the stride time, respectively. The swing phase takes approximately 40% of the whole gait cycle, while the double support period occurs twice in one cycle, and takes around 20% of the normal gait cycle [16].

In this chapter, the term "signal" always refers to the time series that contains a series of discrete values representing the time intervals of gait cycles recorded continuously from walking trials. For instance, a swing time signal is a time series of swing time durations of multiple continuous gait cycles. An example of the stride time, swing time, and double support signals collected from a PD patients walking for 150s are illustrated in Fig.6.1. Each data point represents the duration of its corresponding channel in one gait cycle. As the stance time can be computed by subtracting the swing time from stride time, it is therefore not plotted in the figure. The red crosses are the outliers, which are two standard deviations greater or less than the median value of the whole series.

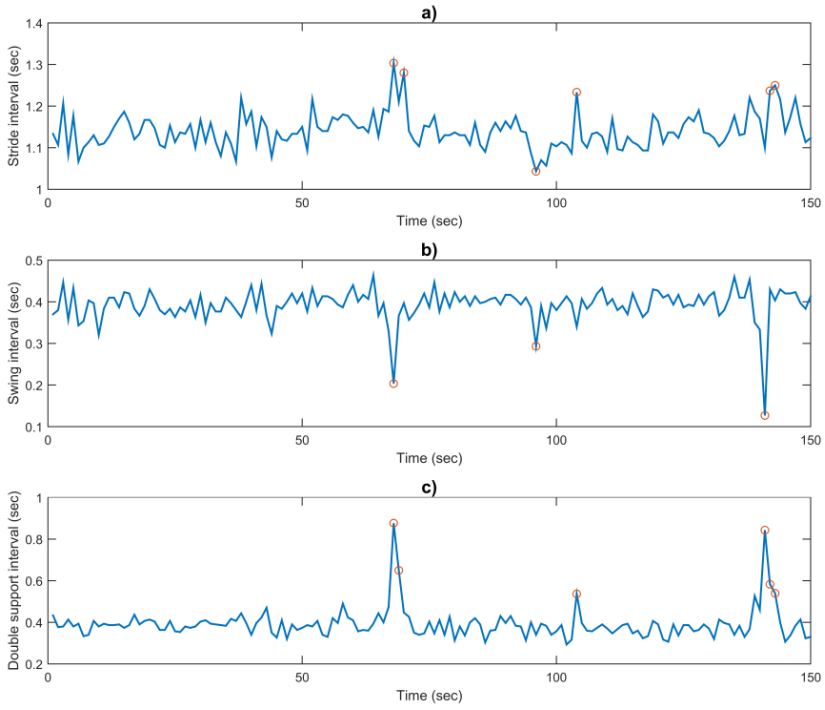


Figure 6.1. Example of raw stride time, swing time and double support time interval signals.

It can be seen that the signals have noticeable fluctuations, indicating the variability between the gait cycles in the temporal domain. The stride interval signals can be considered as time series, and the features can be extracted using statistical approaches by analyzing the fluctuations and used as inputs for analyzing and distinguishing the gait patterns between different groups, especially the groups with gait disorders caused by neurological diseases.

6.1.2 State-of-the-art and Limitations

Stride interval related parameters are the most important temporal parameters for gait analysis, since they carry the dynamic and variability information of walking. The rhythm and stability of walking are commonly investigated by studying the temporal parameters.

The gait rhythm in the patient with PD has been comprehensively studied in [66], where the gait rhythm standard deviation parameters computed from the probability density functions of stride interval related parameters, and the other statistical parameters are extracted and utilized for a SVM-based classification. The best classification accuracy achieved is 90.32% for distinguishing the PD from the HC group. This study has provided very useful information for extracting new features for performing more complex classifications, such as multiclass classification that this chapter focuses on.

A more recent study performed by Peng Ren et al. has proposed phase synchronization and condition entropy as dominate features for distinguishing three pathological groups from the HC group [65]. The promising classification results achieved by this study have proved the importance of the stride-related parameters for gait classification, and brought new insight into the relationship between those parameters and the corresponding gait patterns.

In another study, automatic diagnosis of neurodegenerative disease using the gait dynamics derived from stride related signals was conducted [110]. Statistical features, such as maximum, minimum, mean, and SD were extracted and served to the SVM for distinguishing patients from the healthy. Overall, a 90.63% accuracy was realized, and it was concluded that the double support interval signals are the most effective features for diagnosis.

Stride interval signals are very often used for classification of various gait patterns. For instance, temporal parameters, such as stride time and stance time, were used as features, and classifications of patients with PD and HC were conducted using the multiple regression approach; an accuracy of 92.6% was achieved [60]. The temporal fluctuation in gait dynamics was investigated by Wei Zeng et al., and ANN with RBF kernel was employed for classifying NDD gait from healthy gait [50]. SVM and basic temporal-spatial parameters were used for identifying cerebral palsy gait for diagnosis and evaluation of treatment outcomes in [111], and a promising accuracy of 96.80% was obtained.

Most of the previous studies focused on revealing the statistical characteristics of temporal gait parameters by analyzing the overall signals. In general, classifications were carried out on subject level directly. Owing to the limited number of study participants (e.g., 31 subjects in [66], 64 subjects in [65], 48 subjects in [110], 49 subjects in [60], and 64 subjects in [50]), the advantages of machine learning techniques could not be fully realized. Besides, no study has tried, so far, to solve classification problems on stride interval signals by combining multiple machine learning techniques. Additionally, those studies mainly focused on solving binary classification problems, but the effectiveness of the proposed methods on solving multiclass classification problems were rarely discussed.

In summary, previous research works have provided lots of possible solutions on gait classification problems using temporal parameters. Based on these studies, we would like to move one step further by solving more complex multiclass classification problem by taking advantage of the proposed machine learning framework.

6.2 Gait Classification Using Statistical and Likelihood Features

Figure 6.2 illustrates the work flow of the classification procedures conducted on stride interval signals based on the proposed machine learning framework.

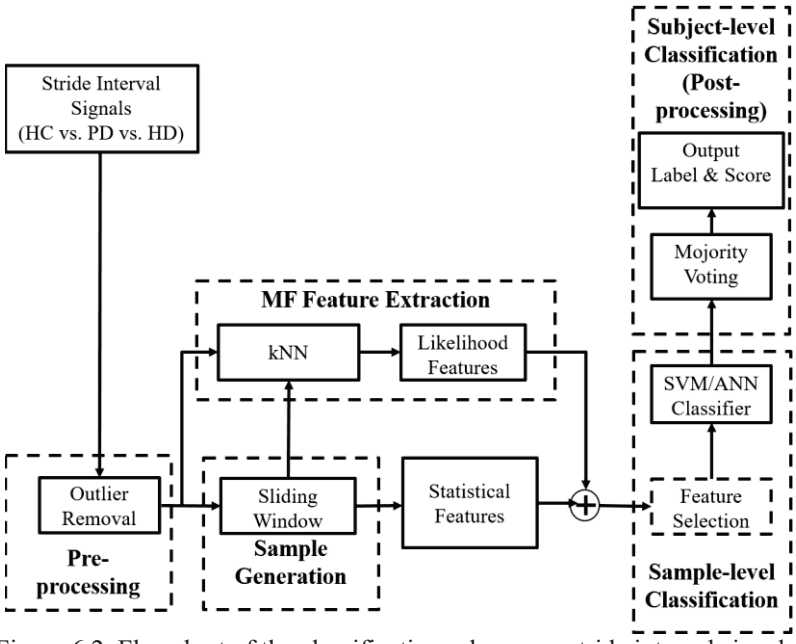


Figure 6.2. Flowchart of the classification scheme on stride interval signals.

The raw data are the stride time signal, swing time signal, double support time signal, the swing percentage signals, and the double support percentage signal. After a simple outlier detection step, the signals are segmented using a sliding window approach that is similar to what has been introduced in Chapter 5. The statistical features are extracted from each window. In parallel, the pre-processed signals are conveyed to the kNN classifier for extracting the machine learning features; the kNN likelihood for each class are combined with statistical features as the full feature sets for high level classification. SVM and ANN are employed as the high-level classifiers for sample level and subject level classifications, respectively. The multiclass classification problem is the focus of this chapter, and the One vs. Rest (OvR) scheme is used to solve such problems with the SVM classifier. The MV approach is implemented in the final stage to make decisions on the final output labels and scores to each test subject.

6.2.1 Data Pre-processing

The pre-processing procedure for each of the five stride interval signals consisted of two simple steps in this chapter: in the first step, the first 5% and last 5% of the signals were removed, as they corresponded to the starting and ending phases of the walking cycle when the subject was adapting their walking speed to reach a stable speed. This is done to minimize the startup effect in the gait; in the second step, the values that are two standard deviations larger or smaller than the median value of the entire signal are replaced with the median value. The outliers are related to the turning strides happening when subjects reach the end of the hallway. According to the "three-sigma rule," [112] around 95% and 99.7% of the normally distributed probability values lie within 2-SD and 3-SD distances from the mean value, respectively. 2-SD is chosen in this chapter, since some of the visible outliers cannot be detected using the 3-SD rule, according to the investigation in [66]. The median value is considered instead of the mean value because, in many cases, the values of the outliers are very large and might greatly affect the mean value of the entire series. An example of the swing interval signals before and after replacing the outliers is illustrated in Fig. 6.3.

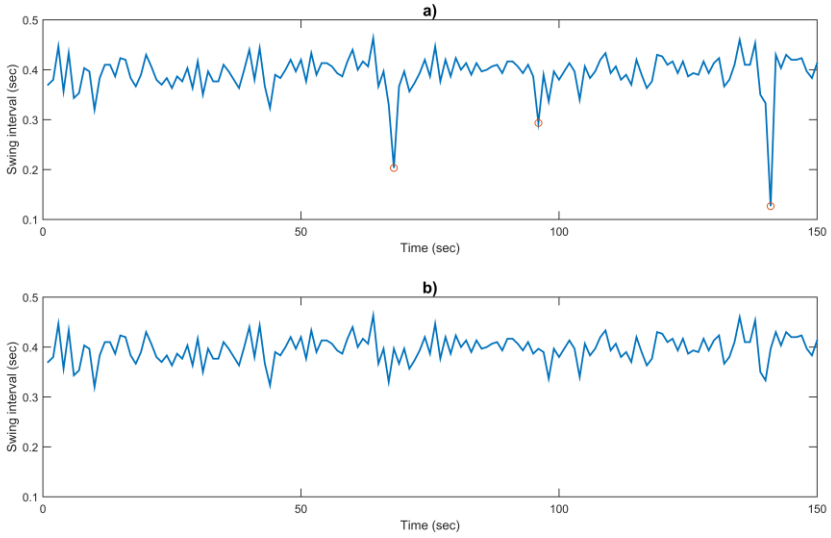


Figure 6.3. Sample swing interval series before (upper plot) and after (lower plot) replacing outliers.

6.2.2 Sliding Window and Sample Generation

Gait cycle segmentation is not necessary for stride interval signals, since the stride interval parameters conceptually correspond to gait cycles. However, in order to have enough information for extracting the statistical features, a sliding window approach is applied to each of the five channels to generate samples for classification. A window is defined in this chapter with a fixed length of 30. For a signal consisting of n data

points, $\{T(i)\}, i = 1,2,3, \dots n\}$, the i^{th} window is defined as the fragment of data starting from data point $T(i)$ and ending at data point $T(i + 29)$. Hence, there is 1/30 of overlap between two successive windows.

The main advantage of using this sliding window method for sample generation is that, it allows for an analysis of stride related parameters in a certain statistical scale instead of point by point, based on the knowledge that the fluctuation in strides of these temporal parameters are very important and can reflect the rhythm of the gait. Besides, analyzing the data within a group makes more sense than analyzing standalone gait cycles for generalizing the overall performances of walking, and offers enough scale of data for understanding walking rhythm in other dimensions.

6.2.3 Statistical Features Extraction

Similar to the method described in Chapter 5, the statistical features are first extracted in the window level. Five statistical features are considered, which are the maximum, the minimum, the mean, SD, and the signal turn count (STC). The first four features, i.e., maximum, minimum, mean, and SD, are self-explanatory, while the STC was first proposed in [113] for analyzing the EMG signal recorded from a patient with a myopathy, whose signal contains more turnings than that of a healthy subject. This feature has been later used in [114] as a dominant feature for searching knee joint disorders. Besides, Wu et al. recently applied the STC approach to evaluate the gait fluctuations in patients with amyotrophic lateral sclerosis (ALS), and it has been proved that the STC, which represents the fluctuation level, is higher in the swing interval of ALS group than in that of the HC group [115]. Based on the outcome of those state-of-the-art studies, the STC is utilized in this chapter as one important feature.

Signal Turns Count

For a time series, $\{T(i)\}, i = 1,2, \dots, n\}$, a data sample is considered as a signal "turn" if two conditions are fulfilled at the same time: 1) the sample indicates an change in direction, namely a change in the sign of the derivative either from positive to negative, or from negative to positive; and 2) the absolute difference between the amplitude of the current sample and the subsequent sample is larger than a threshold. The STC detection criteria can be expressed with the following equations:

$$T(i)\ is\ a\ STC \text{ if} \begin{cases} \big(T(i) - T(i-1)\big)\big(T(i+1) - T(i)\big) < 0 \\ |T(i+1) - T(i)| \geq Th \end{cases},\ 2 \leq i \leq n-1 \qquad (6.1)$$

Practically, the signals refer to the time series after pre-processing. Different from the approach proposed in [66], where the optimal thresholds are decided as fixed values that apply for all subjects, in this thesis, the threshold values are defined as subject-specific parameters, i.e., $Th = 1.5 \times \mu$, where μ is the mean difference value computed from all sample data with their subsequent sample data in the current signal. For instance, the Th for the stride time interval signal for one subject is computed as

$$Th_{stride} = 1.5 \times \frac{1}{n-1}\sum_{i=1}^{n-1}|T(i+1) - T(i)| \quad (6.2)$$

This approach is used instead of the fixed Th value primarily because the inter-subject variability of those temporal parameters can be very significant; hence the threshold that distinguishes the STC of two subjects successfully might not be optimal for other subjects. Therefore, it makes sense to define a subject-specific threshold that represents the characteristics of the corresponding signal individually.

Five statistical features are extracted from each signal. Considering that five signal channels are used in this study, there are 25 features in total extracted from each walking trial of a subject. A summary of all the statistical features is listed in Table 6.1.

Statistical Features	Definition
$Max_{St}, Max_{Sw}, Max_{PerSw}, Max_{Do}, Max_{PerDo}$	Maximum value of the window.
$Min_{St}, Min_{Sw}, Min_{PerSw}, Min_{Do}, Min_{PerDo}$	Minimum value of the window.
$\mu_{St}, \mu_{Sw}, \mu_{PerSw}, \mu_{Do}, \mu_{PerDo}$	Mean value of the window.
$\sigma_{St}, \sigma_{Sw}, \sigma_{PerSw}, \sigma_{Do}, \sigma_{PerDo}$	Standard deviation of the window.
$STC_{St}, STC_{Sw}, STC_{PerSw}, STC_{Do}, STC_{PerDo}$	Signal Turns Count (STC)

Table 6.1. Statistical features.
(St-stride; Sw-swing; PerSw-swing phase percentage; Do-double support time, PerDo-double support percentage)

6.2.4 Likelihood Features Extraction

The kNN has been widely applied to gait classification during the last few years (e.g., [116] [117]) owing to its advantages, such as it is simple to implement, flexible to features, and can handle multiclass problems naturally. As described above, the statistical features are parameters extracted by analyzing the fluctuations in the individual signal channels. Additionally, we would like to analyze the data from another perspective; in other words, we would like to build a model by combining the information from all the five channels, and investigate if a single stride can be classified using the trained model and its values of all five channels. Based on the characteristics and advantages of kNN classifier, the kNN classifier is believed to be suitable as a low-level classifier for extracting necessary machine learning features.

The three-class classification problem is the focus of this chapter, with HC, PD, and HD representing the healthy control class, Parkinson's diseases class, and Huntington's disease class, respectively. According to the theory of the kNN method, the values of all five channels at a certain incidence can be regarded as a sample, and the likelihoods of the sample belong to each of the classes are the final output, representing the probability of the test sample belonging to the distinguished classes. For the data from one walking

trial of one subject, which contains five channels, assume the number of data points, that is, the number of gait cycles, is n; the feature matrix that can then be generated as an n by 5 matrix, where n indicates n samples generated for classification, and 5 means there are five input features for training the kNN classifier.

The kNN classifier is trained using the data from all the classes, while the new samples are predicted using the k nearest neighbors. Assume that the whole data set contains data recorded from m subjects, including m_1 HC subjects, m_2 subject with PD, and m_3 subjects with HD, the number of samples (stride intervals) are labeled with $\{n_k\}, k = 1,2, \dots, m$. The whole procedure for training, testing, and optimizing a kNN classifier can be summarized as follows for a three-class classification problem:

1) Construct the output labels for all the subjects. All samples from the HC group are labeled with "HC", while all samples from the PD and HD are labeled with "PD" and "HD," respectively.
2) For $l = 1, 2, \dots m$, all the samples of the l^{th} subject are reserved as the test set. The data of all the rest subjects are reserved as the training set. The training and testing procedure in steps 3) and 4) are repeated for all subjects in order.
3) Train the kNN classifier using all the training set data. The distance function is the standardized Euclidean distance. The values from each channel are scaled for both the test set and training set, meaning that the values are divided by a scale value s, which is the standard deviation computed from the corresponding channel.
4) Test the samples in the test set. The distances of each sample in the test set to each sample in the training set are computed using the distance function, and the k nearest neighbors are considered. The final output of the kNN classifier for an n_l ×5 input data set is an n_l×3 matrix, with each row corresponding to one sample and each column representing the likelihood that the current sample belongs to each of the three classes, computed as the posterior probabilities that the sample is of each class. The predicted label for the testing subject is decided as the class for which the likelihood is the maximum.
5) The steps 3) and 4) are repeated for all subjects. The final results are reported as the percentage of subjects correctly classified for each individual class.
6) Tune the parameter k to optimize the classifier so that the percentage of overall subjects correctly classified, as well as the percentage of pathological subjects correctly classified, can be maximized.

After the optimization of the classifier, the machine learning features are defined as the likelihood of the samples to belong to each of the three classes. Therefore, three additional features are added as machine learning features in order to boost the classification performances of each single sample, which is the collection of data from 5 channels in one gait cycle. They are denoted as ρ_{HC}, ρ_{PD}, and ρ_{HD}, representing the probability that the present sample is of class HC, PD, and HD, respectively. The values of three parameters add up to 1 for one single sample.

The main advantage of employing the kNN as a low-level machine learning method for extraction features is that it is able to model the data by combining all the five channels as a whole; moreover, the kNN classifier is simple and robust to noisy training data. By

modelling the data in this way, the similarities between the samples are fully discovered and passed to high-level classifier as generalized information in the form of probability values. The effectiveness of the kNN features will be analyzed along with the statistical features by performing the mentioned three-class classification using two strategies.

6.2.5 Feature Analysis and Classification

The proposed 25 statistical features are analyzed by comparing the means and SDs of the three classes. Classification is conducted on both the sample level and subject level, similar to what was introduced in Chapters 4 and 5. Two classification strategies have been utilized to validate the performance of the proposed framework on multiclass classification:

- The first strategy deploys ANN as the high-level classifier, with which the multiclass classification problem can be naturally solved by training one single unified model. The classification results on samples are passed to the MV block for the subject-level classification. The final results for one single subject are reported as one label and one score. The label is the major class that has received the most voting. The score is the percentage of the samples that are predicted as belonging to the major class. For instance, assuming that the total number of samples for one PD subject is 100, as a sample level classification results, if 10 samples, 60 samples, and 30 samples are predicted as belonging to the HC, PD, and HD classes, respectively, the final label is determined as PD, with a final confidence score of 60%.
- The second strategy deploys the SVM as the high-level classifier. As the SVM is only capable of dealing with binary (2-class) classification problems, a OvR strategy [3] is used. The OvR strategy transforms the multiclass classification problem into several binary classification problems. Compared with the one-vs.-one (OvO) strategy, from which $\frac{K(K-1)}{2}$ binary classifiers are needed for a K-class classification problem, only K binary classifiers are necessary for the OvR strategy. In our study, the OvR strategy was employed. Three classifiers were trained using the SVM for sample level classification, and were required to distinguish HC vs. Rest (PD + HD), PD vs. Rest (HC + HD), and HD vs. Rest (HC + PD), respectively. The label of a single sample is determined by the one classifier out of the three that yields the highest confidence. It is not necessary for the confidence score of any classifier to be larger than 50%. For example, for one sample of a subject, the output confidence scores from the HC vs. Rest, PD vs. Rest and HD vs. Rest SVM classifiers are 20%, 10%, and 5%, respectively. The final label for the current sample is determined as HC, even though the confidence score for the HC class is not higher than 50%. Regarding the subject level classification, the LOSO and MV are applied. To be specific, the entire data set of one subject is reserved as test set, and all the rest of the subjects' data are defined as the training set. The three SVM classifiers are trained based on the OvR policy. Afterwards, all the samples of the test set are passed to each of the classifiers

separately, and the labels are predicted for each sample. The final label for this corresponding test subject is decided based on the class which has the largest percentage of samples as which the test set are predicted. It has to be stressed here that, as the samples are passed to the three distinguished classifiers separately, all the percentages are not required to sum up to 100%. In order to explain this strategy more precisely, an introduction and example are given below.

For each strategy, three feature sets are tested: the features set that only contains the original data after pre-processing, which are the processed stride time, swing time, swing time percentage, double support time, and double support time; the original data after processing plus the extracted 25 statistical features; and the original data after processing plus the extracted 25 statistical features plus 3 kNN likelihood features. As the number of windows is equal to the number of data points, the number of samples for each subject equals the number of data points after processing.

One vs. Rest classification

The OvR strategy involves training a single classifier per class with the samples of the certain class as positive samples and all other samples as negatives. This strategy requires each classifier to output instead of one single label, a real-valued confidence score, indicating the extent to which the test sample belongs to such a class. The training and testing procedures for an OvR learner constructed from a binary classification learner L is as follows:

Inputs:

- L, a learner, which is SVM in our scenario
- Samples X
- Output labels y where $y_i \in \{1,2,\dots,O\}$ is the output label for the sample X_i. Assuming the labels for HC class is 1, the labels for PD class is 2, and for HD class is 3.

Output:

- 3 classifiers $f_l, l = 1,2,3,$ trained with l^{th} class as positive class, and the rest 2 classes as negative. The class labels for all samples are relabeled according to this policy.

The decision-making procedure is accomplished by applying all the 3 classifiers to each of the test samples, and predicting the label of the samples for which the corresponding classifier reports the highest confidence score.

$$\hat{y} = \underset{l\in\{1,2,3\}}{\operatorname{argmax}} f_l(x) \tag{6.3}$$

Based on the sample level classification, the subject level classification is done using the introduced MV approach. The final label of the subject is decided based on the class that has the highest percentage of samples predicted.

6.3 Experimental Results

The proposed approach is evaluated with a classification study conducted on a famous database, which contains gait stride interval parameters from healthy subjects and patients with different types of neurodegenerative diseases.

6.3.1 Database

The database that were utilized in the studies described in this chapter was the PhysioBank database [69], which is a large and fast growing archive of physiological data. In this study, we used the gait dynamics in neurodegenerative diseases database [118], which is a database within the PhysioBank database. The database was built to provide a better understanding of the pathophysiology of neurodegenerative diseases and improve the ability to measure responses to therapeutic interventions. The database aims to help quantify gait dynamics accurately. The records in this database contain stride interval signals from 16 HC subjects, 15 PD patients, and 20 HD patients.

The raw data from this database were obtained using force-sensitive resistors, with the output proportional to the under feet force. By deriving the stride-to-stride measures of the heel contact moments, parameters such as stride time, stance/swing time, and double support time could be calculated.

6.3.2 Feature Analysis

The statistics of all the features, as well the five original data channels are listed in Table 6.2, with the mean and SD for the three classes given separately. The significance of the features and original data are indicated as follows: * indicates a significant feature, with $p < 0.01$, while ** indicates a very significant feature with $p < 0.001$, and the rest of features are non-significant features. The significances were computed by comparing the PD and HD groups with the HC group using the two-sampled t-test. By observing the statistics on the data, several findings can be seen:

- For distinguishing PD and HC classes, one significant feature and 29 very significant features were identified; while two significant features and 29 very significant features were identified between HD and HC groups.
- The stride time T_{St} for the PD and HD groups were very significantly higher than the HC group, indicating a much slower preferred walking speed of the pathological subjects.
- The percentages of the swing phase times for the PD and HD were very significantly lower than those of the HC group, meaning that the patients preferred to spend longer time in the stance phase to ensure better balance. This is consistent with previous findings [16].

Features for stride interval signals						
Feature	**HC**		**PD**		**HD**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
T_{St}	1.0848	0.0852	1.1080**	0.1347	1.1204**	0.1865
T_{Sw}	0.3935	0.0380	0.3713**	0.0489	0.3907	0.0833
P_{Sw}	36.2608	1.9385	33.5892**	3.0122	35.1516**	4.9801
T_{Do}	0.3042	0.0384	0.3736**	0.0966	0.3474**	0.1072
P_{Sw}	28.0487	3.0534	33.4848**	5.5094	30.7473**	7.4011
Max_{St}	1.1443	0.0992	1.2354**	0.3082	1.3060**	0.2867
Min_{St}	1.0413	0.0777	1.0209**	0.1054	0.9645**	0.1083
μ_{St}	1.0848	0.0813	1.1080**	0.1135	1.1204**	0.1520
σ_{St}	0.0243	0.0091	0.0489**	0.0550	0.0819**	0.0733
STC_{St}	4.6016	2.0477	4.3271**	2.4541	5.0096**	2.4041
Max_{Sw}	0.4192	0.0402	0.4224*	0.0463	0.4854**	0.1253
Min_{Sw}	0.3694	0.0356	0.3180**	0.0481	0.3053**	0.0712
μ_{Sw}	0.3935	0.0359	0.3713**	0.0405	0.3907*	0.0574
σ_{Sw}	0.0121	0.0034	0.0251**	0.0121	0.0445**	0.0424
STC_{Sw}	4.9915	2.3579	4.9864	2.6236	5.1318*	2.0987
Max_{PerSw}	37.9996	1.6160	36.9918**	1.8514	40.9073**	3.9913
Min_{PerSw}	34.4454	2.0568	29.8497**	3.5743	29.5194**	6.3932
μ_{PerSw}	36.2608	1.7151	33.5892**	2.3700	35.1516**	3.5453
σ_{PerSw}	0.8733	0.2850	1.7347**	0.7519	2.8368**	2.1458
STC_{PerSw}	5.0068	2.2582	5.1977**	2.4348	5.3247**	1.9848
Max_{Do}	0.3419	0.0405	0.4811**	0.3203	0.4587**	0.1588
Min_{Do}	0.2739	0.0322	0.3153**	0.0597	0.2626**	0.0636
μ_{Do}	0.3042	0.0344	0.3736**	0.0717	0.3474**	0.0911
σ_{Do}	0.0165	0.0057	0.0371**	0.0543	0.0467**	0.0334
STC_{Do}	4.3417	2.1707	4.4370	2.3146	4.5980**	2.4096
Max_{PerDo}	30.6896	3.0696	39.3451**	6.4594	38.1506**	8.7361
Min_{PerDo}	25.6791	2.7536	28.8255**	4.2910	24.2752**	5.4986
μ_{PerDo}	28.0487	2.7820	33.4848**	4.8221	30.7473**	6.3259
σ_{PerDo}	1.2223	0.3793	2.5436**	0.9350	3.4307**	1.8696
STC_{PerDo}	4.9303	2.2649	4.8680	2.4374	4.9967	2.3245
ρ_{HC}	0.4461	0.2500	0.2229**	0.2681	0.2536**	0.2991
ρ_{PD}	0.2138	0.2062	0.3579**	0.2548	0.3376**	0.2643
ρ_{HD}	0.3399	0.1950	0.4190**	0.2419	0.4087**	0.2690

Table 6.2 Statistics on original data and extracted features for stride interval signals.

- Regarding the double support time and percentage, PD and HD subjects spent very significantly larger times in this phase.
- Five statistical features extracted from stride time signals were seen as very significant features for both the PD and HD groups. The pathological groups tended to have larger maximums, means, and SDs, showing a larger variability. PD subjects tended to have less STC, while HD subjects tended to have more than the HC class.
- For the five features extracted from the swing time signals, the PD and HD subjects tended to have smaller swing times and larger SDs. No significant difference was observed in the STC for the PD group, while the STC was significantly larger for the HD group when compared to the PD group.
- All the 5 swing percentage features were identified as very significant features. The PD and HD subjects showed on average a smaller swing percentage but a larger deviation, as was observed for the swing time. The STCs for both pathological groups were very significantly larger than HC.
- Pathological subjects spent a longer time in the double support phase, while having larger stride-to-stride variability, as indicated by the double support time features. More STC could be found for HD subjects compared with HC subjects.

Similar differences were seen from the double support percentage features, but no significant difference existed between the STCs of the groups.

- All the three kNN features were identified as being very significant features, with the HC group having the highest likelihood value for ρ_{HC}, the PD group having the highest likelihood value for ρ_{PD}, and the HD group having the highest likelihood value for ρ_{HD}.

All the major findings mentioned above are consistent with previous findings [66] [50] [65]. Based on those findings, successful extraction of features can be concluded.

6.3.3 Classification Results

6.3.3.1 Results with ANN

The three-class classification results performed on the stride interval signals of 51 subjects using the mentioned database are summarized in Tables 6.3 and 6.4, consisting of sample-level and subject-level classifications.

From the sample-level classification results obtained using ANN-based strategy, it can be observed that, by using only the original data, only 59.49% of total samples were correctly predicted. The overall accuracy increased greatly to 86.58% with the inclusion of extracted statistical features. The highest accuracy, 96.88%, was achieved by using all features, including the original data, the 25 extracted statistical features, and the three likelihoo features. This finding shows the effectiveness of all the extracted features.

		Classification Output		
		HC	**PD**	**HD**
1. Only original data after processing				
Actual Label	**HC**	864 (77.49%)	125 (11.21%)	126 (11.30%)
	PD	236 (24.13%)	519 (53.07%)	223 (22.80%)
	HD	386 (29.56%)	281 (21.52%)	639 (48.93%)
Overall Accuracy: 59.49% of the samples				
2. Original data and statistical features				
Actual Label	**HC**	1013 (94.85%)	21 (1.97%)	34 (3.18%)
	PD	137 (13.10%)	835 (79.83%)	74 (7.07%)
	HD	91 (7.08%)	99 (7.70%)	1095 (85.21%)
Overall Accuracy: 86.58% of the samples				
3. Original data and statistical features and likelihood features				
Actual Label	**HC**	1062 (98.15%)	9 (0.83%)	11 (1.02%)
	PD	9 (0.90%)	956 (95.98%)	31 (3.11%)
	HD	19 (1.44%)	27 (2.04%)	1275 (96.52%)
Overall Accuracy: 96.88% of the samples				

Table 6.3 Classification results on sample level using ANN as classifier.

		Classification Output		
		HC	**PD**	**HD**
1. Only original data after processing				
Actual Label	**HC**	9 (56.25%)	4 (25.00%)	3 (18.75%)
	PD	6 (40.00%)	6 (40.00%))	3 (20.00%)
	HD	9 (45.00%)	8 (40.00%)	3 (15.00%)
Mean Score		0.9534	0.7696	0.6133
Overall Accuracy: 35.29% of the subjects				
2. Original data and statistical features				
Actual Label	**HC**	14 (87.50%)	1 (6.25%)	1 (6.25%)
	PD	2 (13.33%)	6 (40.00%)	7 (46.67%)
	HD	2 (10.00%)	5 (25.00%)	13 (65.00%)
Mean Score		0.7471	0.9586	0.9265
Overall Accuracy: 64.71% of the subjects				
3. Original data and statistical features and likelihood features				
Actual Label	**HC**	14 (87.50%)	1 (6.25%)	1 (6.25%)
	PD	2 (13.33%)	10 (66.67%)	3 (20.00%)
	HD	1 (5.00%)	5 (25.00%)	14 (70.00%)
Mean Score		0.7739	0.8152	0.9079
Overall Accuracy: 74.51% of the subjects				

Table 6.4 Classification results on subject level using ANN as classifier.

By analyzing the three classes individually, it can be seen that the percentage of correctly predicted HC class samples increased from 77.94% to 94.85% and 98.15% after including the statistical and likelihood features, respectively. The percentage for the PD class increased from 53.07% to 79.83% and 95.98%, and the percentage for the HD class increased from 48.93% to 85.21% and 96.52% after involving the statistical features and likelihood features, respectively. The significant increases were consistent with the overall accuracy and have further proved the contributions of all the extracted features. As discussed in Chapter 4, it is essential in gait classification that the proportion of subjects with pathological gait misclassified as being healthy is as low as possible. These numbers for the PD and HD classes were 24.13% and 29.56% respectively, using the first feature set, declining to 13.10% and 7.08%, respectively, with inclusion of 25 statistical features, and further decreasing to 0.90% and 1.44%, respectively after utilizing the three likelihood features for classification. A lower misclassification rate for the PD and HD class samples ensured the robustness and feasibility of the proposed classification scheme and feature extraction method.

Regarding the subject-level classification results, significant increase in the overall accuracy, which is the overall number of subjects correctly classified, was observed for the 2nd and 3rd feature sets compared with the 1st feature set. The total numbers of subjects correctly classified were 18, 33, and 38 for the three feature sets, respectively. This shows the effectiveness of the feature extraction approach and the hybrid classifier. When observing each class separately, the number of HC subjects correctly predicted increased from 9 to 14 and remained the same from the 1st feature set to the 2nd and then the 3rd feature set; the number of PD subjects remained six and increased to ten after including the statistical features and likelihood features, respectively; this number for the HD subjects increased from three to 13 and 14 after involving the statistical and likelihood features, respectively. Consistent with the sample-level classification results, it

can be concluded that the extracted features contribute to the robustness of the classifiers and the promising results of the classification. Similarly, a decrease in the number of PD and HD subjects misclassified as HC subjects can be observed. To be specific, the number of misclassified PD and HD subjects, which were six and nine, respectively, decreased to two and two, respectively, then further decreased to two and one, respectively.

The mean confidence scores, computed as the average of scores of subjects correctly predicted as belonging to their actual classes, are listed as well. For the first feature set, the mean score for the HC class was significantly higher than that of the PD and HD classes, showing a large gap in the confidence level of the classifier between the HC class and the rest of the two classes. The correctly predicted HC subjects are more likely to be correctly and easily predicted than subjects in the other two classes; however, this is not exactly the best expected result, as a higher confidence level on pathological groups is what is desired. After including the extracted statistical features, as well as the likelihood features, the confidence level of the PD and HD groups increased significantly, which makes us believe a higher robustness of the classifiers was achieved.

To summarize, the overall performance on solving the 3-class classification problems significantly improved after involving the extracted statistical and likelihood features. The best results were achieved when all features were used.

6.3.3.2 Results with Features Selection

The PCA-based feature selection method was applied on the full feature set. The feature sets, including all five original data, 25 statistical features, and three likelihood features, were transformed using the PCA into mapped feature matrices with principle components. The principle components were ranked in descending order, and the top $x, x = 1,2, \dots ,33,$ components were selected for classification. The results are reported as the number of selected components vs. the number of subjects correctly predicted, which consists of the numbers for overall classes, as well as the number of pathological subjects. The results are plotted in Fig. 6.4.

The number of overall correctly predicted subjects reached a maximum, 41, when the top 32 components were utilized, and the number of pathological subjects reached a maximum, 26, as well. If we only observe the performance on the pathological subjects, the maximum number of correct predictions was achieved with the top 25 components. If we compare the best results achieved with the PCA-based feature selection method with the best results achieved without it, it can be seen that one more HC subject and two more pathological subjects were correctly predicted with feature selection. It can be concluded that more promising results were obtained with the inclusion of the PCA-based feature selection procedure.

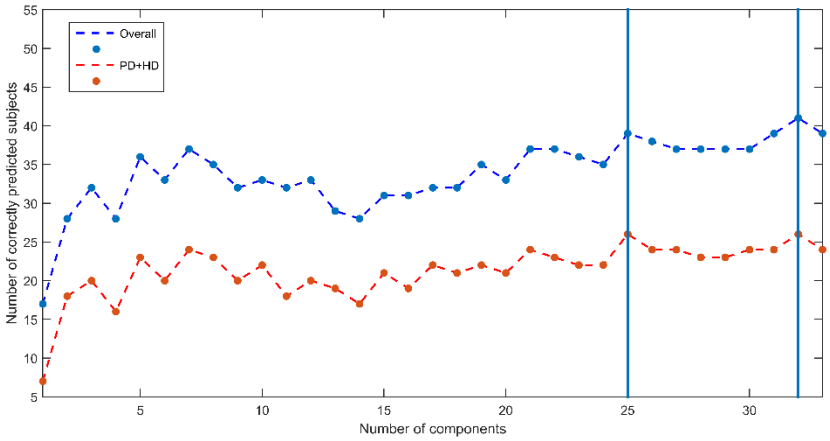


Figure 6.4 Classification results with PCA based features selection.

6.3.3.3 Results with the SVM

The classification results obtained by using the SVM as the high-level classifier and the OvR scheme for samples and subjects are summarized in Table 6.5 and Table 6.6.

According to the sample-level results, the percentage of HC sample correctly predicted increased from 79.62% to 99.92% and 99.48% after including the statistical and likelihood features, respectively. The percentages for PD and HD samples also showed significant increases, i.e., from 48.57% to 99.51% and 99.26% for PD samples, and from 63.69% to 99.39% and 99.39% for HD samples. A similar trend can be seen in all the corresponding "Rest" classes. The promising results prove the effectiveness of the classification scheme and lay a good foundation for the subject-level classification.

From the subject-level results, it is obvious that the overall accuracy increased after statistical features and likelihood features were included. The number of correctly predicted HC subjects increased by one and two after the 1st feature set when the 2nd and 3rd feature sets were used. This number increased from six to eight and then remained the same for PD subjects, while it increased from 15 to 16 and 17 for HD subjects. The number of subjects misclassified as HC subjects for PD and HD classes dropped from six to three. The best results were achieved when all features and the SVM were used, and the results were comparable with those achieved using all features and the ANN, namely, more HC and PD subjects and less HD subjects were correctly predicted.

Regarding the scores for individual classes, there were significant improvements for all classes using the extracted features compared with using only the original data. The improvement was seen when comparing the 3rd feature set with the 2nd feature set for HC and PD classes, and a decrease was seen in the HD class. This indicates that the value of the score is not proportionally associated with the accuracy.

The major findings obtained from the SVM-based classification are consistent with those from the ANN-based classification. Both strategies have proved the necessity of feature extraction, and the effectiveness of the proposed features. The best classification result

was achieved when all features were used along with the PCA-based feature selection methods.

		Classification Output	
1. Only original data after processing			
		HC	**Rest**
Actual Label	**HC**	8745 (79.62%)	2239 (20.38%)
	Rest	3258 (14.16%)	19748 (85.84%)
		PD	**Rest**
Actual Label	**PD**	4826 (48.57%)	5110 (51.43%)
	Rest	1732 (7.20%)	22322 (92.80%)
		HD	**Rest**
Actual Label	**HD**	8321 (63.69%)	4744 (36.31%)
	Rest	3673 (17.55%)	17252 (82.45%)
2. Original data and statistical features			
		HC	**Rest**
Actual Label	**HC**	10976 (99.92%)	9 (0.08%)
	Rest	137 (0.60%)	22868 (99.40%)
		PD	**Rest**
Actual Label	**PD**	9885 (99.51%)	49 (0.49%)
	Rest	3 (0.01%)	24053 (99.99%)
		HD	**Rest**
Actual Label	**HD**	12987 (99.39%)	80 (0.61%)
	Rest	59 (0.28%)	20864 (99.72%)
3. Original data and statistical features and kNN features			
		HC	**Rest**
Actual Label	**HC**	10927 (99.48%)	57 (0.52%)
	Rest	95 (0.41%)	22911 (99.59%)
		PD	**Rest**
Actual Label	**PD**	9860 (99.26%)	74 (0.74%)
	Rest	37 (0.15%)	24019 (99.85%)
		HD	**Rest**
Actual Label	**HD**	12982 (99.39%)	80 (0.61%)
	Rest	80 (0.38%)	20848 (99.62%)

Table 6.5 Classification results on sample level using the SVM.

		Classification Output		
		HC	**PD**	**HD**
1. Only original data after processing				
Actual Label	**HC**	10 (62.5%)	2 (12.5%)	4 (25.00%)
	PD	2 (13.33%)	6 (40.00%)	7 (46.67%)
	HD	4 (20.00%)	1 (5.00%)	15 (75.00%)
Mean Score		0.5452	0.4372	0.4756
Overall Accuracy: 60.78% of the subjects				
2. Original data and statistical features				
Actual Label	**HC**	11 (68.75%)	1 (6.25%)	4 (25.00%)
	PD	1 (6.67%)	8 (53.33%)	6 (40.00%)
	HD	2 (10.00%)	2 (10.00%)	16 (80.00%)
Mean Score		0.5716	0.7297	0.8171
Overall Accuracy: 68.63% of the subjects				
3. Original data and statistical features and kNN features				
Actual Label	**HC**	13 (81.25%)	3 (18.75%)	0 (0.00%)
	PD	1 (6.67%)	8 (53.33%)	6 (40.00%)
	HD	2 (10.00%)	1 (5.00%)	17 (85.00%)
Mean Score		0.6129	0.7484	0.6467
Overall Accuracy: 74.51% of the subjects				

Table 6.6 Classification results on subject level using the SVM.

6.4 Conclusions

In this chapter, the proposed machine learning framework for gait classification has been validated on stride interval signals by solving complex 3-class classification problems. Traditional statistical features, as well as kNN-based machine learning features, were extracted and served as input for conducting the classification. Two strategies were implemented in this chapter according the requirement of the gait classification problem, including one ANN-based strategy and one SVM- and OvR-based strategy. The validation results, consisting of the sample-level results and the subject-level results, have suggested a promising effectiveness of the extracted features, as well as the efficiency of the proposed framework.

7. Conclusions and Outlook

7.1 Thesis Summary

The aim of the thesis is the development, implementation and validation, of a machine learning framework for solving different types of gait classification problems. The framework contains several novel functions that are designed for overcoming some of the existing limitations in the field of machine learning-based gait classification. The feasibility of this framework has been evaluated on three types of gait signals, and the several clinical case studies have proved its usability in clinical applications such as gait rehabilitation and management of treatment of some diseases that cause gait disorders such as Parkinson Disease.

Chapter 1 pointed out the motivation, main focus, and contribution of the dissertation. Chapter 2 provided a technical background of the use of machine learning methods to conduct classification. Chapter 3 discussed the current directions and the state-of-the-art approaches for machine learning based classifications using gait classification as the primary research subject of this thesis. The classification framework is hereby proposed emphasizing some of its main novelties. These are: it is considered as a general framework that can be applied to different types of gait data taking advantage of the semi-periodic characteristic of gait signals; it contains a hybrid classifier that can boost the overall classification performance by extracting additional features using machine learning techniques; it has a sample generation block that segments the data series into smaller units for performing a sample level classification using a larger scale of training data; it outputs an additional confidence score, which is numerical value that reflects the level of gait quality.

In Chapters 4 to 6, the proposed classification framework was used to solve practical gait classification problems for three types of gait signals. In Chapter 4, the joint angle signals were studied. Based on the characteristics of these signals, novel gait variability features were extracted using four distance functions, which measure the differences between the shapes of two gait cycles of joint angle trajectories. In addition to representation of the differences between the cycles gait trajectories, the shapes of individual cycles of the joint angle trajectory were modelled with GMM, and the parameters derived from the model served as machine learning features for the sample-level classification conducted on a pair of cycles using the gait pairing method. The results showed the feasibility of the framework for determining joint angle signals. In Chapter 5, the proposed classification framework was applied to trunk acceleration signals for the purpose of gait balance analysis. Features were extracted using statistical and signal processing techniques, and the regression method was used for modelling the accelerations phase plot in two

dimensions. The sliding window approach was utilized for sample generation with a fixed number of steps equal to window size. Achieved classification results further proved the effectiveness of the proposed framework, including the data processing methods developed in this chapter. In Chapter 6, the stride interval signals were processed using the proposed framework for classification. Beside the basic statistical features extracted from each dimension separately, the five dimensional signals were modelled with kNN to investigate the relationship between signal strides. The classifications have been conducted to solve a multiclass problem using two strategies, and both have achieved promising results. A short summary of the main classification work of Chapters 4-6 can be seen in Table 7.1.

Data Type	Gait Patterns	Data Source	ML strategy	Sample Generation	Features	Best Accuracy
Joint Angle	HC vs. PT	AHRS	GMM +SVM&ANN	Segmentation & Paring	Variability +Shape	98.28%
Trunk Acceleration	HC vs. PT PD vs. PNP	AHRS	Regression +SVM&ANN	Segmentation &Sliding Window	Time & Frequency + Ellipse	98.15%
Stride Interval	HC vs. PD vs. HD	Database	kNN +SVM&ANN	Sliding Window	Statistical +kNN	80.39%

Table 7.1. Summary of the classification work in Chapter 4-6.

7.2 Main Contributions

The main limitations of previous work listed in the introduction have been carefully considered in this thesis, and the solutions for overcoming these limitations were proposed. The major technical and scientific contributions of the dissertation can be summarized as follows:

1) **Gait classification framework:** The thesis came up with a gait classification framework that could be used for a wide range of applications and scenarios. Three types of gait data were utilized to validate the framework. This proved the potential for usage of the framework on other gait data types, such as GRF and joint moment, as well. More importantly, as the proposal of the framework is in accordance with the semi-periodic characteristic of gait, it is to expect that some of the proposed procedures are also applicable to data of movements different from walking, which are also semi-periodic, such as running, exercising, and movements related to other kinds of sport.
2) **Solution to lack of samples: In** this thesis a valuable solution to the problem of a lack of samples (observations), which is often encountered in many human motion related research areas, has been proposed. Although the number of subjects is usually restricted for human motion related experiments, such as walking, running, and most of the exercise and sport movements, they all tend to have certain repetitiveness and semi-periodic behaviors. Once the signals collected over a long period of time are segmented into smaller fragments, a larger number of samples can be available for performing classifications.

3) **Indicator for confidence level of classification:** Instead of providing only a binary classification result, the proposed framework can also output a confidence score, which is a representation of the degree of membership to the predicted class. This score is able to gives additional information to assess the quality of gait.
4) **Exploration in clinical applications:** The classification strategies and related feature extraction techniques are mainly developed for the ease and support of human daily life, especially the quality of health and condition of healthcare. The proposed framework was successfully integrated into practical clinical applications, such as the management of rehabilitation and of treatment of patients suffering from illnesses that cause gait disorders. Even though the applications explored in this thesis are restricted to gait analysis, it is to expect that the framework and proposed strategy would have a high potential for use in other application, especially human motion related applications.
5) **Novel feature extraction techniques:** Besides using some of the most well-known signal processing and feature extraction methods, this thesis aimed to create and apply novel methods for processing gait signals for feature extraction. For instance, the least square regression method was implemented to model the trunk acceleration signals in two dimensions, and the GMM was employed to model the shape of joint angle signals. This is novel application of GMM and the least square regression method in the gait analysis field. Furthermore, the sliding window approach was applied in two scenarios for grouping and streaming information and generating samples, which also showed a very promising performance. Moreover, the use of distance functions in Chapter 4 has also introduced new concepts for understanding the variability and fluctuation of gait signals.
6) **Machine learning features for better classification accuracy:** The use of parallel structured hybrid systems is the first application of such systems in the field of gait analysis. The achieved results have provided new insights regarding the future direction of classification, which is the usability of systems with multiple classifiers for gait and other human motion analysis. Different combinations have been validated in this thesis, such as GMM plus SVM/ANN, kNN plus SVM/ANN, and regression plus SVM/ANN. Although more combinations need to be tested in order to make further conclusions, the present results have made a case for their further study.

7.3 Outlook

This thesis is not without some limitations. First, more types of gait data, such as, kinetic parameters and electrical parameters (EMG) should be further validated; secondly, the types and level of severity of diseases were not categorized during the classification. One type of NND can have different types and levels of severity explicitly defined using medical metrics. In this thesis, all subjects who belonged to one class were considered as the same class of sample without distinguishing them according to their specific types. Therefore, in future work, it would be worthwhile to conduct classification not only

between healthy and unhealthy gait, and also gait patterns of different types of the same disease; thirdly, the underlying relationship between the extracted features and the clinical explanations were not discussed in detail. Exploring such associations more comprehensively would be beneficial to the development of gait related clinical diagnoses and rehabilitation approaches.

Based on these current directions of research, and the contributions of this thesis, the following outlook can be made: 1) more efforts should be directed towards the combination and optimization of machine learning strategies. Furthermore, the use of different hybrid classifiers in solving classification problems for a larger scale of data and a broader range of human motions should be explored; 2) owing to the rapid development of MEMS technologies, it will not be long before data recording of human motion is achieved with much faster processing speeds and higher accuracy, therefore more advanced gait analysis hardware, as well as real-time applications shall be developed; 4) based on the results of this thesis, more research work can be continued on other types of gait data, as well as on other human motions, such as those in sports, and exercises; and 5) the current studies mainly focus on clinical conditions, and in future work, more outdoor and at-home scenarios can be investigated using machine learning for further healthcare-related applications .

Abbreviations

Acc	Accuracy
AD	Alzheimer's Disease
AHRS	Attitude and Heading Reference System
ALS	Amyotrophic Lateral Sclerosis
ANN	Artificial Neural Network
AP	Anterior-posterior
AS	Asymptomatic
AUC	Area Under the Curve
BN	Bayesian Networks
CAD	Computer-Aided Diagnosis
COM	Center of Mass
CP	Cerebral Palsy
CRPS	Complex Regional Pain Syndrome
DT	Decision Tree
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FNN	Neural Network
FOG	Freezing of Gait
GC	Gait Cycle
GMM	Gaussian Mixture Model
GRF	Ground Reaction Force
HC	Healthy Control
HD	Huntington's Disease
HO	Hip Osteoarthritis
IMU	Inertial Measurement Unit
KFD	Kernel Fisher Discriminant
kNN	K Nearest Neighbors
kNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LOSO	Leave One Subject Out
MCorr	Maximum-Cross-Correlation
MDTW	Mean-Dynamic-Time-Warping
MEMS	Microelectromechanical Systems
ML	Machine Learning
ML	Medio-Lateral
MLP	Multilayer Perceptron
MS	Multiple Sclerosis
MSub	Mean-Subtraction
MV	Majority Voting
NB	Naive Bayesian
NDD	Neurodegenerative Diseases
NDD	Neurodegenerative Disease
NPV	Negative Predictive Value
OA	Osteoarthritis
OBW	Occupied Bandwidth
OvR	One vs. Rest
PCA	Principal Component Analysis
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
PD	Parkinson's Disease
PFPS	Patellofemoral Pain Syndrome
PNP	Polyneuropathy

PPV	Positive Predictive Value
PSD	Power Spectral Density
PT	Pathological (gait)
RBF	Radial Basis Function
RF	Random Forest
RMSD	Root-Mean-Square-Deviation
ROI	Region of Interest
ROM	Range of Motion
SD	Standard Deviation
Sen	Sensitivity
SLR	Simple Logistic Regression
STC	Signal Turn Count
SVM	Support Vector Machine
TNR	True Negative Rate
TPR	True Positive Rate

References

[1] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning,* vol. 30, pp. 271-274, 1998.

[2] M. Awad and R. Khanna, Efficient learning machines: Theories, concepts, and applications for engineers and system designers, New York: Apress, 2015.

[3] B. Christopher, Pattern recognition and machine learning, Springer, 2006.

[4] A. Ethem, Introduction to machine learning, London: The MIT Press, 2010.

[5] Y. Xia, Q. Gao and Q. Ye, "Classification of gait rhythm signals between patients withneuro-degenerative diseases and normal subjects: Experiments withstatistical features and different classification models," *Biomedical Signal Processing and Control,* vol. 18, pp. 254-262, 2015.

[6] D. T. Lai, R. K. Begg and M. Palaniswami, "Computational intelligence in gait research: A perspective on current applications and future challenges," *IEEE Transactions on Information Technology in Biomedicine,* vol. 13, no. 5, pp. 687 - 702, 2009.

[7] S. Dutta, D. Ghosh and S. Samanta, "Non linear approach to study the dynamics of neurodegenerative diseases by Multifractal Detrended Cross-correlation Analysis—A quantitative assessment on gait disease," *Physica A: Statistical Mechanics and its Applications,* vol. 448, pp. 181-195, 2016.

[8] M. D. Djuric-Jovicic, N. S. Jovicic, S. M. Radovanovic, I. D. Stankovic, M. B. Popovic and V. S. Kostic, "Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 22, no. 3, pp. 685-694, 2014.

[9] E. Baratin, L. Sugavaneswaran, K. Umapathy, C. Ioana and S. Krishnan, "Wavelet-based characterization of gait signal for neurological abnormalities," *Gait & Posture,* vol. 41, no. 2, pp. 634-639, 2015.

[10] S.-H. Lee and J. S. Lim, "Parkinson's disease classification using gait characteristics and wavelet-based feature extraction," *Expert Systems with Applications,* vol. 39, pp. 7338-7344, 2012.

[11] A. N. Akansu and R. A. Haddad, Multiresolution signal decomposition: transforms, subbands, and wavelets, Boston, USA: Academic Press, 1992.

[12] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophical Magazine,* vol. 2, no. 11, pp. 559-572, 1901.

[13] S. Raschka, "MLXTEND," [Online]. Available: https://sebastianraschka.com/pdf/software/mlxtend-latest.pdf.

[14] V. N. Vapnik, The nature of statistical learning theory, New York: Springer-Verlag, 1995.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, no. 3, pp. 273-297, 1995.

[16] M. W. Whittle, Gait analysis: An introduction, Philadelphia, PA, USA: Elsevier, 2007.

[17] R. D. Sanders and P. M. Gillig, "Gait and its assessment in psychiatry," *Psychiatry (Edgmont),* vol. 7, no. 7, pp. 38-43, 2010.

[18] "Gait Abnormalities," Stanford Medicine 25, [Online]. Available: https://stanfordmedicine25.stanford.edu/the25/gait.html.

[19] L. Wolfson, "Gait and balance dysfunction: a model of the interaction of age and disease," *Neuroscientist,* vol. 7, no. 2, pp. 178-183, 2001.

[20] K. Christopher, Clinical gait analysis: Theory and practice, London, UK: Elsevier Health Sciences, 2006.

[21] R. A. Walsh, R. M. de Bie and S. H. Fox, Movement disorders, New York, USA: Oxford University Press, 2013.

[22] A. Shiratsu and H. Coury, "Reliability and accuracy of different sensors of a flexible electrogoniometer," *Clinical Biomechanics,* vol. 18, no. 7, pp. 682-684, 2003.

[23] "Vicon," [Online]. Available: https://www.vicon.com/.

[24] "Kinovea," [Online]. Available: https://www.kinovea.org/.

[25] S. Natarajan, X. Wang, M. Spranger and A. Gräser, "Reha@home - a vision based markerless gait analysis system for rehabilitation at home," in *13th IASTED International Conference on Biomedical Engineering (BioMed)*, Innsbruck, Austria, 2017.

[26] "AIMS: Navigation IMU and VRU product datasheet," [Online]. Available: http://www.aims.se/.

[27] "Xsens: MTi-100 IMU, MTi-200 VRU and MTi-300 AHRS product datasheets," [Online]. Available: https://www.xsens.com/products/mti-100-series/.

[28] "Hillcrest labs: IMU/AHRS FSM-9 module," [Online]. Available: http://hillcrestlabs.com/products/fsm-9/.

[29] T. Seel, J. Raisch and T. Schauer, "IMU-based joint angle measurement for gait analysis," *Sensors,* vol. 14, no. 4, pp. 6891-6909, 2014.

[30] W. Kong, S. Sessa, S. Cosentino, M. Zecca, K. Salto, C. Wang, U. Imtiaz, Z. Lin, L. Bartolomeo, H. Ishii, T. Ikai and A. Takanishi, "Development of a real-time IMU-based motion capture system for gait rehabilitation," in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Shenzhen, China, 2013.

[31] N. Sharma and A. Dani, "Nonlinear estimation of gait kinematics during functional electrical stimulation and orthosis-based walking," in *2014 American Control Conference (ACC)*, Portland, Oregon, USA, 2014.

[32] L. Palmerini, L. Rocchi, S. Mellone, F. Valzania and L. Chiari, "Feature selection for accelerometer-based posture analysis in Parkinson's disease," *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, no. 3, pp. 481-490, 2011.

[33] M. Mancini, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt and L. Chiari, "Trunk accelerometry reveals postural instability in untreated Parkinson's disease," *Parkinsonism and Related Disorders,* vol. 17, no. 7, pp. 557-562, 2011.

[34] A. E. Patla, A. Adkin and T. Ballard, "Online steering: coordination and control of body center of mass, head and body reorientation," *Experimental Brain Research,* vol. 129, no. 4, pp. 629-634, 1999.

[35] C. Hodt-Billington, J. L. Helbostad and R. Moe-Nilssen, "Should trunk movement or footfall parameters quantify gait asymmetry in chronic stroke patients?," *Gait Posture,* vol. 27, no. 4, pp. 552-558, 2008.

[36] J. S. Brach, D. McGurl, J. M. Vanswearingen, S. Perera, R. Cham and S. Studenski, "Validation of a measure of smoothness of walking," *The journals of gerontology. Series A, Biological sciences and medical sciences,* vol. 66, no. 1, pp. 136-141, 2011.

[37] J. J. Kavanagh and H. B. Menz, "Accelerometry: a technique for quantifying movement patterns during walking," *Gait Posture,* vol. 28, no. 1, pp. 1-15, 2008.

[38] E. Sejdić, K. A. Lowry, J. Bellanca, M. S. Redfern and J. S. Brach, "A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 22, no. 3, pp. 603-612, 2014.

[39] M. Kyrarini, X. Wang and A. Gräser, "Comparison of vision-based and sensor-based systems for joint angle gait analysis," in *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Turin, Italy, 2015.

[40] D. H. Sutherland, "The evolution of clinical gait analysis. Part II kinematics," *Gait & Posture,* vol. 16, no. 2, pp. 159-179, 2002.

[41] T. Chau, "A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods," *Gait and Posture,* vol. 13, no. 1, pp. 49-66, 2001.

[42] T. Chau, "A review of analytical techniques for gait data. Part 2: neural network and wavelet methods," *Gait and Posture,* vol. 13, no. 2, pp. 102-120, 2001.

[43] A. Muro-de-la-Herran, B. Garcia-Zapirain and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, Highlighting clinical applications," *Sensors,* vol. 14, no. 2, pp. 3362-3394, 2014.

[44] B. Ao, G. Fang, Y. Wang, L. Song and Z. Yang, "Healthcare algorithms by wearable inertial sensors: a survey," *China Communications ,* vol. 12, no. 4, pp. 1-12, 2015.

[45] S. Chen, J. Lach, B. Lo and G.-Z. Yang, "Toward pervasive gait analysis with wearable sensors: A systematic review," *IEEE Journal of Biomedical and Health Informatics,* vol. 20, no. 6, pp. 1521 - 1537, 2016.

[46] R. Caldas, M. Mundt, W. Potthast, F. B. Neto and B. Markert, "A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms," *Gait & Posture,* vol. 57, pp. 204-210, 2017.

[47] Z. Gandomkar and F. Bahrami, "Method to classify elderly subjects as fallers and non-fallers based on gait energy image," *Healthcare Technology Letters,* vol. 1, no. 3, pp. 110-114, 2014.

[48] D. Joshi, A. Khajuria and P. Joshi, "An automatic non-invasive method for Parkinson's disease classification," *Computer Methods and Programs in Biomedicine,* vol. 145, pp. 135-145, 2017.

[49] S. Bilgin, "The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects," *Biomedical Signal Processing and Control,* vol. 31, pp. 288-294, 2017.

[50] W. Zeng and C. Wang, "Classification of neurodegenerative diseases using gait dynamics via deterministic learning," *Information Sciences,* vol. 317, pp. 246-258, 2015.

[51] C. Pradhan, M. Wuehr, F. Akrami, M. Neuhaeusser, S. Huth, T. Brandt, K. Jahn and R. Schniepp, "Automated classification of neurological disorders of gait using spatio-temporal gait parameters," *Journal of Electromyography and Kinesiology,* vol. 25, pp. 413-422, 2015.

[52] D. Laroche, A. Tolambiya, C. Morisset, J. F. Maillefert, R. M. French, P. Ornetti and E. Thomas, "A classification study of kinematic gait trajectories in hip osteoarthritis," *Computers in Biology and Medicine,* vol. 55, pp. 42-48, 2014.

[53] M. Yang, H. Zheng, H. Wang, S. McClean, J. Hall and N. Harris, "A machine learning approach to assessing gait patterns for Complex Regional Pain Syndrome," *Medical Engineering & Physics,* vol. 34, pp. 740-746, 2012.

[54] W.-H. Wang, Y.-L. Hsu, M.-C. Pai, C.-H. Wang, C.-Y. Wang, C.-W. Lin, H.-L. Wu and P.-C. Chung, "Alzheimer's disease classification based on gait information," in *2014 International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, 2014.

[55] B. R. Greene, S. Rutledge, I. McGurgan, C. McGuigan, K. O'Connell, B. Caulfield and N. Tubridy, "Assessment and classification of early-stage multiple sclerosis with inertial sensors: comparison against clinical measures of disease state," *IEEE Journal of Biomedical and Health Informatics,* vol. 19, no. 4, pp. 1356-1361, 2015.

[56] D. T. Lai, P. Levinger, R. K. Begg, W. L. Gilleard and M. Palaniswami, "Automatic recognition of gait patterns exhibiting patellofemoral pain syndrome ssing a support vector machine approach," *IEEE Transactions on Information Technology in Biomedicine,* vol. 13, no. 5, pp. 810-817, 2009.

[57] L. V. Gestel, T. D. Laet, E. D. Lello, H. Bruyninckx, G. Molenaers, A. V. Campenhout, E. Aertbelieen, M. Schwartz, H. Wambacq, P. D. Cock and K. Desloovere, "Probabilistic gait classification in children with cerebral palsy: A Bayesian approach," *Research in Developmental Disabilities,* vol. 32, no. 6, p. 2542–2552, 2011.

[58] J.-S. Wang, C.-W. Lin, Y.-T. C. Yang and Y.-J. Ho, "Walking pattern classification and walking distance estimation algorithms using gait phase information," *IEEE Transactions on Biomedical Engineering,* vol. 59, no. 10, pp. 2884 - 2892, 2012.

[59] S. Ilias, N. M. Tahir, R. Jailani and C. Z. C. Hasan, "Classification of autism children gait patterns using neural network and support vector machine," in *2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Batu Feringghi, Malaysia, 2016.

[60] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge and D. C. Ackland, "Classification of Parkinson's disease gait using spatial-temporal gait features," *IEEE Journal of Biomedical and Health Informatics,* vol. 19, no. 6, pp. 1794-1802, 2015.

[61] J. Pauk and K. Minta-Bielecka, "Gait patterns classification based on cluster and bicluster analysis," *Biocybernetics and Biomedical Engineering,* vol. 36, pp. 391-396, 2016.

[62] P. Ren, S. Tang, F. Fang, L. Luo, L. Xu, M. L. Bringas-Vega, D. Yao, K. M. Kendrick and P. A. Valdes-Sosa, "Gait rhythm fluctuation analysis for neurodegenerative diseases by empirical mode decomposition," *IEEE Transactions on Biomedical Engineering,* vol. 64, no. 1, pp. 52-60, 2017.

[63] Y. Ma, R. Fallahzadeh and H. Ghasemzadeh, "Glaucoma-specific gait pattern assessment using body-worn sensors," *IEEE Sensors Journal,* vol. 16, no. 16, pp. 6406 - 6415, 2016.

[64] M. Banaie, M. Pooyan and M. Mikaili, "Introduction and application of an automatic gait recognition method to diagnose movement disorders that arose of similar causes," *Expert Systems with Applications,* vol. 38, pp. 7359-7363, 2011.

[65] P. Ren, W. Zhao, Z. Zhao, M. L. Bringas-Vega, P. A. Valdes-Sosa and K. M. Kendrick, "Analysis of gait rhythm fluctuations for neurodegenerative diseases by phase synchronization and conditional entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 24, no. 2, pp. 291-299, 2016.

[66] Y. Wu and S. Krishnan, "Statistical analysis of gait rhythm in patients with Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 18, no. 2, pp. 150-158, 2010.

[67] V. Agostini, G. Balestra and M. Knaflitz, "Segmentation and classification of gait cycles," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 22, no. 5, pp. 946-952, 2014.

[68] N. Mezghani, S. Husse, K. Boivin, K. Turcot, R. Aissaoui, N. Hagemeister and J. A. d. Guise, "Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier," *IEEE Transactions on Biomedical Engineering,* vol. 55, no. 3, pp. 1230-1232, 2008.

[69] "PhysioBank Database," [Online]. Available: https://physionet.org/physiobank/database/.

[70] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," *Neural Computation,* vol. 8, no. 7, pp. 1341-1390, 1996.

[71] M. P. Ponti Jr., "Combining classifiers: from the creation of ensembles to the decision fusion," in *2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials*, Alagoas, Brazil, 2011.

[72] R. Begg and J. Kamruzzaman, "A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data," *Journal of Biomechanics,* vol. 38, no. 3, pp. 401-408, 2005.

[73] H. H. Manap, N. M. Tahir and A. I. M. Yassin, "Anomalous gait detection based on support vector machine," in *IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, Penang, Malaysia, 2011.

[74] N. M. Tahir and H. H. Manap, "Parkinson disease gait classification based on machine learning approach," *Journal of Applied Sciences,* vol. 12, no. 2, pp. 180-185, 2012.

[75] B. Koopman, E. v. Asseldonk and H. v. d. Kooij, "Speed-dependent reference joint trajectory generation for robotic gait support," *Journal of Biomechanics,* vol. 47, no. 6, pp. 1447-1458, 2014.

[76] T. Öberg, A. Karsznia and K. Öberg, "Joint angle parameters in gait: reference data for normal subjects, 10-79 years of age," *Journal of Rehabilitation Research and Development,* vol. 31, no. 3, pp. 199-213, 1994.

[77] D. A. Winter, A. E. Patla and J. S. Frank, "Assessment of balance control in humans," *Medical Progress through Technology,* vol. 16, no. 1-2, pp. 31-51, 1990.

[78] S. Sprager and M. B. Juric, "Inertial sensor-based gait recognition: A review," *Sensors,* vol. 15, no. 9, pp. 22089-22127, 2015.

[79] E. Billauer, "peakdet: Peak detection using MATLAB," 2012. [Online]. Available: http://www.billauer.co.il/peakdet.html.

[80] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," *SIAM Journal on Numerical Analysis,* vol. 17, no. 2, pp. 238-246, 1980.

[81] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 26, no. 1, pp. 43-49, 1978.

[82] S. Petre and R. Moses, Spectral Analysis of Signals, New York: Prentice Hall, 2005.

[83] G. McLachlan and D. Peel, Finite mixture models, Wiley, 2000.

[84] N. Vlassis and A. Likas, "A greedy EM algorithm for gaussian mixture learning," *Neural Processing Letters,* vol. 16, no. 1, pp. 77-87, 2002.

[85] S. Calinon, F. Guenter and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) ,* vol. 37, no. 2, pp. 286-298, 2007.

[86] R. K. Ibrahim, E. Ambikairajah, B. Celler, N. H. Lovell and L. Kilmartin, "Gait patterns classification using spectral features," in *Signals and Systems Conference, 208. (ISSC 2008). IET Irish*, 2008.

[87] H. Zhang, Y. Liu, J. Liang, J. Cao and L. Zhang, "Gaussian mixture modeling in stroke patients' rehabilitation EEG data analysis," in *2013 35th Annual International Conference of the IEEE*, Osaka, Japan, 2013.

[88] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B,* vol. 39, no. 1, pp. 1-38, 1977.

[89] J. Zhuang, X. Ning, X. Yang, F. Hou and C. Huo, "Decrease in Hurst exponent of human gait with aging and neurodegenerative diseases," *Chinese Physics B,* vol. 17, no. 3, pp. 852-856, 2008.

[90] H. Zheng, M. Yang, H. Wang and S. McClean, "Machine learning and statistical approaches to support the discrimination of neurodegenerative diseases based on gait analysis," *Studies in Computational Intelligence,* vol. 189, pp. 57-70, 2009.

[91] J. M. Hausdorff, "Gait variability: methods, modeling and meaning," *Journal of NeuroEngineering and Rehabilitation,* vol. 2, no. 19, 2005.

[92] O. Kuzmicheva, S. M. Focke, U. Krebs, M. Spranger, S. Moosburner, B. Wagner and A. Gräser, "Overground robot based gait rehabilitation system MOPASS - overview and first results from usability testing," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 2016.

[93] J. B. Saunders, V. T. Inman and H. D. Eberhart, "The major determinants in normal and pathological gait," *Journal of Bone and Joint Surgery,* vol. 35, no. 3, pp. 543-558, 1953.

[94] M. J. Floor-Westerdijk, H. M. Schepers, P. H. Veltink, E. H. F. van Asseldonk and J. H. Buurke, "Use of inertial sensors for ambulatory assessment of center-of-mass displacements during walking," *IEEE Transactions on Biomedical Engineering,* vol. 59, no. 7, pp. 2080-2084, 2012.

[95] R. P. Hubble, G. A. Naughton, P. A. Silburn and M. H. Cole, "Wearable sensor use for assessing standing balance and walking stability in people with Parkinson's disease: a systematic review," *PloS One,* vol. 10, no. 4, p. e0123705, 2015.

[96] Y.-L. Hsu, P.-C. Chung, W.-H. Wang, M.-C. Pai, C.-Y. Wang, C.-W. Lin, H.-L. Wu and J.-S. Wang, "Gait and balance analysis for patients with Alzheimer's disease using an inertial-sensor-based wearable instrument," *IEEE Journal of Biomedical and Health Informatics,* vol. 18, no. 6, pp. 1822-1830, 2014.

[97] M. Demonceau, A.-F. Donneau, J.-L. Croisier, E. Skawiniak, M. Boutaayamou, D. Maquet and G. Garraux, "Contribution of a Trunk Accelerometer System to the Characterization of Gait in Patients With Mild-to-Moderate Parkinson's Disease," *IEEE Journal of Biomedical and Health Informatics,* vol. 19, no. 6, pp. 1803-1808, 2015.

[98] M. Yoneyama, Y. Kurihara, K. Watanabe and H. Mitoma, "Accelerometry-based gait analysis and its application to Parkinson's disease assessment- part 2: a new measure for quantifying walking behavior.," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 21, no. 6, pp. 999-1005, 2013.

[99] L. Rocchi, L. Palmerini, A. Weiss, T. Herman and J. M. Hausdorff, "Balance testing with inertial sensors in patients with Parkinson's disease: assessment of motor subtypes," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 22, no. 5, pp. 1064-1071, 2014.

[100] E. Sejdic, K. A. Lowry, J. Bellanca, S. Perera, M. S. Redfern and J. S. Brach, "Extraction of stride events from gait accelerometry during treadmill walking," *IEEE Journal of Translational*

Engineering in Health and Medicine, vol. 4, 2015.

[101] M. Yoneyama, Y. Kurihara, K. Watanabe and H. Mitoma, "Accelerometry-based gait analysis and its application to Parkinson's disease assessment--part 1: detection of stride event," *IEEE Transactions on Neutal Systems and Rehabilitation Engineering,* vol. 22, no. 3, pp. 613-622, 2014.

[102] T. E. Prieto, J. B. Myklebust, R. G. Hoffmann, E. G. Lovett and B. M. Myklebust, "Measures of postural steadiness: Differences between healthy young and elderly adults," *IEEE Transactions on Biomedical Engineering,* vol. 43, no. 9, pp. 956-966, 1996.

[103] K. J. Kim, J. Lučarević, C. Bennett, I. Gaunaurd, R. Gailey and V. Agrawal, "Testing the assumption of normality in body sway area calculations during unipedal stance tests with an inertial sensor," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Florida, USA, 2016.

[104] P. Esser, H. Dawes, J. Collett and K. Howells, "Insights into gait disorders: Walking variability using phase plot analysis, Parkinson's disease," *Gait and Posture,* vol. 38, no. 4, pp. 648-652, 2013.

[105] O. Gal, "fit_ellipse," 2003. [Online]. Available: https://de.mathworks.com/matlabcentral/fileexchange/3215-fit-ellipse?requestedDomain=www.mathworks.com.

[106] C. E. Bauby and A. D. Kuo, "Active control of lateral balance in human walking," *Journal of Biomechanics ,* vol. 33, no. 11, pp. 1433-1440, 2000.

[107] K. A. Lowry, A. L. Smiley-Oyen, A. J. Carrel and J. P. Kerr, "Walking stability using harmonic ratios in Parkinson's disease," *Movement Disorders,* vol. 24, no. 2, pp. 261-267, 2009.

[108] C. H. Chou, C. L. Hwang and Y. T. Wu, "Effect of exercise on physical function, daily living activities, and quality of life in the frail older adults: a meta-analysis," *Archives of Physical Medicine and Rehabilitation,* vol. 93, no. 2, pp. 237-244, 2012.

[109] T. Valenzuela, "Efficacy of progressive resistance training interventions in older adults in nursing homes: a systematic review," *Journal of the American Medical Directors Association,* vol. 13, no. 5, pp. 418-428, 2012.

[110] M. R. Daliri, "Automatic diagnosis of neuro-degenerative diseases using gait dynamics," *Measurement,* vol. 45, pp. 1729-1734, 2012.

[111] J. Kamruzzaman and R. K. Begg, "Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait," *IEEE Transactions on Biomedical Engineering,* vol. 53, no. 12, pp. 2479-2490, 2006.

[112] G. J. Hahn and S. S. Shapiro, Statistical models in engineering, New York: Wiley, 1994.

[113] R. G. Willison, "Analysis of electrical activity in healthy and dystrophic muscle in man," *Journal of Neurology, Neurosurgery, and Psychiatry,* vol. 27, no. 5, pp. 386-394, 1964.

[114] R. M. Rangayyan, Biomedical signal analysis: A case-study approach, New York: IEEE Wiley, 2002.

[115] Y. F. Wu and S. Krishnan, "Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis," *Medical & Biological Engineering & Computing,* vol. 47, no. 11, pp. 1165-1171, 2009.

[116] E. Dolatabadi , B. Taati and A. Mihailidis, "Automated classification of pathological gait after stroke using ubiquitous sensing technology," in *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA, 2016.

[117] Y. Wang and M. Yu, "A Study for A study for gender classification based on gait via incorporating spatial and temporal feature matrix," in *2013 Fifth International Conference on Computational and Information Sciences (ICCIS)*, Shiyang, China, 2013.

[118] "Gait Dynamics in Neuro-Degenerative Disease Data Base," [Online]. Available: https://physionet.org/physiobank/database/gaitndd/.

List of Figures

Figure 1.1. Traditional procedures of gait classification. 2
Figure 2.1. High-level flow of supervised learning. 12
Figure 2.2. A 3-level DWT 15
Figure 2.3. An artificial neuron. 20
Figure 2.4. A basic feedforward neural network. 20
Figure 2.5. Illustration of a separating hyperplane with a maximum margin by SVM. 22
Figure 3.1. a) The human anatomical position. b) The movement of hip and knee joints in sagittal plane. 24
Figure 3.2. a) Definition of hip and knee angles. b) Normal hip, knee and ankle angle trajectories in one complete gait cycle. 27
Figure 3.3. (a) Setup of AHRSs on subject for the measurement of the joint angles; (b) Dimensions and orientations of the FSM-9 unit. 32
Figure 3.4. Experimental setup to measure trunk acceleration. a) Placement of FSM-9 AHRS; b) Subject wears a belt for stabilization; c) Overground walking for gait measurement. 33
Figure 3.5. Proposed machine learning framework for gait classification. 40
Figure 3.6. Two architectures for combined classifiers. 43
Figure 4.1. Illustration of hip and knee joint angle signals for 13 gait cycles. 48
Figure 4.2. Two gait cycles with same peak values and ROM but inequivalent shapes. 50
Figure 4.3. Flowchart of the classification scheme for joint angle signals. 51
Figure 4.4. Gait cycle segmentation for knee joint angle signal. 53
Figure 4.5. A gait cycle of hip and knee joint angle trajectories before and after resampling using PCHIP. 54
Figure 4.6 Gait cycle pairing for sample generation. 55
Figure 4.7. Two gait cycles with phase shifting before and after applying DTW. 58
Figure 4.8. Influence of number of data points on the results of distance measures 60
Figure 4.9. Illustration of GMM as shape descriptor for hip and knee joint angle signals 63
Figure 4.10. Illustration of feature matrix. 64
Figure 4.11. Block diagram of the LOSO validation. 66
Figure 4.12. Histograms of the most significant variability features for HC and PT classes. 68
Figure 4.13 Averaged trajectories for HC and PT classes and the centers of Gaussians. 70
Figure 4.14 Classification performance comparisons for feature selection methods. Left: t-test; Right: PCA. 74
Figure 4.15. Experimental setup with functional orthosis and IMUs. 77
Figure 4.16. All the hip and knee cycles for three walking tests of one subject. 78
Figure 4.17. Standard form used in clinics for evaluating the daily gait quality of PD patients. 79
Figure 4.18. HSP patient during SRT-based exercise 81
Figure 4.19. MOPASS gait rehabilitation system from frontal and back sides. 83
Figure 4.20. Joint angles of subject performing slow walking with and without MOPASS platform. 84
Figure 5.1. Example of trunk acceleration signals collected from a PD patient. 86
Figure 5.2. Flowchart of the classification scheme for trunk acceleration signals 88
Figure 5.3. Sample acceleration signal in V direction. a) raw signal; b) signal after removing mean component; c) signal after applying median filter; d) signal after normalization. The unit for signals in a) – c) is $m/s2$. 89
Figure 5.4. Example of step segmentation results on a trunk acceleration signal in vertical direction. 90
Figure 5.5. Sliding window for sample generation. 91
Figure 5.6. Fitting acceleration excursion in ML and AP directions with an ellipse using least square based regression 96
Figure 5.7. Experiments for collecting trunk acceleration signals. 97
Figure 5.8. Ellipse for HC and PT groups. 101
Figure 5.9. Classification results with t-test based features selection. 104
Figure 6.1. Example of raw stride time, swing time and double support time interval signals. 110
Figure 6.2. Flowchart of the classification scheme on stride interval signals. 112
Figure 6.3. Sample swing interval series before (upper plot) and after (lower plot) replacing outliers. 113
Figure 6.4 Classification results with PCA based features selection. 124

List of Tables

Table 2.1 Confusion matrix of a binary classification. 10
Table 3.1. Definition of sub-phases in one gait cycle and their intervals. 26
Table 3.2. Summary of four types of gait parameters. 28
Table 3.3. Common pathological gait patterns.. 29
Table 3.4. Comparison between different IMU models.. 31
Table 3.5. Summary of the state-of-the-art researches for machine learning-based gait classification. 39
Table 4.1. Summary of state-of-the-art researches for gait classification using kinematic data 49
Table 4.2. Indexes and denotations for variability features. 61
Table 4.3. Indexes and denotations for all shape features. 63
Table 4.4 Statistics on features for joint angle signals. 69
Table 4.5. Intra-subject and inter-subject variability. 71
Table 4.6. Classification results on samples-level and subject-level using SVM. 72
Table 4.7 Classification results using two feature selection methods. 75
Table 4.8. Classification results on simulated impaired gait 77
Table 4.9. Evaluation results of medication effect on PD patients using the framework. 80
Table 4.10. Classification results and clinical test results on HSP patient during SRT. 82
Table 5.1 Time and frequency domain features. 94
Table 5.2. Statistics on features for ML dimension. 98
Table 5.3. Statistics on features for AP dimension. 98
Table 5.4. Statistics on features for V dimension. 99
Table 5.5. Statistics on features for M dimension. 99
Table 5.6. Statistics on contour features. 99
Table 5.7 Classification results on sample and subjects 102
Table 5.8 Results for classification of PD and PNP classes. 105
Table 5.9. Classification results on PD patients 1 for monitoring medication effect. 105
Table 5.10. Classification results on PD patients 2 for monitoring medication effect. 105
Table 5.11. Classification results on PD patients 3 for monitoring medication effect. 106
Table 5.12. Classification results on PD patients 4 for monitoring medication effect. 106
Table 6.1. Statistical features. 115
Table 6.2 Statistics on original data and extracted features for stride interval signals. 120
Table 6.3 Classification results on sample level using ANN as classifier. 121
Table 6.4 Classification results on subject level using ANN as classifier. 122
Table 6.5 Classification results on sample level using the SVM. 125
Table 6.6 Classification results on subject level using the SVM. 126
Table 7.1. Summary of the classification work in Chapter 4-6. 128